

# EXAMINING THE DIFFUSION OF USER-GENERATED CONTENT IN ONLINE SOCIAL NETWORKS

Jeong-ha Oh

Anjana Susarla

Yong Tan

Department of Information Systems and Operations Management,  
Michael G. Foster School of Business, University of Washington, Seattle  
{jhoh, asusarla,ytan}@u.washington.edu<sup>1</sup>

July 2008

---

## Abstract:

This paper is motivated by the success of You Tube, which is attractive to content creators and media companies for the potential to rapidly disseminate digital content. The tremendous variation in the success of videos posted online and the networked structure of interactions on You Tube lend itself to an inquiry about the role of social influence on content diffusion. Using a unique data set of video information and user information collected from You Tube, we find that evidence for a number of mechanisms by which social influence is transmitted, such as a preference for conformity, social learning and the role of innovators or opinion-makers. Such mechanisms of social interactions can play a huge role not only in the success of user-generated content, but also on the magnitude of that impact. 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 of content. Econometrically, the problem in identifying social influence is that individuals' choices depend in great part on 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 robust to potential self-selection according to user tastes, temporal heterogeneity and the reflection problem. Implications for researchers and managers are discussed.

**Keywords:** Diffusion, Social Contagion, User-Generated Content, Social Networks, Reflection Problem

---

<sup>1</sup>Authors' names listed in alphabetical order. We thank seminar participants at the Workshop in E-Business 2007, Singapore Management University, Towson University, the University of Washington Information School and the Symposium on Statistical Challenges in eCommerce Research for comments on earlier versions of this paper.

## 1. Introduction

Enabled by Web 2.0 technologies, social computing models that allow companies to tap into the “wisdom of the crowds” (e.g., Surowiecki 2004), are proliferating in a number of consumer markets. Social computing has increased consumers’ access to information and greatly expanded the available set of choices to consumers. At the same time, user-generated content creation is shifting power to the edge of the network, allowing users to participate in innovation in a way that would have been unthinkable even a decade ago. This paper is motivated by the success of YouTube, which is a venue for content creators to interact with networked communities of users. YouTube provides a particularly attractive context to explore social influence. Unlike earlier online communities that only had a few contributing content while most people were consumers of content, the usability and functionality, well as the ease with which videos can be shared on YouTube has made it tremendously popular. The networked patterns of interaction in YouTube also differ from other phenomenon of user-generated content such as online reviews, since the very nature of participation on YouTube becomes a method of social interaction.

YouTube provides a wealth of opportunity for content creators to express themselves and promises to be a conduit for socially engaged individuals to share their preferences with others, ultimately promising to transform how consumers engage with popular culture. The YouTube model therefore has the potential to fundamentally alter the structure of industries that deal with digital products such as media and entertainment. While this model is attractive to content creators and media companies alike for the potential to disseminate product-related information and digital content effectively in a short period of time, evidence suggests, however, that all content created on YouTube is not equal. A handful of clips acquire Internet superstar status while most videos languish in obscurity. The unpredictability of success in cultural markets has been attributed to the role of social influence (Salganick et al. 2006), prompting us to question whether social influence has a similar role to play in the diffusion of content on YouTube. Indeed, the well-publicized failure of media companies’ web initiatives, such as the Innertube entertainment portal of CBS (Barnes 2007), to attract users is the considerable uncertainty in predicting the success of web content. The uncertainty in predicting user interest suggests the potential for social interaction effects, where an individual’s preferences or actions depend on the decisions on others.

Social contagion refers to a phenomenon whereby an actor's decision on the adoption of a new product is dependent on other actors' attitudes, knowledge, or adoption of the new product (Van den Bulte and Lilien 2001). Dodds and Watts (2004) argue that individuals most susceptible to social contagion are enormously influential in the dynamics of diffusion. Prior research has highlighted a number of mechanisms by which social influence occurs. Given that individuals value conformity with others (Bernheim 1994), "peer group" effects, or social influence proximate others can substantially influence individual behavior (e.g., Sacerdote 2001). Another mechanism by which individual choices are impacted by others is through social learning. When faced with substantial uncertainty in sampling of new products, social learning occurs through observing choices of neighbors (Ellison and Fudenberg 1993). This paper is intended to be a step forward in understanding the characteristics of digital content diffusion within a social network structure. Using a unique data set of video information and user information collected from You Tube, we find that mechanisms through which social interactions are structured plays an important role not only in the success of user-generated content, but also on the magnitude of that impact. The questions we seek to answer are:

1. What are the effects of a user's social network structure on the diffusion of You Tube content?
2. Among several embedded social network structures, which has the most influence on diffusion?

Identifying social influence effects poses a number of statistical and econometric challenges. First, individual user preferences may be subject to popularity surges that create a serial correlation. Second, identifying social influence is complicated by the fact that an individual's choices reflect the choices of a social group of which an individual belongs to. The difficulty in estimating social influence 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 social influence in economics (Manski 1993). A third problem is that of unobserved user heterogeneity that reflects the unobserved social preferences of users or demographic characteristics such as age. In other words if the distribution of user tastes results in a greater propensity to be loyal to a particular channel on You Tube, then the self-selection into groups may dictate the popularity of content on YouTube. We therefore need to consider potential user self-selection depending on unobserved social preferences.

Prior research has suggested that the growth of Internet has had tremendous impact on consumer-facing activities and industries. File-sharing technologies have shifted consumer demand for albums, as documented by the survival of albums on the Billboard charts (Bhattacharjee et al. 2007). The Internet has been instrumental in promoting online dissemination of product reviews and buyer feedback, with tremendous implications for customer-facing activities such as customer acquisition and retention (e.g., Dellarocas 2003). There has been a considerable amount of research that has examined the impact of online word of mouth on Internet commerce (e.g., Chavalier and Mayzlin 2006, Dellarocas 2003). The growth of online social communities poses new challenges to researchers in explicating the role of social structures driving the success of user-generated content.

This paper can make the following contributions to literature. We integrate perspectives from prior research on diffusion, social networks and the research on user-generated content. Fichman (2000) notes that a considerable amount of literature in IS has considered the impact of diffusion on individuals' adoption decisions rather than on aggregate dynamics. By contrast, we explicitly consider the role of the social network structure as a mechanism for the transmission of information and for the dynamics of contagion. Since we are interested in the context of a market for experience goods such as music and entertainment, there is considerable amount of uncertainty regarding the actual experience that is likely to result, which makes the mechanism of information transmission especially relevant. Further, most models of diffusion, such as the Bass model (Bass 1969), do not identify the mechanism by which social influence occurs. Another contribution of this paper is to highlight the role of various mechanisms through which social influence spreads. For instance, the S-shaped diffusion curves could result from user heterogeneity in the propensity to adopt (Van den Bulte and Stremersch 2004). By contrast, we can disentangle the impact of social contagion from that of user heterogeneity, which is considered to be a limitation of a significant amount of prior literature on diffusion (e.g., Van den Bulte and Stremersch 2004). Further, in line with prior literature on diffusion suggests that the diffusion process before a critical mass is reached may be subject to different influences from that after a critical mass, we can distinguish between influencers in the different stages of diffusion.

The structure of the paper is as follows. We review the prior theory and present the hypotheses and

research model in Section 2. In Section 3, we discuss the data collection process and operationalization of measures. Section 4 presents the empirical approach, Section 5 presents a discussion of the results and Section 6 provides concluding remarks.

## **2. Theory and Hypotheses**

### **2.1 Prior Literature**

The Bass model (Bass 1969) has been extensively used to model the diffusion process in the marketing literature (e.g., Mahajan et al. 1990). Researchers have also examined the impact of word-of-mouth effects on the diffusion of a new product, such as Dodson and Muller (1978) who develop a diffusion process model to incorporate the communication effect between customers and the advertising effect. In sociology, social network methods have provided an important framework to study the diffusion of innovations (e.g., Wejnert 2002). A classic study in this steam is that of Coleman et al (1966) that analyzes the social contagion effects on physicians' adoption of new drugs. Burt (1987) emphasizes the importance of structural equivalence to the process of contagion. Strang and Tuma (1993) find that network centrality does not affect the intrinsic propensity of physicians to adopt an innovation, but does have social contagion effects via increased susceptibility to others' adoptions.

Researchers across several fields such as marketing, economics and sociology have been interested in the analyses of social interactions where individuals' actions or behavior depends upon the actions or choices of other actors (e.g., Banerjee 1992, Bala and Goyal 1998, Ellison and Fudenburg 1993, Granovetter 1973, Manski 1993, Van den Bulte and Lilien 2001). Social interactions underlying individual behavior have been extensively analyzed in economics in the economics literature on "peer effects" (e.g. Bandiera and Rasul 2006, Sacerdote 2001). Increasingly, marketing researchers have incorporated an analysis of social interactions into the diffusion of innovations. Van den Bulte and Lilien (2001) find that when marketing efforts are controlled for, contagion effects disappear. With the explosion of online communities, researchers in a variety of fields are interested in the phenomenon of information diffusion over the Internet and the role of social interactions underlying such processes of information transmission (e.g., Dodds et al. 2003).

## 2.2 The Context

As a YouTube user, one can establish a relationship with other users either as a friend or a subscriber. A friend relationship is initiated by an invitation from one person to another, and for the relationship to be confirmed, it requires an agreement from the other person. Thus, it is more likely that a friend relationship on YouTube represents the pre-existing social ties between individuals, either from friendship based on real life, or friendship developed online. Since the friend network we observe is the result of such mutual agreements, we characterize friendship networks on YouTube as an undirected network (e.g., Newman 2003). The friend network is a community of peers and the set of friends act as proximate actors driving content adoption on YouTube.

A subscriber network, by contrast, results from the network of users who subscribe to a channel (a user). Subscriber relationships do not require mutual agreements. A user can freely add another user into their list of subscriptions, and the action of subscription indicates the willingness to visit and watch the contents of a channel, which is how interested they are in the videos uploaded on the channels, i.e., based on the likes and dislikes of a user rather than social ties. Since the decision on whether to add a channel to a subscription or not is based on each user's taste preferences, subscriber network structure can be considered to be representative of user tastes. When a new video is posted, all the subscribers of a channel are alerted through email or RSS feeds. While friend relationship is based on mutual social agreement, a subscriber relationship is a one-way, or directional, relationship based on tastes of users.

