12-2013

Two Formulas for Success in Social Media: Social Learning and Network Effects

Liangfei QIU
University of Texas at Austin

Qian TANG
Singapore Management University, QIANTANG@smu.edu.sg

Andrew B. Whinston
University of Texas at Austin

Two Formulas for Success in Social Media:

Social Learning and Network Effects

Liangfei Qiu

Department of Economics, The University of Texas at Austin

liangfeiqiu@utexas.edu

Qian Tang

School of Information Systems, Singapore Management University

qian.tang@smu.edu.sg

Andrew B. Whinston

McCombs School of Business, The University of Texas at Austin

abw@uts.cc.utexas.edu

November 19, 2013

Abstract

This paper examines social learning and network effects that are particularly important for online videos, considering the limited marketing campaigns of user-generated content. Rather than combining both social learning and network effects under the umbrella of social contagion or peer influence, we develop a theoretical model and empirically identify social learning and network effects separately. Using a unique data set from YouTube, we find that both mechanisms have statistically and economically significant effects on video views, and which mechanism dominates depends on the specific video type.

Keywords: Social Learning, Network Effects, User-Generated Content, Social Contagion, Social Media

"You've got to create images they won't accept. Make them foam at the mouth. Force them to understand that they're living in a pretty queer world."

— Andre Malraux, *"Picasso Mask"* (1976), page 110.

"A wealth of information creates a poverty of attention."

— Herbert A. Simon, (1982), page 40.

1. INTRODUCTION

With new products such as consumer goods, food, pharmaceutical goods, financial services, and movies constantly flooding the markets, consumers face an already overwhelmingly large and rapidly growing choice set. Meanwhile, with the prolific use of social media, consumers obtain information about products from social sources in the forms of product reviews and friends' recommendations. Therefore, marketers tend to use these forms of social contagion to influence consumers' perception and behavior.

Several studies have shown the presence of social contagion in new product adoption (e.g., Godes and Mayzlin 2009; Goldenberg et al. 2009; Iyengar, Van den Bulte, and Valente 2011; Katona, Zubcsek, and Sarvary 2011). We aim to take a step further to test different mechanisms of social contagion: How are individuals' decisions affected by their peers? In our context, social contagion happens mainly through two channels: (1) Social learning, the process in which consumers obtain knowledge about a product's quality through peers (Tucker and Zhang 2011); (2) Network effect, the phenomenon that the value of a product increases as the number of its users increases (Katz and Shapiro 1985). Which mechanism exists or dominates depends on the specific product in question. When choosing a mobile network operator, network effects may dominate because of free mobile-to-mobile calling. When purchasing an HDTV, social learning becomes the primary force because consumers are mainly concerned with quality.

The two mechanisms have different implications for marketing strategies: For products with strong network effects, creating a large user base is crucial in attracting new adopters. In contrast, generating positive word-of-mouth (WOM) is the key for products with prevailing social learning.

In this study, we differentiate between social learning and network effects for social media content consumption, especially in the context of YouTube, the largest online video sharing website. Selecting online videos to watch is one of the most common choices viewers make every day. According to ComScore, the average user spend about 43 minutes watching online videos in June 2013, and Google websites (primarily YouTube) account for approximately 40% of that time, about 17 minutes.[1] According to YouTube statistics, 100 hours of video are uploaded to YouTube every minute. What these numbers mean is that, given the vast reservoir of online videos, choosing videos to watch can become a complicated issue. On the one hand, consumers receive various information from friends and infer video quality through social learning. On the other hand, frequent social sharing creates direct or indirect network effects where a video becomes a fad. For example, when a video goes viral, users have strong incentives to watch it so they have something to discuss in social encounters.

Figure 1 shows the conceptual framework of our study. Most existing studies on social contagion focus on the Manski problem (Manski 1993): distinguishing general social contagion from homophily — the tendency of individuals to associate with similar others (Aral et al. 2009, Aral and Walker 2011, Bapna and Umyarov 2012). Few of them differentiate between the two mechanisms of social contagion: social learning and network effects. Social learning affects consumers through the quality information conveyed by peers, whereas the network effects

[1] See http://www.comscore.com/insights/Press_Releases/2013/7/comScore_Releases_June_2013_US_Online_Video_Rankings.

influence consumers according to the size of the user base. Although these mechanisms lead to similar empirical outcome, their implications are vastly different. If social contagion is generated mainly by network effects, then seeding strategies, which determine the initial set of targeted consumers, would by implication have a strong influence on the success of viral marketing. A firm can amplify social contagion and accelerate product purchases by offering introductory discounts (Ho et al. 2012). If social learning is the dominant effect, however, seeding would not be effective unless the initial consumers generate positive word-of-mouth. Consumers can infer that the high demand of their peers is caused by the introductory discount rather than the high product quality (Qiu and Whinston 2012). Both cases are theoretically plausible and need to be empirically distinguished.

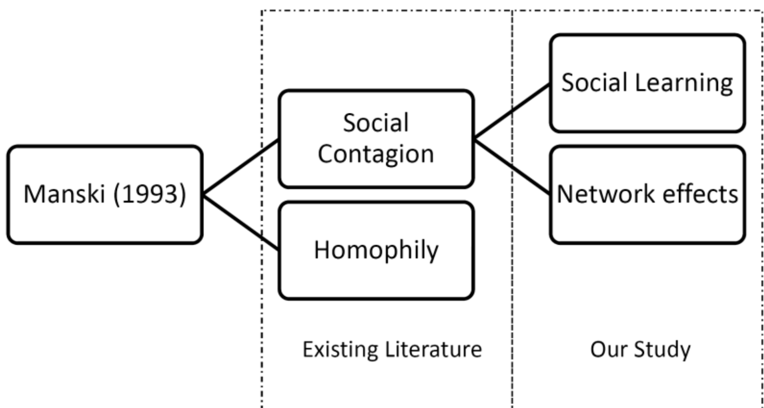

Figure 1. The Conceptual Framework of Social Learning and Network Effects

To the best of our knowledge, our paper is the first to disentangle social learning and network effects in the context of online video sharing. Because of the lack of pre-release marketing effort, these two types of social contagion are particularly important for user-generated content (UGC). Our empirical results suggest that both mechanisms affect the diffusion of social media content significantly, with social learning influencing high-quality content more and network effects influencing attention-grabbing content more. We find that social learning is more pronounced when consumers are less certain about video quality and for videos belonging to unpopular categories. Eliaz and Spiegler (2011) examine the incentives of a firm to offer low quality products as “pure attention grabbers.” Our study confirms their theoretical results: attention grabbers provoke discussions and go viral through network effects. The implications derived from studying YouTube can carry over to other consumer choice problems as well.

The rest of the paper is organized as follows: We review related literature in Section 2. In Section 3, we outline an analytical model that motivates our empirical hypotheses. Section 4 describes the YouTube data. In Section 5, we present the identification strategy and empirical results. Some applications are explored in Section 6. Section 7 concludes the paper.

2. Literature Review

Manski (1993) discusses an econometric challenge of identifying social contagion: Is a person’s behavior caused by his social reference group, or does it simply reflect the same movement in his reference group? The observation that individuals belonging to the same group tend to behave similarly might result from social contagion, exogenous contextual effects, or homophily.[2] Failure to account for contextual effects or homophily might lead to an overestimation of the effect of social contagion.

These confounding effects are difficult to distinguish, and the identification of social contagion often requires strong parametric assumption or rich data collection. Aral et al. (2009) distinguish influence-based contagion from homophily-driven diffusion using a dynamic matched sample of global instant messaging users. Iyengar et al. (2011) distinguish social contagion from homophily and exogenous contextual effects in prescribing behavior among

[2] Among these three effects, only social contagion can generate “social multiplier” with a positive feedback (Manski 1993).

networks of doctors.