We therefore identify three distinct channels of social influence on YouTube. First, we observe that there are networks of friends within the community of interest, which we use to build a friend network. Second, we observe friendship ties between users across different communities. Third, we observe that there are networks of subscribers within the community of interest, denoting social networks based on instrumental ties, i.e., a similar purpose in exploring content. The friend networks on YouTube denote links between individuals who identify themselves as friends, as distinct from the social networks of subscriber relationships, which have a different pattern of affiliation based on shared interests in viewing videos. Subscribers and friends can also rate and comment on videos in addition to adding videos to a list

of favorites, and these choices can also serve as signals to other users, driving the popularity of content. The YouTube setting thus naturally lends itself to a study of social influence in that the social network structure dictates the diffusion of content.

It has been suggested that friendship ties are characterized by greater frequency in interaction compared to other types of ties (Granovetter 1973); therefore, we expect that influence from friends within a network, that we characterize as a local network of friends (e.g., Watts et al. 2002)<sup>2</sup>, as well as influence from friends outside the community, or non-local friends, is qualitatively different from instrumental ties characterizing subscriber networks. We do not observe substantial overlap in membership between these networks, which bolsters our argument about the difference between these networks. The greater interaction and interpersonal influence characterizing these networks increases the opportunity for the transmission of social cues through the network structure (Salancik and Pfeffer 1978). Further, we expect that friendship ties between individuals who are also linked through shared tastes, i.e., the network of friends within a community, exert substantially different type of social pressures compared to patterns of friendship alone. Friendship between individuals within a community of interest on YouTube may arise from similarity in personal characteristics and denotes that individuals may have consistent interests and tastes, which in turn promotes homophily (e.g., Newman 2003), also known as assortative mixing. The strong sense of identification resulting from such patterns of interaction provides greater potential for friendship networks in persuading or influencing the perceptions of actors connected through network ties (Rogers and Kincaid 1981).

### **2.3 Hypotheses**

We consider the impact of social influence in the diffusion of user generated content. The social influence models that we anchor upon consider a rich set of explanations for social influence, such as the desire for social conformity, learning from neighbors (Bala and Goyal 1999) and peer effects (Sacerdote 2001) that dictate preferences. In addition to the diversity of factors driving content diffusion, we also consider the temporal heterogeneity in the diffusion process, which refers to the time-varying influence of factors that

---

<sup>2</sup>Watts et al. (2002) argue that group membership is a primary basis for social interaction, and possess local information about a network, i.e., knowledge about the immediate circle of acquaintances.

drive diffusion (Strang and Tuma 1993). In particular, we consider the potential for the dissipation as well as magnification of social influence over time.

While the question of social influence has been explored in earlier literature (e.g., Armstrong and Sambamurthy 1999, Kraut et al. 1998), there are some crucial distinctions between prior diffusion literature and this study. First, we differ in the role played by early adopters. The literature on product diffusion distinguishes between well-informed early adopters and the late adopters who observe others' actions, wherein early adopters learn by doing while late adopters learn from others, also referred to as social learning (e.g., Bandiera and Rasul 2006 who explicitly distinguish between the two models of learning). By contrast, we posit that the influence of early adopters is derived from their relative position in the social network. Specifically, our focus is on the network structure in that it enables information transmission. Second, we consider diffusion in an online community context where interactions are structured through a social network. Third, a substantial stream of product diffusion models does not distinguish between social influence and the differences in tastes of the users. Fourth, we consider the role of the structural position of actors in promoting content; it has been suggested, for instance, that marketing efforts can confound the role of social contagion in the diffusion process (Van de Bulte and Lilien 2001). Finally, unlike the models of diffusion, we attempt to distinguish between different explanations for contagion, such as social learning, conformity preference and communication.

### **2.3.1 Conformity or Normative Preferences**

It has been suggested that peers exert considerable social pressure that influences individual behavior (Sacerdote 2001). Individuals care deeply about the opinions of others they interact with, creating a pressure to conform to choices of others (Bernheim 1994) and face dissonance when they do not adopt the choices of individuals whose approval they seek (Coleman et al. 1966). Since an actor may influence proximate actors' opinions on a new product, we consider the centrality of a user in the friend network in determining social contagion due to conformity. Central actors act as conduits of information (Wasserman and Faust 1994) and occupy opinion-making role that influence the perceptions and decisions of others who value their judgment (Ibarra and Andrews 1993). Actors who are connected to a greater number of actors also have greater awareness of the choices of their peers, and signal their preferences to other



actors through their choices, inducing the latter to view the same video; thus central actors influence proximate actors' decisions due to conformity pressures. Individuals who occupy a key position in a network defined by friendship ties therefore may have more power of influencing perception and persuading others to follow their choices. Thus, we hypothesize that:

**Hypothesis 1:** *Social influence as a result of conformity preferences significantly affects the diffusion of content.*

Paradoxically, while central actors can be influential in persuading others, their central network position can also constrain the forms of influence. The close interactions characterizing friendship ties also create a reputation cost that deters individuals from expressing inappropriate opinions or behavior (Granovetter 1985). In linking to a new video, central actors in a friend network risk losing their social capital in recommending un-tested or un-proven content. Once a video acquires a critical mass of views, centrally connected actors in a network do not have to jeopardize their reputation by promoting content that may risk alienating them from others. The choices of central actors are then reiterated by other actors who value conformity, and the result is a social multiplier effect, which strengthens the social influence of central actors, enhancing the effect of conformity preferences. Thus,

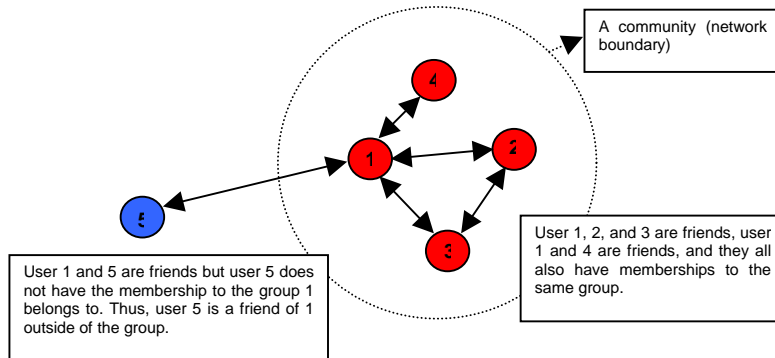
**Hypothesis 2:** *Social influence from centrally connected actors has a positive impact on the diffusion of content in the later stages of the process of diffusion.*

### **2.3.2 Social Learning**

A video posted on You Tube is an experience good, and not a search good; thus, it is characterized by substantial uncertainty in terms of whether viewers will favorably react to it or not. The value of such a product is revealed only after direct experience (Nelson 1970). The uncertainty in predicting user experiences coupled with the range and the depth of offerings in each category and the growth of titles in You Tube create a cognitive overload for a potential viewer. Such assumptions are consistent with empirical findings that consumers in online settings face considerable information overload (Brynjolfsson and Smith 2000). Social information processing or social learning is particularly important in shaping perceptions under conditions of uncertainty or ambiguity (Bala and Goyal 1998). As the uncertainty in product quality increases, boundedly rational consumers can learn from neighbors' actions (Bala and

Goyal 1998, Banerjee 1992). Information transmission occurs through a network structure whereby agents who observe their neighbors' choices use simple rules of thumb, such as popularity weighting, in deciding whether to adopt a product or not (Ellison and Fudenberg 1993). Figure 1 depicts the network relationships within and outside the group boundary.

**Figure 1.** Friend Relationships Inside and Outside a Group Boundary



We consider the impact of non-local friendship ties in facilitating social information transfer between proximate actors. Burt (1992) classifies a network contact as non-redundant if an individual does not share ties with other contacts in a person's immediate social network. The role of non-local friendship ties in disseminating information is then similar to that of word of mouth effects (e.g., Chevalier and Mayzlin 2006). While there is a potential for learning from friendship ties within the local network, i.e., within the community, the strength of identification and conformity preferences characterizing such groups may also increase the redundancy of information within the local network (e.g., Burt 1992). The cognitive overload involved in choosing between different videos, the time required to sample a variety of videos coupled with the substantial uncertainty in the resultant experience from a video strengthen the importance of information transfer or social learning from non-local networks of friends.

**Hypothesis 3:** *Social learning from individuals outside the local network of peers has a significant impact on the diffusion of content.*

The novel information from non-local ties increases the diversity and richness of informational cues (Burt 1992) within the local network, providing an avenue for social learning. Actors that are more connected to individuals outside the local network have access to new and non-redundant content, which

is then disseminated into a cohesive network of peers. Given the potential for conformity preferences within the local network, we hypothesize that the impact of non-local ties is greatest when the local network structure can play an important role. Thus, we hypothesize:

**Hypothesis 4:** *Social influence from individuals outside the local network of peers has a significant impact on the later phase of content diffusion.*

### **2.3.3 Social Influence from Innovators and User Tastes**

On You Tube users can create their own page or channel and subscriber to or accept subscriptions from other channels. In other words, subscribing to a channel is the same as subscribing to a user.<sup>3</sup> A channel is personalized to each user and allows for users to display content that they uploaded, videos from other members, videos favorited by them and channels that they have subscribed to, and channels that subscribe to their content. There can be several motives for content creation and posting content such as peer recognition (Resnick et al. 2000), self-expression (Raymond 2001) and identity affirmation (Forman et al. 2008). The shared interest in sampling music characterizing social networks of subscriber groups provides an opinion-making role to subscribers that occupy key roles in the social network. In particular, we consider the degree centrality of individuals in a subscriber network, which denotes the extent to which individuals are extensively involved with or adjacent to many other actors in the social network (Wasserman and Faust 1994). We also distinguish between the in- and out-degree centrality of actors. Subscribers with several in-degree connections may be more popular, and may be more likely to disseminate preferences among a wider group of actors, increasing the rate of content diffusion. Subscribers with several out-degree connections may be more gregarious, with the result that they can promote awareness to a greater number of actors in the network, increasing the rate of content diffusion.