Within the framework of social contagion, studies have been focusing on distinguishing social learning from other contagion mechanisms such as saliency effect (i.e., observed choices are more salient than alternative choices), and conformity concerns (i.e., the social pressure to adopt the choice made by the majority). Cai et al. (2009) use a field experiment to show that observational learning, rather than saliency effect, affects customers' choices. Van den Bulte and Stremersch (2004) study different social contagion mechanisms using a meta-analysis of publications on new product diffusion and find evidence for status concerns and social-normative pressures but not for social learning under uncertainty. Iyengar et al. (2012) differentiate between social learning and normative influence in the adoption of a new drug, and find both operate as simultaneous yet separate mechanisms for social contagion. Shi and Whinston (2013) study observational learning in the context of location-based networks. Moretti (2011) shows that social learning is a more important determinant of sales in the movie industry than network effects. Our empirical results show that both social learning and network effects are important in the context of online videos. This result highlights the difference between movies and online videos. Consumers have more precise prior information to estimate the quality of a movie, but much less information for social media videos considering the limited marketing campaigns of user-generated content. Therefore, it is unlikely that a movie can go viral solely because of network effects, but social media videos can. We also depart from the literature by further exploring the role of video types on social learning and network effects. Comparing our results with those in Morreti (2011), we find that, in social media, attention grabbing videos go viral using network effects without high-quality content. The fact that high-quality videos focus on social learning and have become more similar to movies is consistent with the trend that social media are imposing competitive pressure on professional videos. From a methodological perspective, we apply a two-step bootstrapping method to correct possible invalid inferences in a number of applied studies, such as Barro (1977) and Moretti (2011).

Online WOM, especially online user reviews, has become an important channel of social contagion (Chevalier and Mayzlin 2006; Duan et al. 2009; Goldenberg et al. 2010). In the context of UGC, Susarla et al. (2012) demonstrate that social networks affect economic outcomes by structuring the information available to other users, which then influences their decisions, perceptions, and behaviors. Goldenberg et al. (2012) show that the stream of people's chatter from social broadcasting networks facilitates social learning among a much broader peer group than has traditionally been possible. Although these studies provide evidence of the existence of social contagion in the diffusion of UGC, none of them look into the two specific contagion mechanisms, social learning and network effects, each of which may have different managerial implications.

For UGC, understanding whether the popularity of the content makes it valuable (network effects) or the value of the content makes it popular (social learning) is pivotal. By distinguishing between network effects and social learning, our paper contributes to the understanding of different social contagion mechanisms in the diffusion of UGC.

3. A Theoretical Framework of Social Learning and Network Effects on YouTube

A video can go viral either because of social learning or network effects. This section examines the different implications of these two mechanisms. We set up two simple analytical models incorporating social learning and network effects separately. Testable hypotheses are developed accordingly based on the predictions of the two models.

3.1 A Model of Social Learning on YouTube

YouTube videos are experience goods whose quality cannot be fully observed by consumers ex ante but can be ascertained upon consumption. Therefore, before consumption, consumers are never completely sure about the quality, but they can always acquire useful information from friends who have already watched the videos. The literature on observational learning (Banerjee 1992) examines the social learning that occurs through observing other people's behaviors. We also consider underlying social networks from which people make inferences about the quality of a video based on the information within their social connections. We capture the learning process with a Bayesian learning model, where each consumer receives feedback from peers and updates the prior belief of the video quality.[3]

Our theoretical model of social learning is based upon Bayesian learning models (Acemoglu et al. 2011), but extending them in one important aspect: we model the probability that a consumer watches a video is the product of the probability that he is aware of the video and the probability of watching the video conditional on being aware of it (Shi, Rui, and Whinston 2013). It's important to consider the awareness probability explicitly in social media. Compared to movie sales studied in Moretti (2011), social media consumption is severely restrained by how well consumers know the existence of a video. Because consumers on YouTube face an almost infinite choice set, it would be unrealistic to assume that they know the existence of each video.[4]

The decision process for YouTube video consumption over time is described in Figure 2. Although Figure 2 depicts only the sequence of events at time 1 and 2, the process proceeds in a

[3] Following Banerjee (1992), the timing of consumption is exogenously given, and we do not consider the strategically behavior of delaying the decision making process to obtain more feedback.

[4] As of August 2012, on average, about 72 hours of video are uploaded to YouTube every minute, and the number is still growing, see http://www.youtube.com/t/press_statistics.

similar manner at time 3, 4, …, T.

We first describe the decision process, conditional on that consumers are aware of the existence of the video. The utility that a representative consumer i obtains from watching some YouTube video j is

$$u_{ij} = V_j + \eta_{ij}, \eta_{ij} \sim N(0,1/\rho_\eta),$$

where V_j is the latent quality of the video, and η_{ij} represents the unobserved taste heterogeneity. At time 1, video j is posted on YouTube. Since there is no specific prior information pertaining to the video before it is posted, consumers share a common prior on the quality of the video, given by

$$V_j \sim N\left(X_j'\beta, 1/\rho_{V_j}\right),$$

where X_j is a vector of the observable characteristics of video j before watching. $X_j'\beta$ is the ex-ante expectation of quality, and ρ_{V_j} is the precision of prior for video j.

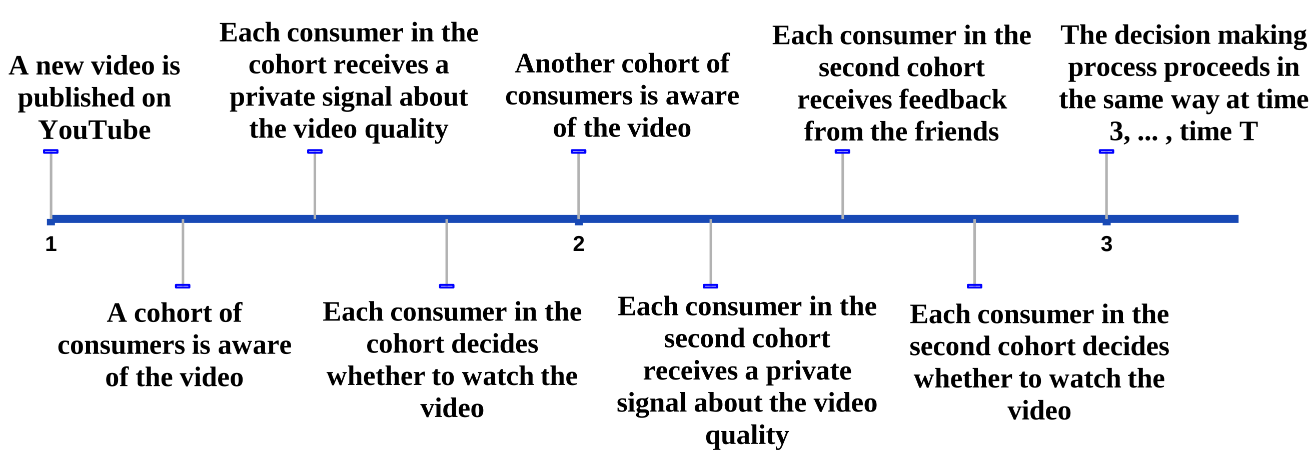

Figure 2. Timeline

Before making a decision, each consumer observes a conditionally independent private signal of the quality:

$$S_{ij} = V_j + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0,1/\rho_{\varepsilon_j}), \tag{1}$$

where ρ_{ε_j} is the precision of consumer i's information source for video j. The signal errors ε_{ij} are independent across consumers. Consumers update the prior according to Bayes' rule. $E_t[u_{ij}|I_t]$ represents consumer i's expected utility of video j at time t given the information set at time t, I_t. Notice that the video is newly published, so no social learning occurs at time 1. Conditional on being aware of video j, a consumer chooses to watch it if the ex-ante expected utility is no less than the opportunity cost of watching video j, c_{ij}. Therefore, if a consumer is aware of video j at time 1, then he watches video j if

$$E_1[u_{ij}|I_1] = \frac{\rho_{V_j}}{\rho_{V_j} + \rho_{\varepsilon_j}} X_j'\beta + \frac{\rho_{\varepsilon_j}}{\rho_{V_j} + \rho_{\varepsilon_j}} S_{ij} + \eta_{ij} \geq c_{ij}.$$

Accordingly, the probability that a consumer watches video j conditional on being aware of it at time 1 is:

$$\Pi_1 = \Pr\left(E_1[u_{ij}|I_1] \geq c_{ij}\right) = \Phi\left(\frac{\frac{\rho_{V_j}}{\rho_{V_j}+\rho_{\varepsilon_j}} X_j'\beta + \frac{\rho_{\varepsilon_j}}{\rho_{V_j}+\rho_{\varepsilon_j}} V_j - c_{ij}}{\sqrt{\frac{\rho_{\varepsilon_j}}{\left(\rho_{V_j}+\rho_{\varepsilon_j}\right)^2} + \frac{1}{\rho_\eta}}}\right).$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution.

We then model the probability that a consumer is aware of video j at time 1, denoted as W_{j1}. W_{j1} is a function of the characteristics of video j at time 1, such as the video ratings, the number of YouTube Favorites, and the number of video comments. The proportion of informed consumers for video j at time 1 is given by: $P_{j1} = W_{j1}$. Given a large number of viewers on YouTube[5], the number of views of video j at time 1 is

[5] Over 800 million unique users visit YouTube each month, see http://www.youtube.com/t/press_statistics. According to the law of large numbers, we can calculate the number of video views. If the number of potential consumers is not sufficiently large,

$$views_{j1} = P_{j1}\Pi_1 N = W_{j1}\Pi_1 N,$$

where N is the number of potential consumers.