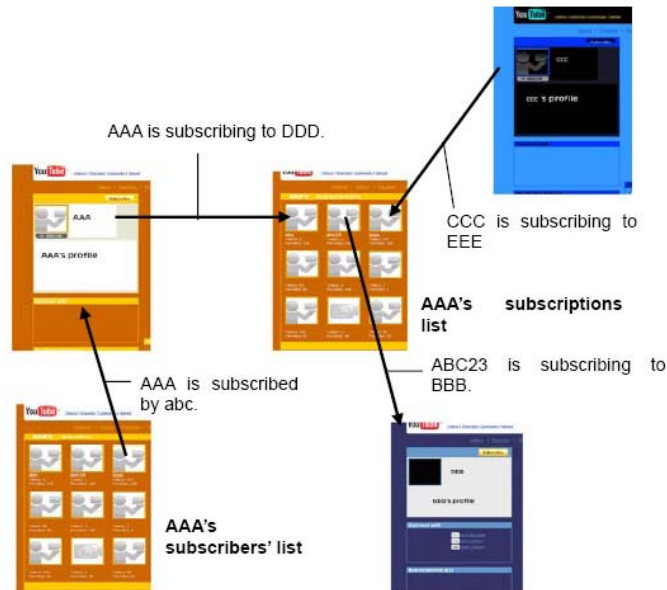
Subscribers that are more connected are (a) characterized by greater openness to new content and thus likely to be exposed to a variety of content, (b) are more likely to be on the *cutting edge* (Rogers 1995) and therefore informed about newer types of content and (c) broadcast their preferences to a group of networked peers, thus disseminating their preferences among the social network. Centrally connected users in a subscriber network are therefore influential in transmitting information about new videos,

---

<sup>3</sup> We use the terms channel and user interchangeably in the paper.

acting as opinion leaders who influence the spread of information about an innovation (Bass 1969, Rogers 1995, Mahajan et al. 1990), which in our context refers to new videos that have not acquired recognition from the overall YouTube audience. While the literature on online word of mouth and the impact of user reviews has investigated how opinions of peers affect the popularity of content, we are interested in the mechanism by which communication links influence the spread of information in a network structure. Figure 2 depicts a representative subset of nodes in a subscriber network.

**Figure 2.** Subscribers' Relationship



**Hypothesis 5:** *Subscribers who are more connected are likely to have a greater impact on the diffusion of content.*

Opinion leaders are more likely to adopt newer ideas or products compared to others, which gives them a pivotal role in influencing others and promoting new ideas or products in the early stages of diffusion (e.g. Rogers 1995). Actors who occupy a central position in the subscriber group have early access to a wide variety of new content, which increases their likelihood of viewing a new video. Thus the long-term success of content depends critically on the choices of initial adopters that are centrally connected.

**Hypothesis 6:** *Subscribers who are more connected are likely to have a greater impact on the initial phase of content diffusion.*

### 2.3.4 Heterogeneity in User Tastes and Group Formation

Prior literature suggests that diffusion curves may also result from differences in user propensities to adopt a new product, rather than social contagion (Bemmar 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 may create a self-selection of users into different channels or interest groups. While we do not hypothesize the impact of user tastes on group formation, we control for this possibility in the empirical estimation.

## 3. The Data

### 3.1 Defining the Network Boundary

Many prior studies of social networks have relied on snowball sampling methods, where referral data is collected starting from an initial target. While this method is efficient to locate hidden populations (Salganick and Heckathorn 2004), it poses challenges in sampling such as selecting the right initial target. Delimiting the sample size and demarcating the network boundary also becomes problematic. Because we cannot exhaustively cover the whole network, one of the problems in snowball sampling is to consider the right approach in demarcating the boundaries of the sample. Artificially limiting the sample size also creates the problem that some of the observable population characteristics may be omitted. Another problem with snowball sampling is that it fails to identify isolated incidences.

We select our sample by focusing on a set where the network boundary is predefined by focusing on a community in YouTube, which is an interest group. A *community* in YouTube is defined as a group with specific video categories (there are total thirteen categories<sup>4</sup> in YouTube). For example, a community listed under the ‘music’ category is a group of people who share the same interest, which is ‘music’, and upload videos related to ‘music’ within the group. Any person interested in ‘music’, and wants to share their videos with other members in the group can join the community. Therefore the interest group itself

---

<sup>4</sup> The existing thirteen categories are Autos & Vehicles, Comedy, Education, Entertainment, Film & Animation, Howto & Style, Music, News & Politics, People & Blogs, Pets & Animals, Science & Technology, Sports, and Travel & Events.

forms the network boundary. Our sample community (or group<sup>5</sup>) is drawn from the ‘music’ and ‘people & blog’ categories, which not only is representative of the characteristics of YouTube as a medium of sharing video clips but also truly representative of the social interactions at work within YouTube. The group of individuals belonging to ‘people & blogs’ is characterized by greater interpersonal interactions and the ‘music’ category is one of the most popular channels for sharing videos, guaranteeing high activity level. As described in Table 1, the group is 2 years old, and has 1558 group members, and 4106 videos. There are 15 new videos on average uploaded every day in this group, and on average 5 new members join the group every day.

**Table 1.** Data Set Descriptions

Age of the Group	2 years
Number of Group Members	1558
Average growth of the number of videos per day	15
Average growth of the number of members per day	5
Number of videos	4106
Number of users (Channels that posted videos)	913
Data collection duration	2 months
Number of observation points in time	11

### 3.2 Data Collection and Description

We use a panel data consisting of video information and user information collected from YouTube.com, over a period of 2 months. Our sample only focuses the videos uploaded within the group of our interest, and the members of that specific group. We have total 4,106 videos posted by 913 users. The data was collected with 11 observation points in time, five days apart. At each observation point, the information on each video and each user within the group was collected in a screen shot manner, and was tracked repeatedly over time every five days. Table 1 summarizes these descriptive measures of the data set. For each user, the complete list of friends, subscribers, and subscriptions has been collected. These lists of friends, subscribers, and subscriptions are also tracked repeatedly over time as multiple events. Since this data collection was repeated for each data point, we get the snapshots of the network structure over time.

The average age of a video, which is 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

<sup>5</sup> In YouTube, the terms group and community are synonyms.

247,455. This indicates a high dispersion of the popularity of video clips varying from minimum 4 clicks to maximum 13,449,210 clicks. To control the skewness, we used a log-transformed number of views, and normality tests confirm that the distribution of the number of views is normal after log-transformation. Although we cannot observe how many times a specific viewer watches a certain video repeatedly, we can assume that there is a reasonable limit to how many times each individual watches a video. Since we take logs, any bias that is caused by repeated viewings is only going to be a slight downward revision of the estimates. We also gather data on the number of links to a video, which are the external links leading to a video clip and which listed by the user posting the video content. With the number of clicks coming from each link, the number of outer links provides partial control over the traffic coming from outside YouTube. Table 2 shows the data summary.

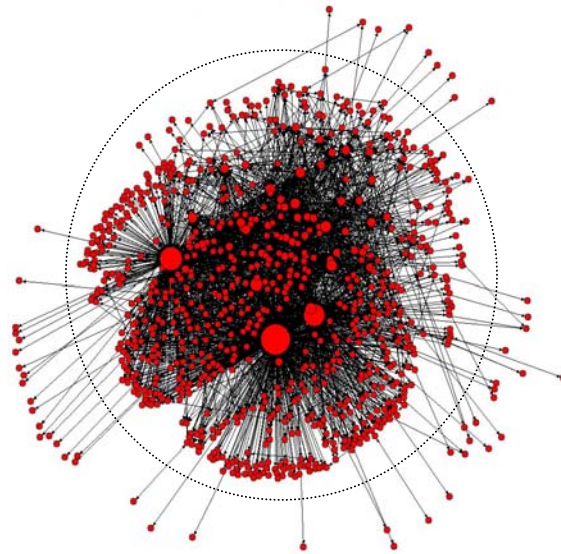
**Table 2.** Data Summary

Variables	Mean	St. Dev.	Min	Max
Age of video (days)	212	144.26	0	731
Number of times a video is watched	14180.80	247455	4	13449210
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
Age of Channel (days)	379.96	171.29	2	846

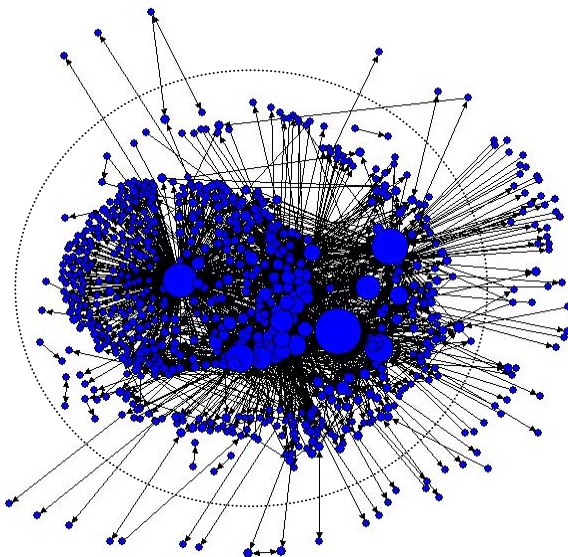
### 3.3 Social Network Structures on You Tube

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 that is indicative of user tastes. The network boundary serves another purpose in addition to defining the sample limit. It also provides a boundary to distinguish close within-group social influence from outside-group influence. Friends interact more if they also share common interests, and their influence on a specific subject of interest will differ from the influence of friends who do not share that specific subject. In other words, the role of social influence will be different depending on the type of network ties. Since we have the total number of friends and subscribers (subscriptions), by constructing the network map within the interest group, we not only know the social influences and users’ preferences coming from within the local network, but also know the direct ties coming from outside the local network. Figure 3 shows a subset of the users forming

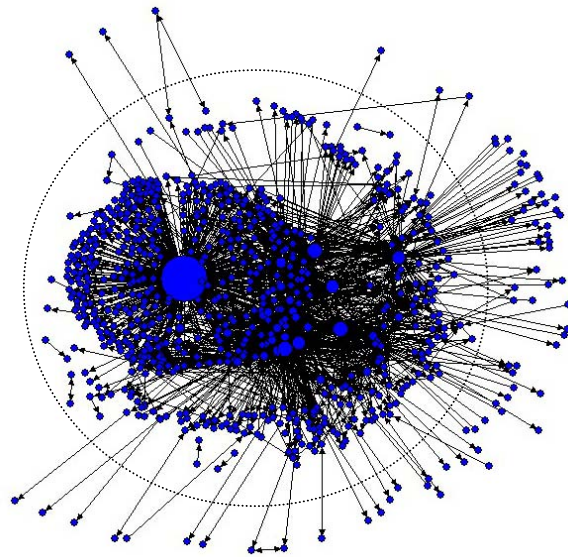
this friend network and Figure 4 shows a subset of the subscriber network.