With social learning, consumers have more information at time 2 because they receive feedback from friends. The underlying social network $\Gamma = (M, L)$ is given by a finite set of nodes $M = \{1,2,\dots,N\}$ and a set of links $L \subseteq M \times M$. Each node represents a consumer. The social connections between the consumers are described by an $N \times N$ matrix denoted by $g \in \{0,1\}^{N\times N}$ such that:

$$g_{ij} = \begin{cases} 1, & \text{if } (i,j) \in L \\ 0, & \text{otherwise} \end{cases}.$$

Let $M_i(g) = \{j \in N: g_{ij} = 1\}$ represent the set of friends of consumer i. We assume that consumer *i* has *k* friends, where $k = \#N_i(g)$. Among them, k_1 friends have watched the video at time 1. These friends communicate to consumer *i* their ex-post utilities after watching the video, u_{mj}, where *m* =1, 2, 3, ... ,k_1. r_1 friends were aware of the video, but decided not to watch the video at time 1, and they are indexed by *m*= $k_1 + 1$, $k_1 + 2$, ... , $k_1 + r_1$. The friends who have decided not to watch the video also provide valuable information: their expected utilities are less than the opportunity cost of watching the video. The number of friends who were unaware of the video at time 1 is $k - k_1 - r_1$.

As a result, at time 2, consumer *i*'s information set consists of the ex-post utilities of some friends, the number of friends who decided not to watch the video, and the number of friends who were unaware of the video. Combining these three pieces of information, consumer *i* estimates the quality by maximizing the likelihood of the observed evidence:

$$LH\left[u_{mj}, m = 1, 2, 3, \dots, k_1;\ r_1\ ; k - k_1 - r_1 \middle| V_j\right]$$

Chebyshev's inequality can give us a bound on views.

$$= \Pi_{m=1}^{k_1} f(u_{mj}) \cdot \Pi_{m=k_1+1}^{k_1+r_1} \Pr\left(E_1[u_{mj}|I_1] < c_{ij}\right) \cdot \Pi_{m=1}^{k_1+r_1} P_{j1} \cdot \Pi_{m=k_1+r_1+1}^{k} (1 - P_{j1}),$$

where $f(u_{mj})$ is the likelihood of observing u_{mj}. The maximum likelihood estimator, G_{ij2}, is an estimate of V_j. It is unbiased and asymptotically normal:

$$G_{ij2} \sim N(V_j, 1/d_{i2}),$$

where $d_{i2} = -E\left[\left(\frac{\partial \ln L}{\partial V_j}\right)^2\right]$ (Amemiya 1973).

If a consumer is aware of video j at time 2, his expected utility becomes the weighted average of the prior mean, his private signal, and the maximum likelihood estimator:

$$E_2[u_{ij}|I_2] = \frac{\rho_{V_j}}{\rho_{V_j} + \rho_{\varepsilon_j} + d_{i2}} X_j'\beta + \frac{\rho_{\varepsilon_j}}{\rho_{V_j} + \rho_{\varepsilon_j} + d_{i2}} S_{ij} + \frac{d_{i2}}{\rho_{V_j} + \rho_{\varepsilon_j} + d_{i2}} G_{ij2}.$$

Note that as time goes on, consumers place less weight on the prior mean. Because consumers receive more information at time 2, the prior becomes a less important factor in the decision making process. The probability that consumer i watches video j at time 2 is:

$$\Pi_2 = \Pr\left(E_2[u_{ij}|I_2] > c_{ij}\right) = \Phi\left(\frac{\frac{\rho_{V_j}}{\rho_{V_j}+\rho_{\varepsilon_j}+d_{i2}} X_j'\beta + \frac{\rho_{\varepsilon_j}+d_{i2}}{\rho_{V_j}+\rho_{\varepsilon_j}+d_{i2}} V_j - c_{ij}}{\sqrt{\left(\rho_{\varepsilon_j} + d_{i2}\right) / \left(\rho_{V_j} + \rho_{\varepsilon_j} + d_{i2}\right)^2 + 1/\rho_\eta}}\right),$$

If a consumer is aware of video j at time T, the decision making process proceeds in the same way. Consumer i has k_t friends who decide to watch the video at t, r_t friends who decide not to watch the video at time t, and $k - k_t - r_t$ friends who are unaware of the video, where $t = 1, 2, 3, \ldots, T-1$. The probability that consumer i watches video j at time T is:

$$\Pi_T = \Pr\left(E_T[u_{ij}|I_T] > c_{ij}\right) = \Phi\left[\frac{\alpha_T X_j'\beta + (1-\alpha_T)V_j - c_{ij}}{g_T}\right]. \tag{2}$$

where $g_T = \sqrt{\left(\rho_{\varepsilon_j} + \sum_{t=2}^{T} d_{it}\right) / \left(\rho_{V_j} + \rho_{\varepsilon_j} + \sum_{t=2}^{T} d_{it}\right)^2 + 1/\rho_\eta}$, and $\alpha_T = \frac{\rho_{V_j}}{\rho_{V_j}+\rho_{\varepsilon_j}+\sum_{t=2}^{T} d_{it}}$.

The proportion of informed consumers at time T is given by: $P_{jT} = P_{jT-1} + (1 - P_{jT-1})W_{jT}$, where $1-P_{jT-1}$ is the proportion of uninformed consumers at time $T-1$, and W_{jT} is the probability that a consumer is aware of video j at time T. The number of views of video j at time T is:

$$views_{jT} = (P_{jT} - P_{jT-1})\Pi_T N, \tag{3}$$

where $P_{j0} = 0$.

A few remarks need to be made here. In the process of social learning, α_T is the weight that consumers put on the ex-ante prior. It is evident that α_T decreases with T. As time T grows, the probability of watching videos relies less on the ex-ante prior and more on social learning. If the revealed quality of the video is higher than the mean of the ex-ante prior, $V_j > X_j'\beta$, we call it a positive surprise. If the revealed quality of the video is lower than the prior mean, $V_j < X_j'\beta$, we call it a negative surprise. Unanticipated variables entering the formulation of a model has been documented in a large body of literature (Baillie 1987). Social learning is a process of adjusting beliefs about the quality according to the surprises. Thus, we have the following proposition:

Proposition 1. *In the presence of social learning, if a positive surprise is sufficiently large ($V_j \gg X_j'\beta$), then Π_T is increasing in T. If a negative surprise is sufficiently large ($X_j'\beta \gg V_j$), then Π_T is decreasing in T.*

Proof. From equation (2), we can obtain:

$$\Pi_{T+1} - \Pi_T$$
$$= \Phi\left(\frac{\alpha_{T+1}X_j'\beta + (1-\alpha_{T+1})V_j - c_j}{g_{T+1}}\right) - \Phi\left(\frac{\alpha_T X_j'\beta + (1-\alpha_T)V_j - c_j}{g_T}\right).$$

If $V_j - X_j'\beta > \frac{1}{\xi_T}\left(\frac{1}{g_T} - \frac{1}{g_{T+1}}\right)\left(X_j'\beta - c_j\right)$, where $\xi_T = \frac{1-\alpha_{T+1}}{g_{T+1}} - \frac{1-\alpha_T}{g_T} > 0$, we have $\Pi_{T+1} - \Pi_T > 0$. Thus, if a positive surprise is sufficiently large, then Π_T is increasing in T.

Similarly, we can show that if a negative surprise is sufficiently large, then Π_T is decreasing in T. ■

Our model shows that consumers learn about the surprise over time, and a positive surprise increases the expected quality of the video as time passes. Therefore, the probability of watching the video increases. Similarly, a negative surprise reduces the expected quality over time, and the probability of watching the video decreases.

Other things being equal, we consider some video j with a large positive surprise and some video j' with a large negative surprise. According to equation (3), we find that:

$$\ln views_{jT} - \ln views_{jT-1} > \ln views_{j'T} - \ln views_{j'T-1}.$$ [6]

Therefore, we have the following testable hypothesis from the theoretical prediction:

Hypothesis 1. *In the presence of social learning, the growth rate of views of a video that has a positive surprise is higher than the growth rate of views of a video that has a negative surprise.*

In our model of social learning, we can also consider the impact of the consumers' prior. The intuition is that the effect of social learning is more pronounced for videos with less precise priors. If a consumer is very uncertain about the quality of a video, the value of social learning is large: The additional information he learns from his friends should be more valuable than the case when he knows the quality precisely.