**Figure 3.** Friend Network, Group Boundary, and Outside Network



**Figure 4.1.** Subscriber Network (Node size based on in-degree centrality)



**Figure 4.2.** Subscriber Network (Node size based on out-degree centrality)

### 3.4 Dependent Variable

The standard Bass model estimates the growth of aggregate demand or diffusion rate as a function of the aggregate demand of prior time period as well as the time elapsed from the initial launch of a new product. Following the standard the diffusion model, we set the dependent variable as the diffusion rate of each video. The diffusion rate is 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  $t-1$ ,  $\Delta v_{ijt} \equiv v_{ijt} - v_{ijt-1}$ . Our measure of aggregate demand is the popularity of each video, which is represented by  $v_{ijt}$  (Bass 1969). Here,  $v_{ijt}$  is the aggregate number of times a video has been watched, in other words, the aggregate number of times a video has been clicked. The subscripts  $i, j, t$  are for video  $i$ , user  $j$ , and observation time  $t$ . Thus, the dependent variable measures the popularity growth of video  $j$  posted by user  $i$  from time  $t-1$  to time  $t$ .

### **3.5 Independent Variables**

#### **3.5.1 Social Network Measures**

We calculate social network measures using UCINET 6 (Borgatti et al. 2002). The degree centrality (Wasserman and Faust 1994), which is calculated from the number of direct ties, measures the size of the proximate network of an actor. We present the graph theoretic interpretation of this measure in the appendix. The degree centrality of a certain node in a network (user channel, in this case) is the node (channel) characteristic that captures the ability and the opportunity of a node<sup>6</sup> (a user in the social network) to diffuse information (videos, in this case). Wasserman and Faust (1994) argue that the more central the network position of the actor, the more the actor is a channel of relational information to others. An actor with a high degree centrality denotes where “the action is” in the network (Wasserman and Faust 1994:179). The actors with higher degree centralities have greater opportunities to disseminate information because they have more ties and more choices. Having more ties also makes them less dependent on any specific other node. Thus, central actors occupy a position of social influence and the centrality measures denote the social capital of an actor, consistent with Burt (1987).

For the friend network, we calculate two measures to denote the social influence of an actor. First, we calculate the degree centrality measure (Wasserman and Faust 1994), which indicate the importance of a node (channel) in the social network, as discussed above. We also calculate the Bonacich power (Bonacich 1987), which is based on the insight that (i) an individual’s status is a function of the status of other actors who the actor is connected to, and (ii) connections to many prominent (powerful) others

---

<sup>6</sup> Consistent with social networks literature, a user is an ‘actor’ or ‘node’ while the relationship between users is a ‘link’ or a ‘tie.’

reduces the power of a node. For instance, a channel A and channel B have  $n$  connections but B's connections are well connected with others in the network while A's connections are more isolated. While B is clearly more central, she may have less influence over her proximate connections that are less dependent on B for access to information. The Bonacich power captures the insight that power can be increased as well as reduced depending on the structure of the network by connections to powerful others. The Bonacich power measure is a modified version of closeness centrality (Freeman 1979) and calculated iteratively by assigning each actor an estimated centrality equal to their own degree, plus a weighted function of the degrees of the actors to whom they were connected. Since the subscriber network is a directional network, we calculate the in- and out- degree centralities of users (Wasserman and Faust 1994).

The social network measures therefore are  $Frn\_NrmDeg_{jt}$  as the degree centrality of user  $j$  in the friend network at time  $t$ ,  $Subs\_NrmOutDeg_{jt}$  as the out-degree centrality of user  $j$  in the friend network at time  $t$ , and  $Subs\_NrmInDeg_{jt}$  as the in-degree centrality of user  $j$  in the friend network at time  $t$  and  $frn\_NrmBP_{jt-1}$  as the Bonacich power. A user's social network reaches may reach beyond the boundary of the interest group, with the result that 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. Since we have the complete list of friends and subscribers, by constructing the friendship and subscription networks within the group, we can also distinguish the number of ties – either friends or subscribers – established outside of the group. To capture the social influence coming from outside of the group boundary, we measure  $\log NumFrn\_outside_{jt}$ , which is the log-transformed number of friends of user  $j$  at time  $t$ , which corresponds to the degree centrality of user  $j$  outside of the group boundary, as it is a measure of the number of proximate ties. The overall Friend Network Centralization is 35.13% and the descriptive statistics on this social network in our consideration are shown in Table 3. While the minimum number of degree centrality is 0, the maximum degree centrality is 521, indicating that some users are highly connected in their networks, while others are not as connected.

**Table 3. Friend Network Description**

	Degree Centrality	Normalized Degree Centrality
Mean	3.852	0.261
St. Dev.	20.557	1.395
Sum	5682.000	385.482
Min	0.000	0.000
Max	521.000	35.346
Network Centralization = 35.13%		

The descriptive statistics on the subscriber social network are shown in Table 4. To capture a user  $j$ 's taste preferences going outward of the group, we measure  $\log \text{NumOutSubs\_outside}_{jt}$ , which is the number of subscriptions (outward directional ties) user  $j$  forms outside of the group at time  $t$ . To measure the traffic due to user preference coming from outside of the group, we measure  $\log \text{NumInSubs\_outside}_{jt}$ , which is the number of incoming subscribers (receiving ties) of user  $j$  at time  $t$  from outside of the group. Since subscriber relationship is not by mutual agreement, this network is directional, and the in- and out- degree centralities are calculated.

**Table 4. Subscriber Network Description**

	OutDegree Centrality	InDegree Centrality	Nrm OutDegree Centrality	Nrm InDegree Centrality
Mean	1.352	1.352	0.092	0.092
St. Dev.	12.415	3.960	0.842	0.269
Sum	1994.000	1994.000	135.278	135.278
Min	0.000	0.000	0.000	0.000
Max	429.000	71.000	29.104	4.817
Network Centralization (Outdegree) = 29.032%				
Network Centralization (Indegree) = 4.728%				

### 3.5.2 Video Characteristics

Similar to the standard Bass model, we estimate the diffusion rate of a video as a function of time and the cumulative demand for a video. The cumulative demand,  $v_{ijt}$ , for video  $i$ , posted by user  $j$ , at a certain time  $t$  is the cumulative number of clicks to video  $i$  at time  $t$ . As the Bass model estimates the diffusion process as a function of time elapsed from the launch of a new product, we capture the time elapsed as the age of video  $i$ , represented by  $VAge$ , i.e., the number of days since a video has first been posted online.

### 3.5.3 Control Variables

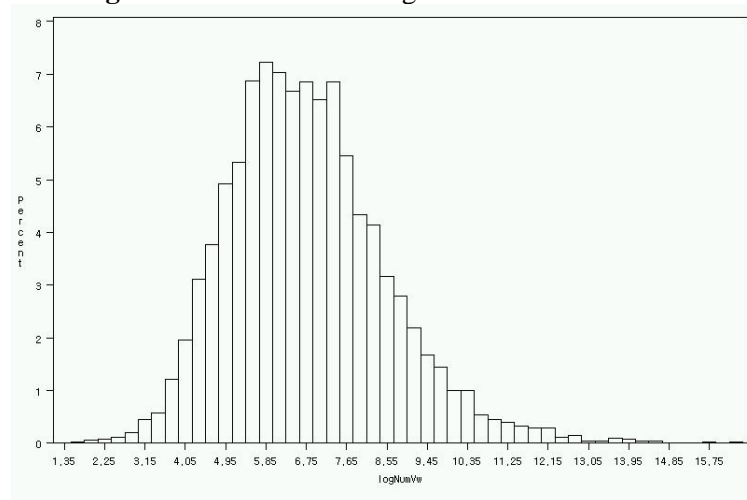
Since the characteristics of a video affect its popularity, we control for the number of external links, *NumOfLinks*, which enable accesses to the videos from blogs, MySpace, Facebook, or other online communities and forums outside YouTube. This variable provides information of the traffic coming from 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.

## 4. Empirical Approach

### 4.1 Baseline Model and Variable Descriptions

Figure 5 represents the distribution of views of a video by the percentage of viewers. To capture the diffusion characteristics of the popularity of a video, we estimate the growth rate  $\Delta v_{ijt}$  as the dependent variable. The popularity of a video  $i$ , posted by user  $j$  at time  $t$ , is measured by the total number of times a video was watched,  $v_{ijt}$ , and the growth rate  $\Delta v_{ijt}$  is the difference of the number of views at time  $t$  and time  $t-1$  ( $\Delta v_{ijt} \equiv v_{ijt} - v_{ijt-1}$ ). This definition is also consistent with Bass model (Bass 1969) and the prior literature on new product diffusion (e.g., Talukdar et al. 2002).

**Figure 5.** Distribution of log-Number of Views



The Bass model estimates the growth in aggregate demand or the rate of diffusion as depending on the ratio of number of innovators and imitators, and the time elapsed since a new product is launched. There are several assumptions underlying the standard Bass model. First, it assumes that diffusion process to be

binary, where the individuals only have a choice of whether to adopt or not, and that the population is homogeneous. Second, in the basic diffusion model, the size of the total population is fixed and known. Third, it assumes that the parameters of external and internal influence do not change over time. In our context, since a video is an experience good, a viewer watching a video on You Tube is equivalent to a consumer adopting a new product and the choice facing a viewer is whether to sample a video or not. We consider the issues of heterogeneity in our structural estimation.

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 influence from within the group ( $Frn\_NrmDeg_{jt-1}$ ,  $Subs\_NrmOutDeg_{jt-1}$ ,  $Subs\_NrmInDeg_{jt-1}$ ) and that from outside of the group ( $\log NumFrn\_outside_{jt-1}$ ). Our baseline model is the standard Bass model that takes into account the social network structure. Subscript  $i$  denotes the  $i$ th video, subscript  $j$  denotes the  $j$ th user, and  $t$  is the time.  $Y$  is a set of covariates that represent the video characteristics, the age of video  $i$  at time  $t$  uploaded by user  $j$ ,  $VAge$ , the number of outer links,  $NumOfLinks$ , and the ratings  $Rating_{ijt}$ . The ratings are posted by registered YouTube users, and may 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. The diffusion equation is stated in the following equation 1, and Figure 5 depicts the distribution of the log-transformed total number of views of the videos. The normal degree centrality of user  $j$  within the friend network, and the In- and Out- degree centrality of user  $j$  within the subscriber network are all user specific. Social ties and user preferences outside of the network are also user specific at time  $t$ . Table 5 summarizes the description of variables and the correlations are presented in Table 6.

$$\Delta v_{ijt} = \beta_0 + \beta_1 v_{ijt-1} + \beta_2 v_{ijt-1}^2 + \beta_3 \log(VAge_{ijt}) + Y_{ijt-1} \cdot \alpha + X_{jt-1} \cdot \gamma + \varepsilon_{ijt} \quad (1)$$

$$X_{jt-1} = \left[ Frn\_NrmDeg_{jt-1} \quad Frn\_NrmBP_{jt-1} \quad \log NumFrn\_outside_{jt-1} \quad Subs\_NrmOutDeg_{jt-1} \quad Subs\_NrmInDeg_{jt-1} \right]$$

$$Y_{ijt-1} = \left[ Rating_{ijt-1} \quad \log NumOfLinks_{ijt-1} \right]$$

**Table 5.** Description of Variables

	Variable No.	Variable Name	Description of Variables
Video Characteristics	1	logNumOfViews	Log number of times a video is watched ( $v_{ijt}$ )
	2	(logNumOfViews) <sup>2</sup>	The square term of the log number of views ( $v_{ijt}^2$ ).
	3	logVAge	(log-transformed) Time elapsed since a video has been posted ( $\log(VAge_{ijt})$ )
	4	Rating	Video Rating (0 to 5) ( $Rating_{ijt}$ )
	5	logNumOfLinks	Number of links, which lead to the video, placed outside YouTube
Social Network Measures	6	Frn_NrmDegree	Degree centrality of a user in the friend network within the group
	7	Frn_NrmBP	Normalized Bonacich' Power in the friend network within the group
	8	Subs_NrmOutDegree	Out-degree centrality of a user in the subscriber network within the group. Connection initiated by the user.
External Social Network Measures	9	Subs_NrmInDegree	In-degree centrality of a user in the subscriber network within the group. Connection initiated by others.
	10	logNumFrn_Outside	Number of friends outside the group
	11	logNumOutSubs_Outside	Number of outgoing subscriptions outside the group
	12	logNumInSubs_Outside	Number of incoming subscribers from the outside of the group

**Table 6.** Correlation Matrix

Variable Number	1	2	3	4	5	6	7	8	9	10	11	12
1	1.0000											
2	0.4041	1.0000										
3	0.3435	0.0268	1.0000									
4	0.2559	-0.1113	0.0723	1.0000								
5	0.5610	-0.0288	0.3009	0.2036	1.0000							
6	0.1197	0.0266	0.0105	0.0435	0.0828	1.0000						
7	-0.0468	0.0130	-0.1227	-0.0120	-0.0460	-0.1169	1.0000					
8	0.0030	-0.0393	0.0017	0.0684	-0.0737	0.1347	-0.1200	1.0000				
9	0.2513	0.1200	0.0424	0.1305	0.0837	0.4554	-0.0880	0.4434	1.0000			
10	0.2374	0.0503	0.0264	0.1809	0.1486	0.3495	-0.1865	0.2866	0.4269	1.0000		
11	0.0614	-0.0364	-0.0529	0.1473	-0.0240	0.1835	-0.0759	0.4646	0.4201	0.6086	1.0000	
12	0.4032	0.1566	0.1347	0.2431	0.1918	0.2832	-0.1002	0.2144	0.6029	0.5573	0.4669	1.0000

\* Variables 1 and 2 are mean-centered

## 4.2 Problems with the Bass model

Identifying social contagion effects poses a number of statistical and econometric challenges that are not taken into account in the standard Bass model. First, individual user preferences, such as watching a video on YouTube, are influenced by the aggregate preferences of all users in YouTube. 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. This 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 a social group of which an individual belongs to. 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. Further, social influence in the You Tube context occurs not only through membership in social networks but also from actors outside the social network. For instance, users communicate not only with networks of friends but also through individuals outside the social network. When an individual interacts with others outside her friend network as well as individuals within the friend network, individual decisions are indicative not only of the social influence of the group an individual belongs to (i.e., the friend network) but also the social influence of others from outside the friend network.

A third problem is that of unobserved heterogeneity in user tastes, which may result from an unobserved demographic characteristic such as age that reflects the unobserved preferences of users. For instance, some users 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 may then result from heterogeneity rather than contagion due to social influence (e.g., Bemmaor and Lee 2002). The structure of network formation in You Tube allows us to address this problem since we can distinguish between a user's memberships in a group of social influence (the friend network) from membership in groups dictated by user tastes. However, we face an additional challenge in that such user heterogeneity in taste could likely affect the composition of social networks. In that case, users with a high degree of incoming subscriptions may 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 are most interested in seeking out a variety of content. The measured impact of a user's connectedness in the subscriber

networks then reflects a systematic pattern of self-selection rather than the result of opinion making or information transmission roles of users that occupy a central role in subscriber networks. In order to rule out self-selection of social network composition driven by heterogeneity in user preferences, we need to understand the factors that dictate the composition of social networks. Given the multiplicity of factors influencing an individual, we distinguish between factors that affect membership in a social network from other types of social influence by conducting exclusion restrictions on the group composition.

### **4.3 Structural Model Specification**

As mentioned earlier, the Bass model does not consider heterogeneity in the susceptibility to diffusion. Further, the presence of serial correlation and endogeneity need to be addressed in the econometric approach. We also need to consider is that of unobserved user and video level heterogeneity. The 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. Third, we conduct Hausman-Taylor estimation (Wooldridge 2002: 225-228) to address potential self-selection into social groups and unobserved relationships between users and videos. We now explain the different stages in detail.

#### **4.3.1 Addressing Temporal Heterogeneity**

Strang and Tuma (1993) note that the population-level diffusion models assume that each individual is equally susceptible to external factors. However, in reality, each individual has different tendencies to react toward contagious factors. Strang and Tuma (1993) introduced a class of individual-level models of diffusion that allow heterogeneity within population and over time by decomposing the diffusion process into two components - the number of individuals at risk of adoption and the hazard rate of adoption for each individual. We similarly address the heterogeneity of each user's susceptibility to contagious factors. 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. We also follow Strang and Tuma (1993) and Wejnert (2002) to incorporate temporal effects.

#### **4.3.2 Rho-Differencing to Remove Serial Correlation**

We now discuss the structural model specification. First, we remove first-order autoregressive effects



from the error term. In the Bass model, when there is a serial correlation,  $\rho$ , in the error term:

$$\Delta v_{ijt} = \beta_0 + \beta_1 v_{ijt-1} + \beta_2 v_{ijt-1}^2 + \beta_3 \log(\text{VAge}_{ijt}) + \varepsilon_{ijt},$$

where,  $\varepsilon_{ijt} = \eta_{ijt} + \rho \varepsilon_{ijt-1}$ . Through a serial correlation adjustment, the autocorrelation effect,  $\varepsilon_{ijt-1}$ , is removed, only leaving random shock,  $\eta_{ijt}$  as the error term.

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

Now, after estimating  $\rho$  from the equation above, we remove the serial correlation by taking the first-order difference, which is the  $\rho$ -differencing procedure. This leaves us with the variables corrected for serial correlation, which are  $\Delta v_{ijt}' = \Delta v_{ijt} - \hat{\rho} \Delta v_{ijt-1}$ ,  $v_{ijt-1}' = v_{ijt-1} - \hat{\rho} v_{ijt-2}$ ,  $(v_{ijt-1}^2)' = v_{ijt-1}^2 - \hat{\rho} v_{ijt-2}^2$ , and  $\log(\text{VAge}_{ijt})' = \log(\text{VAge}_{ijt}) - \hat{\rho} \log(\text{VAge}_{ijt-1})$ .

#### 4.3.3 Hausman-Taylor Estimation with Exclusion Restrictions

One possible estimation problem we need to consider is that users could self-select into different subscriber networks based on their tastes. We address this issue using exclusion restrictions. There could also be unobserved relationships between users and videos that are correlated with the error term. We address these issues using Hausman Taylor estimation (Hausman and Taylor 1981) rather than a fixed effects estimation. In contrast with either fixed effects or random effects estimation, the HT approach assumes that some and not all of the regressors are correlated with individual effects (Hausman and Taylor 1981). This is because, in our case, the exploratory variables we consider are all time-varying. Fixed effects estimation, by contrast, removes all sources of time-invariant variation in the exploratory variables. The exclusion restrictions are part of the Hausman-Taylor (hereafter, HT) approach.

As recommended by Manski (1993), to identify social effects, we need to understand the factors that dictate the composition of social groups. To control for the potential of self-selection according to user taste preferences, we therefore conduct exclusion restrictions using  $\text{Subs\_NrmOutDeg}_{jt}$  and  $\text{Subs\_NrmInDeg}_{jt}$  as instruments for out-degree and in-degree centralities of the subscriber network.