Proposition 2. *(a) If the positive surprise is sufficiently large, $\Pi_{T+1} - \Pi_T$, is decreasing in the precision of the prior, ρ_{V_j}. (b) Similarly, if the negative surprise is sufficiently large, $-(\Pi_{T+1} - \Pi_T)$ is decreasing in ρ_{V_j}.*

Proof. From equation (2), we can obtain:

[6] The logarithm growth rates are widely used in economic modeling and empirical study. They are good approximations for percentage growth rates.

$$\frac{\partial}{\partial \rho_{V_j}}(\Pi_{T+1} - \Pi_T) < 0,$$

when the positive surprise $(V_j - X_j'\beta)$ is sufficiently large. Part (b) can be proved similarly. ■

The incremental probability, $\Pi_{T+1} - \Pi_T$, can measure the effect of social learning. An increase in the precision of the prior makes the additional information from friends less valuable. Thus, social learning should be more valuable among videos that are less familiar to consumers, and we have the following empirically testable Hypothesis 2:

Hypothesis 2. *In the presence of social learning, the positive surprise has a greater impact on videos with less precise priors.*

3.2 A Model of Network Effects on YouTube

Besides social learning, social contagion can also be driven by network effects. Network effects in our context refer to only the direct network effect where the utility of watching a video directly depends on the number of consumers who have watched the video, irrespective of their reasons for the choice of watching the video (Katz and Shapiro 1985). While indirect network effects have been widely studied for two-sided platforms (Tucker and Zhang 2010), they are beyond the scope of this paper. The underlying mechanism for network effects is different from that of social learning. Essentially, network effects are payoff externality, which implies that the value of the service depends directly on the consumption choices made by some other consumers. For example, a consumer might enjoy discussing a video with his peers. In this case, the actions of other consumers do not convey any quality information about the video but still increase the consumer's probability of watching. On the contrary, social learning leverages information externality instead of payoff externality. In other words, social learning influences consumer decision through the conveyed sentiment towards the video, while network effects affect the decision through the total number of viewers irrespective of their attitudes towards the video.

To model network effects, we introduce another component to reflect the impact of total number of viewers such that the utility consumer i obtains from watching YouTube video j at time T is given by:

$$u_{ijT} = V_j + \eta_{ij} + \delta \sum_{t=1}^{T-1} (P_{jt} - P_{jt-1}) \Pi_t N \, , \; \eta_{ij} \sim N\left(0{,}1/\rho_\eta\right),$$

where $(P_{jt} - P_{jt-1})\Pi_t N$ is the number of consumers who have watched the video at time *t*. The consumer derives direct utility from the total number of consumers who have watched the video. The parameter δ measures the magnitude of network effects. If $\delta > 0$, then network effects exist. If $\delta = 0$, then the impact of network effects is insignificant. For pure network effects model, we assume no social learning, and, consequently, consumers do not receive feedback from peers.

Under network effects, the probability that a consumer watches video j at time *T* is given by:

$$\Pi_T = \Phi\left(\frac{\frac{\rho_{V_j}}{\rho_{V_j}+\rho_{\varepsilon_j}} X_j'\beta + \frac{\rho_{\varepsilon_j}}{\rho_{V_j}+\rho_{\varepsilon_j}} V_j + \delta \sum_{t=1}^{T-1} (P_{jt} - P_{jt-1}) \Pi_t N - c_{ij}}{\sqrt{\rho_{\varepsilon_j} / \left(\rho_{V_j} + \rho_{\varepsilon_j}\right)^2 + 1/\rho_\eta}}\right)$$

It is evident that Π_T is increasing in *T* under network effects irrespective of the values of $X_j'\beta$ and V_j. What matters in the case of network effects is the initial viewer base. With strong network effects, a video that shows a negative surprise but has generated more initial views because of high ex-ante quality expectation may perform better than a video that shows large positive surprise. In fact, if social contagion is purely driven by network effects, surprise does not matter, and we should not observe the empirical pattern described in Hypothesis 1.

One way to empirically identify network effects is to examine sequential correlation in video views such that the views at time *t* is positively correlated with the lagged cumulative video view, $(P_{jt} - P_{jt-1})\Pi_t N$. However, we would not be able to uniquely identify network effects without controlling for social learning. We use a two-stage test to distinguish the two confounding mechanisms. Recall that in Hypothesis 1, with social learning, a video with a positive surprise has a higher growth rate than a video with a negative surprise. However, under network effects, a significant difference in the growth rates resulting from surprises would be absent. Therefore, if the empirical evidence supports Hypothesis 1, then it would show the existence of social learning. If Hypothesis 2 is also confirmed, it would provide additional evidence for social learning. We also look into whether other diffusion mechanisms, such as social pressure and conformity concerns, could result in a similar pattern described in Hypothesis 1 and 2. According to Iyengar et al. (2012), there is no correlation between social conformity and product quality. Therefore, the impact of "quality surprise" should be insignificant under social conformity.

In the second stage, we test whether network effects exist using instrumental variable as a source of exogenous variation for existing levels of views. If social contagion is purely driven by social learning, the growth rate of video views should remain unchanged when the shock does not reflect information about video quality. Such uninformative surprises do not incur any social learning. However, in the presence of network effects, a negative shock that does not contain any quality information reduces the viewer base, and thus leads to a lower growth rate of video views. Therefore, we can identify network effects by testing Hypothesis 3. The detailed description of the two-stage test is given in Section 5.

Hypothesis 3. *In the presence of network effects, the negative surprise lowers the growth rate of video views even if the negative surprise does not reflect information about video quality.*

4. Data

To empirically test the theoretical model, we look at new videos that were posted during our data collection period. As the world's largest video viewing and sharing website, YouTube has enormous numbers of videos, which makes random sampling infeasible. Instead, we focus on the most active providers by selecting the top 1,000 YouTube providers (in terms of total video views) identified for June 2011.[7] We collected a daily panel of data on these providers for one month, from March 1, 2012 to March 31, 2012. Our sample includes a total of 302 new videos published on March 1, 2012 by these top providers. We use one day as the time unit of analysis to capture the fast-changing nature of online videos.

The provider level data include provider ID, data collection date, date when the provider joined YouTube, number of subscribers to the provider's channel, total views of all the provider's videos, total views of the provider's channel page, number of videos, number of friends, number of subscriptions the provider has to other providers, channel views rank, and video views rank. The video level data include video ID, data collection date, date when the video is posted, the provider of the video, number of views, category in which the video belongs, video length, whether the video has an in-stream ad, average rating, number of times the video is favorited by viewers, and number of comments. All videos in our sample were published on March 1, 2012. Because YouTube Analytics data is updated daily, the first day in our analysis is March 2, 2012. Summary statistics of the video characteristics at the beginning of our data collection period are reported in Table 1. We assume that each viewer watches a video only once. Although consumers may repeatedly watch a video, the bias caused by repeated viewings is small if logs of views are used instead of views (Susarla et al. 2012). Table 2 provides summary

[7] In this study, we focus on social learning and network effects given that consumers are aware of the video. We do not study how consumers become aware of a video. That is why we select the videos published by the top 1,000 YouTube providers as our sample.

statistics of the characteristics of our YouTube providers.

Table 1. The First-Day Video Characteristics

Variables	Mean	Std. Dev.	Min	Max
Number of video views	2,497.20	8,524.969	2	107,628
Video rating	4.692	0.5066	1.76	5
Number of times the video being favorited by viewers	139.507	495.910	0	6,800
Number of comments	384.173	1030.539	0	9,832
In-stream ads (yes-1, no-0)	0.5359	0.4995	0	1

Some YouTube providers also post their video links on Twitter. We control for these personal marketing efforts when estimating social contagion. Using Twitter application programming interface (API), we collected a random selected sample of all Twitter data containing the key word #YouTube or video. Using the collected tweets, we analyzed the included shortened uniform resource locator (URL) link related to YouTube and recovered the unique YouTube video ID. Then we used a simple natural language processing on tweet content to identify the tweets posted by video providers as new video announcements such as "I uploaded a new video …, please check out."