The exclusion restrictions therefore can remove the endogeneity in systematic matching across users and

social networks depending on unobserved taste preferences. As a robustness check we also 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. The test rejected the joint significance of the variables. We estimate the following second-stage models:

$$Subs\_NrmOutDeg_{jt} = f(\log NumOutSubs\_outside_{jt}) \quad (3)$$

$$Subs\_NrmInDeg_{jt} = f(\log NumInSubs\_outside_{jt}) \quad (4)$$

We use the symbols  $X$ ,  $Y$  and  $Z$  to represent:

$$X_{jt-1} = \left[ frn\_NrmDeg_{jt-1} \quad frn\_NrmBP_{jt-1} \quad \log NumFrn\_outside_{jt-1} \right]$$

$$Y_{ijt-1} = \left[ Rating_{ijt-1} \quad \log NumOfLinks_{ijt-1} \right]$$

$$Z_{jt-1} = \left[ Subs\_NrmOutDeg_{jt-1} \quad Subs\_NrmInDeg_{jt-1} \right]$$

In the third stage, we apply a HT estimation procedure. This procedure works as follows: we estimate a random effects model using (i) exogenous time-varying variables as instruments for the endogenous time-varying variables (the exclusion restrictions in the second stage) and (ii) the means of the exogenous time varying variables as instruments for the endogenous time-invariant variables (from the estimation procedure detailed in Wooldridge 2002, pp. 225-228). In other words, the Hausman-Taylor estimation procedure allows us to remove the unobserved heterogeneity of the pairs of video×user relations that may be correlated with the error term, i.e., a user’s time invariant propensity to enjoy a particular type of music that is likely to be correlated with a particular type of channel. We now estimate the diffusion equation with estimated values from the second stage:

$$\hat{Z} = [Subs\_NrmOutDeg_{jt}, Subs\_NrmInDeg_{jt}],$$

$$\Delta v_{ijt} = \beta_0 + \beta_1 v_{ijt-1} + \beta_2 (v_{ijt-1}^2) + \beta_3 \log(VAge_{ijt}) + Y_{ijt-1} \cdot \alpha + \hat{X}_{jt-1} \cdot \gamma + W_{jt-1} \cdot \delta + \log VAge_t \cdot \hat{Z}_{jt-1} \cdot \xi + \eta_{ijt} \quad (5)$$

$$X_{jt-1} = \left[ frn\_NrmDeg_{jt-1} \quad frn\_NrmBP_{jt-1} \quad \log NumFrn\_outside_{jt-1} \right]$$

## 5. Results

### 5.1 Parameter Estimates of the Baseline Model

Parameter estimates for the baseline model are summarized in Table 6. The baseline model finds that local friendship networks (*Frn\_NrmDegree*) and non-local friendship networks (*logNumFrn\_Outside*) have a significant but negative impact on diffusion. The influence from subscriber networks is mixed. The degree to which connections are initiated into a channel (*Subs\_NrmInDegree*) is positive but insignificant, while the connections initiated by a channel (*Subs\_NrmOutDegree*) are significantly negative. The rating of a video and the number of external links has a strong positive impact on the diffusion process. The adjusted R-Squared values are not very high, which is unsurprising, given the diversity of factors that can influence the diffusion of digital content in online settings. We have therefore reported coefficients as significant only if they are significant at the 5% level or better.

The baseline model does not account for a number of econometric challenges. First, the baseline model does not examine serial correlation, since we assume that the difference in the number of views between time periods is a function of the total popularity of a video. However, it is possible that there are unobservable exogenous events that trigger a surge of popularity, which also becomes the driving force of the diffusion, i.e.,  $\Delta v_{ijt}$  is influenced by  $\Delta v_{ijt-1}$ . Second, the baseline model does not take into account the interaction between variables that may impact the diffusion process (e.g., Wejnert 2002) and the nature of group formation due to user preferences. We therefore consider a structural model with interaction terms, which we discuss next.

**Table 6.** Parameter Estimates – Baseline model  
(standard error in parenthesis)

Variable	Parameter Estimates	
	Benchmark	
Intercept	0.2876	(0.003544)***
LogNumOfViews	0.001317	(0.000378)***
(LogNumOfViews) <sup>2</sup>	0.000625	(0.000094)***
Log(vAge)	-0.05003	(0.000596)***
Rating	0.001981	(0.000295)***
logNumOfLinks	0.002463	(0.000861)***
Frn_NrmDegree	-0.00080	(0.000203)**
Frn_NrmBP	-0.00003	(9.959E-6)**
Subs_NrmOutDegree	-0.00308	(0.000788)**
Subs_NrmInDegree	0.000182	(0.001749)
logNumFrn_Outside	-0.00063	(0.000261)**
Fit Statistics		
Adjusted R Squared	0.17	

\*\*\*  $p = 0.01$ ; \*\*  $p = 0.05$

## 5.2 Structural Model Parameter Estimates

The results from the structural model are shown in Table 6. In the early stages of the life of a video, the influence from the incoming subscribers is the most important factor while friend networks appear to play an important part in later adoption. In both the baseline model and the structural model the video rating and the number of external links have a significant impact on video diffusion, indicating that, although the peer group of friends' network helps a video to reach a critical mass in the early stage, the perceived video quality and strategic linkages with other online channels of influence such as blogs positively influence diffusion. The R-Squared values are greater for the baseline model, probably due to serial correlation, which is removed in the structural model.

**Table 6.** Parameter Estimates – Structural Model  
(standard error in parenthesis)

Variable	Parameter Estimates			
	Structural w/o Interaction		Structural w/ Interaction	
Intercept	0.2317	(0.003536)***	0.2525	(0.006278)***
LogNumOfViews	0.001963	(0.000348)***	0.002068	(0.000350)***
(LogNumOfViews) <sup>2</sup>	-0.00014	(0.000085)	-0.00015	(0.000085)*
Log(vAge)	-0.03806	(0.000590)***	-0.04196	(0.001165)***
Rating	0.000766	(0.000274)***	0.000773	(0.000274)***
logNumOfLinks	0.004641	(0.000791)***	0.004759	(0.000791)***
Frn_NrmDegree	-0.00069	(0.000174)***	-0.00439	(0.001651)***
Frn_NrmBP	-0.00001	(9.107E-6)	0.000172	(0.000075)**
Subs_NrmOutDegree	-0.00209	(0.001795)	0.09068	(0.01258)***
Subs_NrmInDegree	0.01514	(0.002672)***	0.04580	(0.01971)**
logNumFrn_Outside	-0.00158	(0.000293)***	-0.01317	(0.002174)***
logVAge*Frn_NrmDegree			0.000700	(0.000312)**
logVAge*Frn_NrmBP			-0.00004	(0.000014)**
logVAge *logNumFrn_Outside			0.002126	(0.000402)***
logVAge *logSubs_NrmOutDegree			-0.01751	(0.002380)***
logVAge *log Subs_NrmInDegree			-0.00532	(0.003719)
<b>Fit Statistics</b>				
R Squared	0.072		0.074	

\*\*\*  $p = 0.01$ ; \*\*  $p = 0.05$

## 5.3 Discussion

We find that friendship networks have a significant impact, indicating support for H1, and particularly in the later stages of a video diffusion. The interaction term between degree centrality of the friend network with time is positive, indicating support for H2, and indicating that conformity preferences

play a very important role in the later stages of the diffusion process. We also find that actors that are more connected to individuals outside the local network (i.e., have more friends from outside the group) have significant influence on the diffusion of content, validating H3. This is due to the fact that such individuals have access to new and non-redundant content. Central actors that can enrich the set of experiences and ideas through contact with non-local actors are more likely to disseminate new content within their peer group. The interaction term between the friends from outside the network with time also has a significant impact on diffusion, validating H4. The magnitude of the impact of non-local ties is three times stronger than that of local ties alone, suggesting that social learning may play an important role in content diffusion. Since we find that the number of friends from outside the network has a stronger impact than the degree centrality of the actor in a friend network, the local network effects appear to be less influential compared to the impact of access to non-redundant information from individuals outside a local network. We also find that the impact of non-local ties and within-group friendship networks affect different stages of the diffusion process. The centrality of users in subscriber networks is highly significant, and plays a dominant role in the earlier stages of diffusion, indicating support for H5 and H6. Early adopters, by virtue of their positions of influence in the subscribers network, are pivotal in persuading others in order that a video acquires a critical mass of viewers.

Figure 6 depicts the process of diffusion with and without taking into account the social structure of interactions on YouTube. 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. One, the dynamics whereby the number of adopters reaches a critical mass is likely to be different when we factor in the role of early adopters in persuading proximate actors in a social network. The initial diffusion rate is very sensitive to the number of incoming subscriptions of a channel, which implies that the rate at which a video diffuses through the population depends on actors' willingness to sample new videos. A video acquires momentum and spreads through the population as a function of taste preferences of users in the initial stages. Second, since we find that the role of conformity preferences and social learning is greater in the later stages of the diffusion process,

diffusion through friendship networks occurs only after some proportion of the initial population of users is *infected*, or when a certain proportion has already viewed a video, i.e., when a social threshold level of adoption is reached. Interestingly, we find that the combined effect of friendship networks and subscriber networks reduces the time for the diffusion curve to reach the familiar S-shape. While the effects of social influence from friendship networks is not strong in the early stages of the life of a video, it is likely that adoption by a small number of highly influential nodes (e.g., Dodds and Watts 2007) in the subscriber networks sets off a trajectory of adoption amongst the friend networks, creating a social multiplier effect that magnifies the impact of user tastes through social preferences such as conformity. The contrasting effects of degree centrality and the Bonacich power measures in a friend network also suggest that conformity preferences and social influence from individuals that are perceived to be of a higher status may have subtle differences.

**Figure 6.** Diffusion of Number of Views Without vs. With Social Network Effects

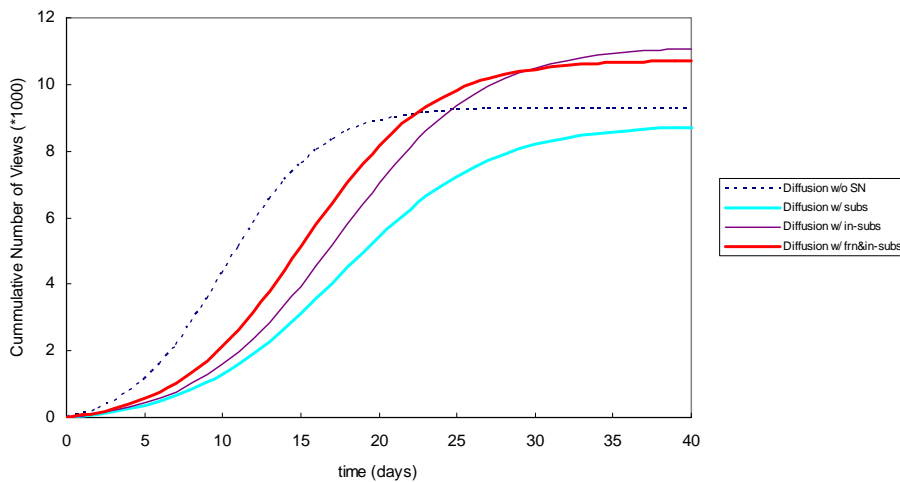
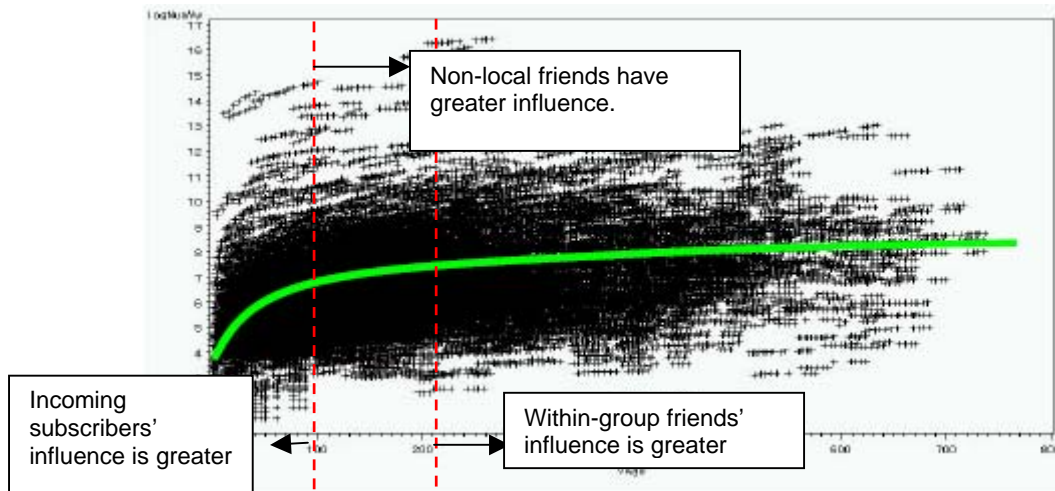


Figure 7 depicts the number of views of a video over time. Prior literature in marketing suggests it is difficult to distinguish between social contagion and the heterogeneity in users, which highlights the difficulty in identifying with precision the factors that influence the structure of consumer demand. Given the multiplicity of factors influencing an individual, we distinguish between factors that affect membership in a social network from other types of social influence and therefore address the reflection problem in inferring social influence. Some of the prior literature on diffusion suggests that what drives the diffusion process before a critical mass is reached may be very different from what drives the

diffusion process after a critical mass is reached. We similarly find differing impacts from the various types of social interactions. We observe that conformity preferences as well as the heterogeneity of users' tastes influence the diffusion process, but affecting different stages in the diffusion of user-generated content. We also rule out the possibility that an individual's viewing pattern either is solely influenced the social group that they belong to or that the viewing patterns are a result of heterogeneity of users. What makes the results interesting is that they demonstrate not only that social influence matters, but that we can identify the variety of social interactions that contribute to diffusion of new products.

**Figure 7.** Video Popularity over Video Age



The results also offer a contrast between the different types of information transmission in different types of social interactions. We find that the influence from the subscriber network is markedly different from that of non-local friendship ties. So see why this is significant, consider that if users were to “learn” without regard to identity, we should observe that the impact of friend networks and subscriber networks is similar throughout the diffusion process. However, theory suggests that social learning by users is sensitive to identity, and users place different weights on information acquired from other users (e.g., Bala and Goyal 1998). The fact that we find that the impact of non-local ties is stronger in the process of diffusion than that of within-group ties also supports our interpretation of social learning. Social learning describes a richer motive for behavior than a simple taste for conformity with others. A pure preference for conformity implies frequent and sustained connections with others in a local network leading to a strong sense of identification, which also has the effect of limiting the amount of non-redundant

information that can be transmitted across the local network. By contrast, social learning plays a role when user faces a plethora of choices and needs to make decisions about adopting uncertain content. The social learning from non-local friends also supports Granovetter's (1973) arguments about the "strength of weak ties". Interaction with friends with similar tastes in the local network promotes homophily and furthers a sense of identification; however, while such sources of influence have a greater ability to persuade a user they also provide fewer sources of external information. Individuals in the non-local network facilitate observational or social learning as well as expose a node to a greater variety of content. The aggregate dynamics of the content diffusion model studied in this paper present substantial evidence for social learning effects. Thus, 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 of content.

Our study makes a number of theoretical and empirical contributions to the literature. The Internet has been instrumental in promoting online dissemination of product reviews and buyer feedback, with tremendous implications for customer-facing activities such as customer acquisition and retention. The research on online word of mouth (e.g., Chavalier and Mayzlin 2006, Dellarocas 2003) has examined the impact of user-generated social information on Internet commerce. Our study can add to this stream of literature by identifying the networked structure of social interactions and interpersonal influence. Forman et al. (2008) find that users in an online social community provide self-descriptive information that reveals their *social identity*, and such identity disclosure plays a significant role in product sales. In a similar manner, networks of interpersonal influence might promote users' identification with a particular channel, influencing content diffusion (similar to users' identification with brand of a company). The results in this study suggest that the transmission of online word of mouth effects should be analyzed taking into account the networked structure of interpersonal interactions. The social capital fostered through networked interactions may also mitigate the potential for information asymmetry in online markets, suggesting a social network based explanation for reputation systems on the Internet (e.g., Resnick et al. 2000).

Prior research on the technology acceptance model (TAM) has posited that acceptance of



technological innovations is driven by the ease of use and usefulness of an innovation, and the TAM model has been extended to consider the role of social influence (Venkatesh and Davis 2000). A perspective based on social networks can enrich this stream of research by highlighting the process by which individuals turn to each other for social cues and the role of social networks in structuring interpersonal influence that shapes perception and behavior. Social multiplier effects arising from social contagion due to interpersonal influence within a social network can be instrumental in shaping perceptions of the usefulness of innovations (e.g., Bandiera and Rasul 2006). Our study suggests that the role of social influence strengthened through conformity preferences and social learning mechanisms play an important role in the diffusion process, including technology acceptance.

#### **5.4 Managerial Implications**

Parameswaran and Whinston (2007) posit that YouTube has “evolved into a pop culture medium that drives rapid dissemination of popular videos worldwide”. Similarly, the potential role of YouTube as a marketing tool is enormous. Several media companies are interested in YouTube as a channel to reach viewers. The question then becomes, should the company sponsor a set of content creators to post their content on YouTube or should they fund a set of video that has already acquired momentum? Companies need to consider how to strategically place videos in order to reach more users. Similarly, advertisers should consider the role of incorporating user-generated content as a channel of advertising. It may be especially valuable to use YouTube as a channel to incorporate user-generated content into an advertising campaign that can then be extended outside YouTube.

The ease with which users can participate in online social networks and seek out digital content such as music and entertainment from a variety of social media is of tremendous interest to marketers and content creators. We find that the number of outgoing subscriptions has a channel 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. This has a number of implications for practitioners. A push model of content creation, where the subscriber tries to push content by initiating connections to several other users, may be less successful than a pull model where the subscriber is highly popular and receives several connections from others. Subscribers with high in-degree centrality can act

as fashion leaders or opinion makers in the diffusion of content. Therefore, another implication for practitioners is to consider the role of the prestige of a subscriber in the diffusion of content. The difference between different mechanisms of social influence, conformity preferences or social learning fostered through friendship ties as compared to networks defined by user tastes, is of immense importance to content creators and to media companies that are interested in YouTube as a channel to reach viewers.

Fundamentally, from a longer-term perspective, the success of a single video posted by a user is not enough to ensure repeat viewing of the content of a channel. In other words, it is the brand recognition of the user's channel that is important. Content creators on YouTube need to work towards loyalty towards the channel, suggesting that user recognition and brand building of channels in user-generated content communities such as YouTube works in a manner analogous to brand-building and marketing efforts by companies. Such brand recognition can be measured by the incoming subscribers, i.e., the incoming traffic to a channel based on channel recognition.

## **5.5 Limitations**

While we capture the network structure changing over time, we do not model how this process occurs. The promotional effects by content creators such as major record labels, which may affect the actions of early adopters and thus trigger the process of diffusion, is unobservable in our estimation. We also do not consider the interaction between offline and online social networks, and in particular, whether social influence through offline interactions bolsters the impact of conformity pressures and status in networked interactions online. We also do not consider whether conformity preferences in groups are enhanced through cohesive social network structures that foster a sense of social identity.

We also do not consider the interaction between network structure and decision-making in groups that is likely to affect a group's susceptibility to diffusion. Bala and Goyal (1993) 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. Agents' decisions could depend upon not only on how many neighbors an agent has and the actions the agents' neighbors choose but also by agent-specific factors.

While we consider individual heterogeneity in the process of diffusion and control for time-invariant heterogeneity, we do not investigate whether such heterogeneity results in different agents updating their beliefs differently. It may be necessary to conduct experimental studies or simulation to tease out the types of social learning and expertise of the influencers.

## **6. Conclusions and Future research**

Our study captures the process by which different mechanisms of social influence affect the trajectory of diffusion in an online social network setting. Our estimation is robust to unobservable self-selection of users and unobserved heterogeneity in user preferences that allows us to identify social influence in a network structure. Our estimation is also robust to contemporaneous shocks that may result in serial correlation. Given the multiplicity of factors that influence the diffusion of content on online social networks, our study makes a first step in disentangling the different means by which social interactions are structured and identifying social influence.

In an increasingly hyper-networked age, individuals have access to informational content from a wide variety of on-line sources such as blogs, consumer forums, podcasts and social media that influence their tastes and preferences. Individuals whose networks bridge across 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 may 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. Future work can also investigate whether pre-existing social linkages affect a user's recognition of and the consequent influence from channels of subscribers. The interaction between interpersonal influence structured through social networks and communication efforts by influentials driving the diffusion process is another area of inquiry.