Table 2. The First-Day Chanel Characteristics

Variables	Mean	Std. Dev.	Min	Max
Total views of the provider's channel page	1.32e+07	2.01e+07	3,175,291	1.75e+08
Total views of all the provider's videos	1.87e+08	2.38e+08	3,690,640	1.55e+09
Number of subscribers to the provider's channel	298,743.8	459,699.2	9,200	5,109,145
Number of subscriptions by the provider to other providers	183.6144	1,063.834	0	17,641
Number of videos	267.9837	289.788	1	969
Number of friends	16,976.32	22,916.99	0	120,570

5. Empirical Framework

5.1 Identification of the Surprise

In our theoretical model, the surprise is defined as the difference between the revealed quality and the prior mean, $V_j - X_j'\beta$. Following Barro (1977), Hirshleifer, Lim, and Teoh

(2009), and Moretti (2011), we empirically define the surprise as the difference between realized video views and predicted video views at time 1 (March 2, 2012).[8] In our study, we consider the prediction error to be a surprise. More specifically, we argue that YouTube video views on the first day can be predicted by the characteristics of YouTube providers (channel). Therefore, we use the residual from a regression of first-day log views on the characteristics of YouTube providers at time 1 and video category as a measure of video-specific surprise. The residual measures the deviation of realized video views from the rational expectation at time 1. The residual may change with different functional forms used for the predicted views. Therefore, as a robustness check, we test our hypotheses with different regression specifications.

The characteristics of YouTube providers are reasonable measures of expected video quality. YouTube allows consumers to subscribe to the providers they would like to follow. By subscribing to a provider, they are informed immediately when the provider posts a new video. The prior information about the provider shapes to a large extent a consumer's expectation of the provider's new videos.

Table 3. Identification of the Surprise: First-Stage Regression

	(1)	(2)	(3)	(4)
lvviews	0.444***	0.445***	0.449***	0.0952***
	[12.83]	[12.89]	[13.0]	[3.112]
lcviews	0.297***	0.297***	0.291***	0.0799*
	[4.809]	[4.798]	[4.710]	[1.661]
lvideos	0.0175	0.0176	0.0179	0.0171
	[0.564]	[0.568]	[0.579]	[1.512]
subs	3.26e-08	3.29e-08		4.43e-08
	[0.138]	[0.140]		[0.530]
subscriptions	2.71e-06			1.33e-05
	[0.0360]			[0.504]
Constant	1.630	1.629	1.410	5.090***
	[0.684]	[0.685]	[0.787]	[5.886]
Observations	302	302	302	302
R-squared	0.545	0.545	0.544	0.528

t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

8 In finance literature, Jegadeesh and Titman (1993) document a momentum phenomenon: Firms reporting positive earnings surprises outperform firms reporting negative earnings surprises.

Table 3 illustrates the first-stage regression results of first-day (March 2, 2012) log video views on provider characteristics at time 1. The predicted value of the dependent variable obtained from this first-stage regression is a measure of expected video quality. The residual is used as a measure for surprises to test Hypothesis 1. The channel characteristics include the log of total views of channel j's videos, $lvviews$; the log of total views of the provider's channel page, $lcviews$; the log number of uploaded videos of the channel, $lvideos$; the number of the provider's subscribers, $subs$; and the number of other providers the provider subscribe to, $subscriptions$. The independent variables also include a set of dummy variables indicating the video category. Column 1 in Table 3 shows the regression results. Column 2 and 3 indicate that the results are robust to other regression specifications.

To check the robustness of our measure of surprises, we also use another empirical measure of surprises: the difference between realized video ratings and predicted video ratings at time 1. Similarly, we regress first-day video ratings on the characteristics of providers and use the residual as a measure of surprises. This result is shown in Column 4 in Table 3. The robustness test results are shown in Section 5.2 and Section 5.4. We also need to check whether the residuals measure surprises rather than simply reflect omitted variables that are observable to the viewer but not the researcher. If it is the case, the residual would be systematically correlated with viewership. However, we regress log views on the residual at time 1, and find the coefficient insignificant (p value = 0.331).

Figure 3 shows a clear example of videos with different surprises. The figure plots the daily video views for a video with a positive surprise (video 2) and a video with a negative surprise (video 1). These two videos belong to the same YouTube video category and have similar initial views, but experience different growth patterns: Video 2, having positive surprise, has a significantly higher growth rate than video 1, having a negative surprise. The first-day views of video 1 and video 2 are roughly the same (304 and 449 respectively). However, at the end of our sample period, views of video 1 and views of video 2 are 1,102 and 25,508 respectively. As stated in our theoretical model, this striking difference is likely to be caused by social learning over time.

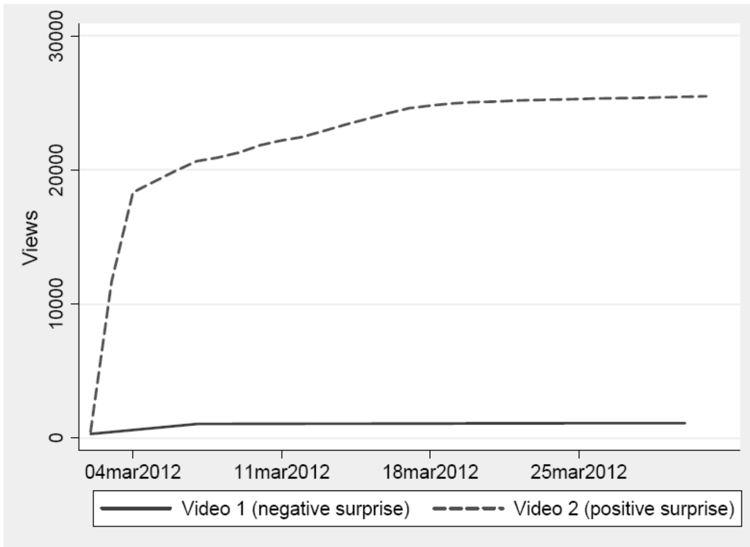

Figure 3. Daily Views for Videos with Different Surprises

5.2 A Test of Social Learning

5.2.1 The Existence of Social Learning

According to Hypothesis 1, in the presence of social learning, the growth rate of views of a video with a positive surprise is higher than the growth rate of views of a video with a negative surprise. If only network effects exist, we should see no significant difference between the growth rate of a video with a positive surprise and that of a video with a negative surprise.

Our empirical approach is based on the literature on treatment effects (Wooldridge 2007). The positive surprise is interpreted as the “treatment”, and views of "treated" videos are compared to views of the “control” videos with negative surprises.

To test whether the difference between the growth rates is significant, we estimate the following model using difference-in-difference:

$$\ln views_{jt} = b_0 + b_1 t + a_j + \Psi_{jt}' b_2 + b_3 (t \times D_j) + b_4 (t \times EV_j) + \mu_{jt}, \quad (4)$$

where $\ln views_{jt}$ is the log of views of video j at time t, t represents the time period, a_j represents unobserved fixed effect of video j, and Ψ_{jt} includes the characteristics of video j that change over time, such as $rating$ (the video ratings), $favs$ (the number of YouTube Favorites), $comment$ (the number of video comments), $lvideo$ (the log number of uploaded videos of the channel), $lvviews$ (the log of total video views of the channel), and $subs$ (the number of channel subscriptions). We control the marketing efforts of YouTube providers on Twitter, which are measured by sum_upload_{jt}, the total number of tweets containing the unique YouTube video ID and the word "uploaded". Unlike the specifications in Duan et al. (2009) and Susarla et al. (2012), equation (4) does not contain lags of accumulative views, because the lag terms do not help distinguish between social learning and network effects. In our context, both social learning and network effects can lead to a positive effect of previous views on current views: (1) Consumers learn from other people's choices. They infer the quality is high when they see a larger number of accumulative views. (2) Consumers can obtain a higher utility from a larger view base because they enjoy discussing a video with their peers.

In the regression, D_j is a dummy variable indicating whether the surprise of video j is positive ($D_j = 1$, if the surprise is positive; $D_j = 0$, otherwise), EV_j is the expected video quality measured by the predicted value of the dependent variable obtained from the first-stage regression in Section 5.1, and μ_{jt} is the error term. Note that we expect that b_4 is not significantly different from 0 because the expected quality should not change the growth rate of video views over time after controlling for other variables in Model (4).

Following the literature on treatment effects (Wooldridge 2007), we make the unconfoundedness assumption: $t \times D_j$ is strictly exogenous. Note that the correlation between $t \times D_j$ and μ_{jr} for any time t and time r causes inconsistency in regression coefficients. Thus, we need to control for the time-varying heterogeneity (Ψ_{jt}), and the unobserved fixed effects in the regression. If the surprise assignment (positive or negative) changes in reaction to past outcomes on $\ln views_{jt}$, strict exogeneity can be violated (Wooldridge 2007). However, because the surprise assignment is determined at time 1 and is independent of the idiosyncratic views shocks in period t, strict exogeneity is a reasonable assumption.