## Appendix: Social network measures and definitions

In a social network context, adjacency denotes that two agents, represented by nodes, are directly connected with one another (Robinson and Foulds 1980). An adjacency matrix is a representation of the adjacency relationships of the actors in a social network used to calculate the social network measures.

### A1. Degree Centrality Measures for Non-Directional Relations

We consider a unipartite graph where  $x_{ij} = 1$  denotes that a link exists between nodes  $i$  and  $j$ . By definition, a node does not link to itself.

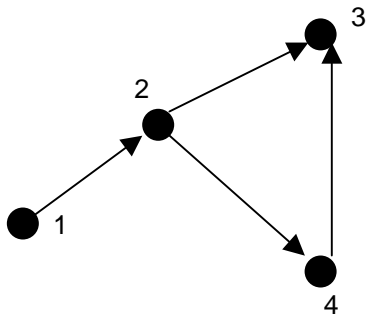
$d(n_i)$  is the degree of the node;  $C_D(n_i) = d(n_i) = \sum_j x_{ij} = \sum_j x_{ji}$  is an actor-level degree centrality index.

For a group of size  $g$ , the maximum value of the actor level centrality measure is  $(g - 1)$ . Standardizing for group size, the standardized measure of degree centrality is the proportion of nodes adjacent to  $n_j$ :

$$C'_D(n_i) = \frac{d(n_i)}{g-1}, C'_D(n_i) \in (0,1) \text{ and varies between } \frac{1}{g-1} \text{ and } 1 \text{ (e.g., Freeman 1978).}$$

### A2. Degree Centrality Measures for Directed Relations

As an illustrative example, consider the graph below



A node is connected when there is at least one arc or set of arcs that relate the actor with another actor (Wasserman and Faust 1994). Each arc from node  $i$  to node  $j$  is denoted by  $l_k$ .  $x_{ij} = 1$  when the directed path  $n_i \rightarrow n_j$  is present. The adjacency matrix (for graphs of one path length) for the above graph then is:

		Node			
		1	2	3	4
Node	1	0	1	0	0
	2	0	0	1	1
	3	0	0	0	0
	4	0	0	1	0

We define out-degree centrality as the row sum for the node in a dichotomous matrix. Thus, out-degree of actor  $i$  is  $x_{+i} = \sum_j x_{ij}$ . The column sum (for a node) in a dichotomous matrix is the in-degree centrality of the node. In-degree of actor  $j$  is  $x_{j+} = \sum_i x_{ij}$ . We normalize the above measures by  $(g-1)$  to remove scale effects.

### **A3. Bonacich Power**

For a matrix  $X$  of relationships, the Bonacich power is defined as  $c_i = \sum_j (\alpha + \beta c_j) X_{ij}$  where  $\alpha$  is a scale parameter defined so that  $\sum_i c_i (\alpha, \beta)^2$  equals the number of units in the network (Bonacich 1987).  $\beta$  can take both positive and negative values: positive values imply that a node is more powerful when it is connected to others that are more powerful, while a negative value implies that a node is more powerful as its connections become weaker.

### **References for calculating measures:**

- Bonacich, P. (1987), "Power and centrality: A family of measures," *American Journal of Sociology* 92(5): 1170-1182
- Freeman, L. C. (1978) Centrality in social networks conceptual clarification. *Social Networks*, 1, 215.
- Robinson, D. F. and Foulds, L.R. (1980) *Digraphs: Theory and Techniques*. Gordon and Breach

## References

- Angrist J., A. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economy* 106(4): 979-1014.
- Armstrong, and V. Sambamurthy, 1999. Information Technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Information System Research*, 10, 304-327.
- Bala, V., S. Goyal. 1998. Learning from Neighbors. *Review of Economic Studies* (65:3) 595-621.
- Banerjee, A.V. 1992. A simple model of herd behavior. *Quarterly Journal of Economics* 107(3): 797-817.
- Bandiera, O., I. Rasul. 2006. Social Networks and Technology Adoption in Northern Mozambique. *Economic Journal* 116: 862-902.
- Bass, F. 1969. A new product growth for model consumer durables. *Management Science* 15(5): 215-27.
- Barnes, B. 2007. Can CBS Put the Net Into Network? --- Broadcaster Launches Plan Syndicating Shows on Web, Admits Old Strategy Failed. *The Wall Street Journal*, 14 May, 2007.
- Bemmaor, A.C., J.C. Lee. 2002. The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. *Marketing Science* 21(2): 209–220.
- Bernheim, D.A. 1994. A Theory of Conformity. *Journal of Political Economy* 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 Science* Vol. 53, No. 9, pp. 1359-1374.
- Bonacich 1987. Power and centrality: A family of measures *American J. of Sociology* 92(5): 1170-82
- Borgatti, S. P., Everett, M.G., L.C. Freeman. 2002. *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Brynjolfsson, E., M.D. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46(4) 563-585.
- Boulding, W., M. Christen. 2003. Sustainable Pioneering Advantage? Profit Implications of Market Entry Order. *Marketing Science* (22:3) pp. 371-392.
- Burt, R. 1987. Social Contagion and Innovation: Cohesion Versus Structural Equivalence. *The American Journal of Sociology* (92:6) pp. 1287-1335.
- \_\_\_\_\_. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard University Press
- Chevalier, J., D. Mayzlin. 2006. The Effect of Word of Mouth Online: Online Book Reviews. *Journal of Marketing Research*.
- Coleman, J.S., E. Katz, H. Menzel. 1966. *Medical Innovation: A Diffusion Study*. New York.
- Dellarocas, C. 2003. The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Systems. *Management Science* 49 (10), 1407-1424.
- Dewan, S. and J. Ramaprasad, 2008, "Consumer Blogging and Music Sampling," Working Paper, University of California at Irvine
- Dodson, Jr., J.A., E. Muller. 1978. Models of New Product Diffusion through Advertising and Word-of-Mouth. *Management Science* (24:15) 1568-1578.
- Dodds, P.S., D.J. Watts, C.F. Sabel. 2003. Information Exchange and Robustness in Organizational Networks. *Proceedings of the National Academy of Sciences*, 100 (21): 12516-12521
- Dodds, P.S., D.J. Watts. 2007. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*. Forthcoming.
- Ellison G., D. Fudenberg. 1993. Rules of Thumb for Social Learning. *Journal of Political Economy*, 101(4): 612-43
- Fichman, R.G. 2000. The Diffusion and Assimilation of Information Technology Innovations, in R.W. Zmud (Ed.) *Framing the Domains of IT Management: Projecting the Future Through the Past*, Cincinnati, OH: Pinnaflex Educational Resources, Inc.
- Forman, C., A. Ghose, B. Wiesenfeld. 2008. Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Information. *Information Systems Research*, September, 19(3).
- Freeman, L.C. 1979. Centrality in social networks: A Conceptual clarification. *Social Networks* 1:215-39
- Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology*, 6: 1360-1380.
- \_\_\_\_\_. 1985. Economic Action and Social Structure: The Problem of Embeddedness. *American Journal*

- of Sociology*, 91(3), 481–511.
- Hausman, J.A., W.E. Taylor. 1981. Panel Data and Unobservable Individual Effect. *Econometrica* 49, 1377-98.
- Hodgson, J. 2007. Yahoo, Bebo Agree To Work Together On Advertising. *Fortune Magazine Online*.
- Ibarra, H., S.B. Andrews. 1993. Power, social Influence, and sense Making: effects of network centrality and proximity on employee perceptions. *Administrative Science Quarterly* 38(2): 277-303
- Kang, S. 2007. MySpace Signs Partners To Expand Video Offerings. *The Wall Street Journal*, 16 May.
- Kraut, RE., Rice, RE., Cool, C. and Fish, RS. "Varieties of social influence: The role of utility and norms in the success of a new communication medium," *Organization Science* (9:4), 1998, 437-453.
- Mahajan, V., E. Muller, F.M. Bass. 1990. New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing* (54:1) 1-26.
- Manski, C.F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60(3) 531-542.
- Nelson, P. 1970. Information and Consumer Behavior. *Journal of Political Economy* 78(2) 311-329
- Newman, M.E.J. 2003. The structure and function of complex networks. *SIAM Review* 45, 167-256
- Parameswaran, M., A.B. Whinston. 2007. Social Computing: An Overview. *Communications of the AIS* (19) 762-780.
- Rogers, E.M., L.D. Kincaid. 1981. *Communication Networks*. New York: Free Press.
- Raymond, E. 2001. *The Cathedral and the Bazaar*. O'Reilly
- P. Resnick, R. Zeckhauser, E. Friedman, K. Kuwabara. 2000. Reputation systems. *Communications of the ACM*, 43(12):45–48.
- Rogers, E. 1995. *Diffusion of Innovations* (4th ed.). New York: The Free Press.
- Sacerdote, B. 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics* (116:2) pp. 681-704.
- Salancik, G., J. Pfeffer. 1978. A social information processing approach to job attitudes and task design. *Administrative Science Quarterly*, 23: 224-253.
- Salganick, M., Dodds, P.S., D. Watts. 2006. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science* (311) pp. 854-856.
- Salganik, M.J., D.D. Heckathorn. 2004. Sampling and estimation in hidden populations using respondent-drive sampling. *Sociological Methodology* 34:193-239
- Strang, D., N.B. Tuma. Spatial and Temporal Heterogeneity in Diffusion. *American Journal of Sociology* (99:3), 1993, pp. 614-639.
- Surowiecki, J. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* Doubleday
- Talukdar, D., K. Sudhir, A. Ainslie. 2002. Investigating New Product Diffusion Across Products and Countries. *Marketing Science* (21:1) 97-114.
- Van den Bulte, C., G. Lilien. 2001. Medical Innovation Revisited: Social Contagion versus Marketing Effort," *American Journal of Sociology* (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 Science* 23.
- Venkatesh, V., F.D. Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science* (46:2), 186-204.
- Watts, D.J., P.S. Dodds, and M.E.J. Newman, 2002, "Identity and search in social networks," *Science*, 296(5571):1302-1305
- Wasserman, S., K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.
- Wejnert, B. 2002. Integrating Models of Diffusion of Innovations: A Conceptual Framework. *Annual Review of Sociology* (28) 297-326.
- Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.