We are interested in the difference-in-difference estimator, b_3. If $b_3 > 0$, then the difference between the growth rates is positive, supporting Hypothesis 1 and the existence of social learning. If $b_3 = 0$, then the growth rates of video views with different surprises are the same, which indicates no significant social learning.

The fixed effects regression results are shown in Table 4. In the table, $interaction = t \times D_j$. Column 1 shows the results from a regression that includes all the coefficients specified in equation (4) except b_4. In this regression, $\hat{b}_3 = 0.0313$ and is significantly positive, which confirms Hypothesis 1. Note that in the regression model, after controlling for online WOM, such as $rating$, $favs$, and $comment$, the impact of a positive quality surprise is still significantly positive. This suggests that social learning is beyond reading comments and reviews.

As expected, Column 2 shows that $\hat{b}_4$ is insignificant, which implies that the anticipated quality does not have a significant impact on the growth rate of video views. This result is reminiscent of rational expectations models: only unanticipated factors affect real economic variables. Barro (1977) examines the rational expectation monetary models and empirically found that only the unanticipated part of money movements has real effects on the unemployment rate.

Table 4. Fixed Effects Regression of Video Views on Surprises: A Test of Social Learning

	(1)	(2)	(3)	(4)	(5)	(6)
Interaction ($\boldsymbol{t} \times \boldsymbol{D_j}$)	0.0313***	0.0271***	0.0313***	0.0480***	0.0218**	0.0176***
	[12.58]	[5.041]	[3.201]	[7.727]	[2.248]	[30.86]
rating	0.262***	0.213**	0.262**	1.542***	1.105***	0.206***
	[2.934]	[2.301]	[2.245]	[2.771]	[2.652]	[2.442]
favs	0.000110***	0.000329***	0.000110*	0.000834***	0.000134	0.000172***
	[2.920]	[3.642]	[1.857]	[6.157]	[0.890]	[4.873]
comment	0.000213***	0.000124***	0.000213***	-0.000406***	0.000423***	0.000134***
	[7.835]	[5.337]	[4.544]	[-3.369]	[4.506]	[5.204]
sum_upload	0.0300	0.0397	0.0300	0.0431	0.0534	0.0241
	[0.461]	[0.224]	[0.211]	[0.241]	[0.521]	[0.390]
Interaction ($\boldsymbol{t} \times \boldsymbol{EV_j}$)		0.00134				
		[0.702]				
Observations	9060	9060	9060	1290	3390	9060
R-squared	0.328	0.384	0.328	0.459	0.564	0.110

t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

Using residuals to measure unanticipated variables (the surprise) has a long tradition in macroeconomics and finance. A number of studies use the following two-step regression procedure to estimate the effect of the unanticipated variables, such as unanticipated inflation (Sargent 1976), unanticipated money growth (Barro 1977), earnings surprises (Hirshleifer, Lim, and Teoh 2009), and unanticipated movie quality surprises (Moretti 2011): First, the residuals from a separate auxiliary regression are used as a proxy for the unanticipated variable and then the residuals are used as an explanatory variable in the equation of interest (Baillie 1987). Simply using ordinary least squares (OLS) without adjusting for the extra variance of the generated regressor term (the surprise) can yield consistent estimates but invalid statistical inferences (Pagan 1984; Wooldridge 2002). The existing literature such as Barro (1977) and Moretti (2011) tends to underestimate the standard errors of coefficient estimates. It is crucial to address this issue in the context of social media because the data tend to be more noisy. We use the two-step bootstrapping method proposed by Cameron and Trivedi (2005) to obtain proper statistical inferences. The asymptotically refined result is presented in Column 3.

5.2.2 Social Learning and Video Popularity

Social learning effect differs across video categories. Column 4 and Column 5 in Table 4 present regression results for subsamples of two categories: "Tech" and "Music."According to a survey by Sysomos Inc. (http://www.sysomos.com/reports/youtube/), Music is the most popular category on YouTube, and Tech is the least popular category. We find that social learning is more pronounced for "Tech" videos than for "Music" videos. This finding indicates that social learning affects videos belonging to unpopular categories more. The implication is that the role of social learning becomes more salient in a niche market than in a mass market. Column 6 shows that the results are robust when we define the surprise as the difference between realized video ratings and predicted video ratings.

A striking pattern in the data is that video views are remarkably skewed. The top 10 videos account for 47.46% of total views, and the top 30 account for 66.81%. Quantile regression analysis is particularly useful when the conditional distribution of video views is heterogeneous and does not have a "standard" shape (Koenker and Hallock 2001), to account for unobserved heterogeneity and heterogeneous covariates effects. A simple differencing strategy used in fixed effects estimation shown in Table 4 is infeasible for quantile regressions since quantiles are not linear operators. So we adopt an estimator that is consistent and asymptotically normal (Canay 2011) to compute the quantile estimates.

In Figure 4, we plot the parameter estimates b_3 of the quantile regressions based on equation (4). There are nine estimated quantile regressions with 0.1, 0.2, …, and 0.9 quantiles. The parameter estimates of the quantile regressions are connected by the solid line, with the shaded area being their 95% confidence intervals. The fixed effects estimate b_3 shown in Table 4 is plotted as horizontal dashed lines in this figure. We find that the quantile regression parameter estimates are significantly positive for all quantiles and decrease with quantiles in general. This suggests that the impact of a positive surprise is higher for less popular videos, because the same magnitude of positive surprises implies more pronounced social learning for less popular videos than for popular videos. Since most of popular videos are published by top content creators on YouTube and consumers have high ex-ante expectations on these videos, the impact of an additional positive surprise (social learning) is relatively small. For example, financed by venture capitalists and grants from YouTube, Maker studios produced a popular sketch comedy show called "AsKassem" for YouTube (Miller 2011). Consumers who have watched AsKassem #1 - #70 form high expectations on the quality of the new episode #71. Tucker and Zhang (2011) find a similar result: Popularity information benefits niche products with narrow appeal more than the mainstream products. Figure 4 also indicates that the panel data model with fixed effects tends to underestimate the impact of a positive surprise for less popular videos and tends to overestimate the impact for popular videos.

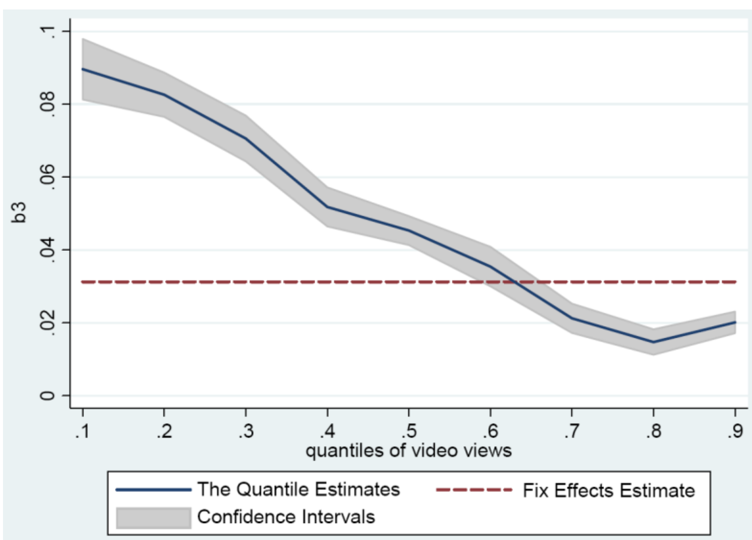

Figure 4. The Estimates of Quantile and Fix Effects Regressions

5.2.3 Social Learning Effects over Time

An underlying assumption of Regression Model (4) is that the effect of social learning is constant. The time trend is assumed to be linear. However, the effect of social learning may vary over time. We estimate Model (5) using fixed effects (FE). The dummy variable day_t indicates the time period *t*. We are interested in the coefficient φ_t.

$$\ln views_{jt} = b_0 + \sum_{t=1}^{29} \rho_t day_t + a_j + \Psi'_{jt} b_2 + \sum_{t=1}^{29} \varphi_t \,(day_t \times D_j) + \mu_{jt}, \tag{5}$$

Table 5. Estimating the Effect of Social Learning over Time Using Model (5)

VARIABLES	FE	VARIABLES	FE	VARIABLES	FE
$day_1 \times D_j$	0.0661***	$day_{11} \times D_j$	0.0465***	$day_{21} \times D_j$	0.0049***
	[9.293]		[11.79]		[7.645]
$day_2 \times D_j$	0.0664***	$day_{12} \times D_j$	0.0431***	$day_{22} \times D_j$	0.0094***
	[11.57]		[11.31]		[7.396]
$day_3 \times D_j$	0.0652***	$day_{13} \times D_j$	0.0396***	$day_{23} \times D_j$	0.0048***
	[11.92]		[10.86]		[7.003]
$day_4 \times D_j$	0.0643***	$day_{14} \times D_j$	0.0354***	$day_{24} \times D_j$	0.0072***
	[11.68]		[10.31]		[6.612]
$day_5 \times D_j$	0.0646***	$day_{15} \times D_j$	0.0319***	$day_{25} \times D_j$	0.0052***
	[13.32]		[10.03]		[6.163]
$day_6 \times D_j$	0.0617***	$day_{16} \times D_j$	0.0269***	$day_{26} \times D_j$	0.0051***
	[12.97]		[9.736]		[5.776]
$day_7 \times D_j$	0.0594***	$day_{17} \times D_j$	0.0229***	$day_{27} \times D_j$	0.0074***
	[12.61]		[9.459]		[5.420]
$day_8 \times D_j$	0.0581***	$day_{18} \times D_j$	0.0183***	$day_{28} \times D_j$	0.0041***
	[12.87]		[8.999]		[5.022]
$day_9 \times D_j$	0.0547***	$day_{19} \times D_j$	0.0141***	$day_{29} \times D_j$	0.0064***
	[12.58]		[8.014]		[4.551]
$day_{10} \times D_j$	0.0503***	$day_{20} \times D_j$	0.0089***		
	[12.21]		[8.132]		
Observations	9060				
R-squared	0.385				

t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

Table 5 shows that φ_t is large when t is small. This implies that the effect of social learning is more pronounced in earlier periods. Our finding is consistent with the empirical facts in Tang, Gu, and Whinston (2012): Most videos receive the majority of views in the first four to five days after being posted, and only marginal views thereafter.

5.2.4 Further Analysis on Social Learning

An issue with our empirical test is that Proposition 1 depends on the condition that the surprise is sufficiently large. In the previous estimations, the sample is divided into two groups in terms of positive or negative surprises. We check for robustness by further dividing our sample into four equally sized groups depending on the magnitude of the surprise. For example, Group 1 includes the videos with the lowest 25% of the level of the surprises, and Group 4 includes the videos with the highest 25% of the level of the surprises. Then we estimate the following specification:

$$\ln views_{jt} = b_0 + b_1 t + a_j + \Psi_{jt}' b_2 + \sum_{g=2}^{4} \delta_g (t \times S_g) + \mu_{jt},$$

where S_g, for g = 2, 3, 4, are dummy variables indicating the video belongs to the surprise group g, and δ_g, for g = 2, 3, 4, are coefficients on $t \times S_g$. Table 6 reports the results, which are qualitatively similar to those in Table 4.

Table 6. Fixed Effects Regression under Different Surprise Groups

	Fixed Effects Regression
$\boldsymbol{t \times S_4}$	0.0424***
	[31.95]
$\boldsymbol{t \times S_3}$	0.0131***
	[17.31]
$\boldsymbol{t \times S_2}$	0.00412***
	[2.920]
Observations	9060
R-squared	0.486

t-statistics in brackets: *** p<0.01, ** p<0.05, * p<0.1

According to Table 6, $\delta_4 > \delta_3 > \delta_2 > 0$. This is consistent with our intuition: In the presence of social learning, the views of a video with a higher level of surprise should increase more rapidly. We also conduct two hypothesis tests: the null hypotheses are δ_4=δ_3 and δ_3=δ_2. The *F*-statistics are 271.72 and 9.47, respectively, which implies that we can reject both null hypotheses with a significance level of 0.05. As a result, we can conclude that the social learning effects are stronger with larger positive surprises.

We also check if quality (rather than social learning) itself can explain all the findings. A high-quality video can easily be forwarded by people who happen to view them first to their friends. Through request and compliance, more and more people will view the video and realize the true quality. However, this cannot explain our result: the impact of a positive quality surprise is significantly positive. We find that when the ex-ante expectation of the quality is extremely high, a high-quality video may have a negative surprise that can slow down its growth rate, because the true quality fails to fulfill the expectation.

Hypothesis 2 indicates that social learning is more important for videos with less precise priors. To test the hypothesis, we estimate the following model:

$$\ln views_{jt} = b_0 + b_1 t + a_j + \Psi'_{jt} b_2 + b_3 (t \times D_j)$$
$$+ b_4 (t \times prior_j) + b_5 (t \times D_j \times prior_j) + \mu_{jt}, \tag{6}$$

where $prior_j$ is a measure of the prior precision of video j. Here we propose the total views of the provider's channel page on the first day (March 2, 2012) to empirically identify which videos have more precise priors. YouTube users upload videos to their YouTube channels. A consumer has a better idea of the quality of a new video published by a high-ranking channel (in terms of channel page views) because the consumer is more likely to have watched another video published by the same channel before. In this case, videos published by higher-ranking channels have more precise priors. We divide the sample into two equally sized groups by channel views: the high-ranking group and the low-ranking group. If a video belongs to the high-ranking group, then the dummy $prior_j = 1$; otherwise, $prior_j = 0$.

Table 7. The Effect of Prior Precision on Social Learning

	(1)	(2)	(3)	(4)
interaction $(\boldsymbol{t} \times \boldsymbol{D_j})$	0.0152***	0.0333***	0.0129***	0.0294***
	[12.15]	[9.581]	[10.76]	[10.20]
tprior $(\boldsymbol{t} \times \boldsymbol{prior_j})$	0.0113***	0.0482***	0.0236***	0.0481***
	[15.15]	[25.47]	[30.88]	[20.96]
tdprior $(\boldsymbol{t} \times \boldsymbol{D_j} \times \boldsymbol{prior_j})$	-0.0151***	-0.0523***	-0.0162***	-0.0466***
	[-8.142]	[-10.16]	[-9.050]	[-8.592]
rating	0.154*		0.0371	0.124
	[1.744]		[0.445]	[0.701]
favs	0.000136***		6.40e-05*	5.84e-05
	[3.677]		[1.823]	[0.882]
comment	0.000160***	0.000871***	0.000265***	0.00104***
	[5.943]	[18.99]	[10.43]	[22.66]
sum_upload	0.0157		0.186***	0.473***
	[0.246]		[3.058]	[3.542]
Observations	9060	9060	9060	9060
R-squared	0.353	0.253	0.222	0.244

t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

The coefficient of interest here is b_5 in the regression model (6). Table 7 illustrates that the empirical evidence is consistent with Hypothesis 2. In the table, $tprior = t \times prior_j$, and $tdprior = t \times D_j \times prior_j$. In Column 1, we find that the coefficient on $tdprior$, $\hat{b}_5$, is -0.0151, which is significantly negative. We can consider two identical videos with the same positive surprise except for the fact that the first belongs to the high-ranking group and the second belongs to the low-ranking group. In the presence of social learning, a negative b_5 implies that the growth rate of views of the second video is higher than the growth rate of views of the first video. In other words, social learning has a greater effect on videos with less precise priors, which supports Hypothesis 2. Column 2 indicates that the estimate of b_5 is robust to a different specification.

We also show that the regression results are robust to different measures of the prior precision. We use two other measures of the prior precision: the number of subscribers and the variance of the ratings of the channel’s past videos. If more YouTube users subscribe to a channel, consumers are more certain about the quality of the videos from that channel. Similarly, we divide the sample into two equally sized groups, based on subscribers rank. If a video belongs to the high-ranking group, then the dummy $prior_j = 1$; otherwise, $prior_j = 0$. We also calculate the variance of the ratings of the channel's videos at day 1 (March 2, 2012), and divide the sample into two equally sized groups: If a video belongs to the low-variance group, then the dummy $prior_j = 1$; otherwise, $prior_j = 0$. If the variance is low, the quality of the videos from that channel is more homogeneous, and consumers are more certain about the quality. Column 3 and Column 4 in Table 7 show that the coefficient on $tdprior$ is significantly negative. Different measures of the prior precision do not affect our key results.

5.3 A Test of Network Effects

In this section, we test the existence of network effects on YouTube using the presence of in-stream ads as a source of exogenous variation for existing levels of video views. YouTube in-stream ads run only on partner videos. Only successful content creators are qualified for the partner program, and videos published by them might contain in-stream ads. It is reasonable to assume that the presence of in-stream ads is exogenous in our context. It is possible that advertisers are more likely to use the channels that have higher viewership and higher quality videos. However, all of our sample videos are published by the top providers on YouTube, and all of them are partners. As a result, the presence of in-stream ads is not likely to be correlated with video quality. Empirically, we find that the presence of in-stream ads is not significantly correlated with viewership in our sample. Stock et al. (2002) propose a method based on the first-stage F statistic for detecting weak instruments. For an instrument to be reliable, the first-stage F statistic in the two-stage least squares (2SLS) regression should be greater than 8.96 when the number of instruments is one. We run a first-stage regression on our instrument variable and find that F statistic = 303.2 > 8.96. Therefore, we can reject the null hypothesis that the presence of in-stream ads is a weak instrument.

Network effects are identified by isolating the surprises caused solely by the presence of in-stream ads. YouTube bundles video content with in-stream ads, which are intrusive to many consumers.[9] The presence of in-stream ads is a negative shock that can reduce the viewer base. If network effects exist on YouTube, the negative shock lowers the growth rate of views at time 1, and then it further lowers the growth rate at time 2. As time goes on, we should see a significantly negative self-reinforcing feedback loop. However, such a negative shock does not contain any information of the video quality. If social learning is the sole form of social contagion, the Bayesian learning process remains unchanged. The negative shock can decrease the viewership at time 1, but the long run growth rate of video views should not be affected significantly (no self-reinforcing feedback loop).[10] If a consumer is shown an ad before the video, one may think this could impact social learning in that this would result in lower consumer satisfaction and more negative word of mouth. However, the additional information about ads from the peers is redundant. When consumers make decisions on whether to watch the video, they know whether the video contains an ad.[11] In summary, if there exists only social learning with network effects absent, the presence of ads is a transitory shock that does not have significant long run effect. If network effects exist, the presence of ads results in a negative self-reinforcing feedback loop.

We re-estimate model (4) to test network effects, using 2SLS regression, and the

---

[9] Wilbur (2008) estimates a two-sided model of the television industry and shows that viewers tend to be averse to advertising. Anderson and Gans (2011) study an advertising-sponsored content provision model and interpret advertising clutter as a "price" paid by viewers who enjoy the content.

[10] Let video j is a video without an ad, and video j' is a video contains an ad. Other things being equal, we can obtain $\ln views_{jT} > \ln views_{j'T}$, $\ln views_{jT-1} > \ln views_{j'T-1}$, and $\ln views_{jT} - \ln views_{jT-1} \approx \ln views_{j'T} - \ln views_{j'T-1}$ from our model of social learning.

[11] When consumers are shown an ad before the video, they can choose to switch to other videos. It is equivalent to not watching the video.

in-stream ads ads_j as instrument for the surprise dummy D_j. ads_j is a dummy, where $ads_j = 1$ if the video has an in-stream ad, and $ads_j = 0$ otherwise. Generally, 2SLS is used to avoid endogeneity. However, we use the first-stage regression to isolate the surprises that are caused solely by the shock of in-stream ads. We are interested in the coefficient b_3 in the regression model (4). If Hypothesis 3 (the presence of network effects) is supported, we expect to see that $b_3 > 0$, which implies that negative surprises lower the future views without revealing any quality related information. If $b_3 = 0$, then network effects are insignificant.

Table 8. A Test of Network Effects: 2SLS

	(1)	(2)	(3)
interaction $(\boldsymbol{t} \times \boldsymbol{D_j})$	0.00670**	0.00667**	0.0036***
	[2.540]	[2.509]	[2.583]
rating	0.436***	0.438***	0.430***
	[5.213]	[5.248]	[5.210]
favs	0.000170***	0.000174***	0.0002***
	[4.794]	[5.587]	[5.582]
comment	0.000138***	0.000141***	0.0001***
	[5.425]	[5.853]	[4.831]
sum_upload	-0.00957		-0.0198
	[-0.157]		[-0.334]
Observations	9060	9060	9060
R-squared	0.241	0.241	0.241

t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

Table 8 shows the results of the test. b_3 is the coefficient on $interaction$ $(t \times D_j)$. Column 1 and 2 represent different model specifications. We find that $\hat{b}_3$ is significantly positive under all specifications. The result suggests that social learning is not the only causal social contagion on YouTube, and network effects also play a critical role. The test confirms Hypothesis 3: the existence of network effects on YouTube. Column 3 shows that the results still hold when the surprise is defined as the difference between realized video ratings and predicted video ratings.

6. Application: How to Go Viral?

Social learning and network effects outline two ways that a video could go viral and gain success. Through examining the most popular videos on YouTube, we are able to categorize them into two distinct groups: the group consists of videos that feature high quality, engaging scenes, articulated story lines (high-quality videos), and the other group of videos often include questionable behaviors that deviate from social norms yet still gain tremendous popularity (attention grabbers). The recent "Pussy Riot" incident in Russia serves as a good example of a typical attention grabber. This Russia-based feminist rock band protested against the political scene in Russia through unorthodox musical performances and produced YouTube videos that went viral overnight. It is worth noting that Pussy Riot did not gain international fame through their musicality per se; instead, most viewers were drawn to those videos out of curiosity and were interested in the messages the music carried.

One type of strategy often adopted is the inclusion of controversial elements in videos. Such instances often provoke controversy and stir heated discussion revolving around those contents. This type of videos tends to attract critical reviews from both sides of the spectrum; viewers feel strongly and emotionally attached to the video in either extremely positive or negative ways. In contrast to those quality-oriented productions, the goal of attention grabbers is primarily to attract attentions or promote ideas. Intuitively speaking, we would not expect too much social learning effect to take place for the popularity of this type of video. In an analytical model, Eliaz and Spiegler (2011) shows that a firm can earn higher profits by employing pure attention grabbers with positive probability. Similarly, we propose that, as suggested by their discussion-provoking nature, videos with attention grabbing content can initiate higher network effects, and viewers find it valuable to watch them because these videos allow them to engage in discussions with their social contacts. Therefore, we hypothesize that this type of videos gains popularity mostly through network effects as opposed to social learning:

Hypothesis 4. *(a) Network effects are more pronounced for videos with attention grabbing content. (b) Social learning is more pronounced for high-quality videos.*

To test Hypothesis 4, we estimate the following two models:

$$\ln views_{jt} = b_0 + b_1 t + a_j + \Psi_{jt}' b_2 + b_3(t \times D_j)$$

$$+b_4(t \times attention_j) + b_5(t \times D_j \times attention_j) + \mu_{jt}, \qquad (7)$$

and

$$\ln views_{jt} = b'_0 + b'_1 t + a_j + \Psi_{jt}' b_2' + b'_3(t \times D_j)$$

$$+b'_4(t \times quality_j) + b'_5(t \times D_j \times quality_j) + \mu_{jt}, \qquad (8)$$

where $attention_j$ is a measure indicating whether video j is a video with attention grabbing content, and $quality_j$ is a measure indicating whether video j is a high-quality video. Here we use review rank and rating rank to empirically identify videos with high quality or attention grabbers. We define high-quality videos as videos with both large numbers of comments and high ratings, and attention grabbing videos as videos with high comment rank but mixed ratings. The co-existence of extremely high and extremely low ratings often suggests controversy. Specifically, if both the number of comments and the rating of video j at the end of our sample period rank among top 25% of total videos, then it is considered as a high-quality video, and the dummy $quality_j = 1$; otherwise, $quality_j = 0$. If the number of comments of video j at the end of our sample period rank among top 25%, but the rating is in the lowest 25%, then it is a video with controversial content, and $attention_j = 1$; otherwise, $attention_j = 0$.

Table 9. High-Quality Videos vs. Attention Grabbers

	(1)	(2)	(3)
$\boldsymbol{t \times D_j \times attention_j}$	0.0145***		
	[6.330]		
$\boldsymbol{t \times D_j \times quality_j}$		0.368**	-0.0208
		[2.014]	[-0.281]
rating	0.0307	0.0525	0.188
	[0.393]	[0.238]	[1.295]
favs	0.000145***	-0.000621*	0.00034***
	[4.416]	[-1.672]	[4.309]
comment	0.000140***	0.000853***	1.50e-06
	[5.821]	[10.21]	[0.0169]
sum_upload	0.0342	0.0715	-0.164
	[0.603]	[0.378]	[-1.254]
Observations	9060	9060	9060
R-squared	0.193	0.117	0.161

t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

To study the network effects for different videos, we estimate the regression model (7) using 2SLS. Similarly, we instrument the surprise dummy D_j using the in-stream ads ads_j. b_5 in equation (7) is the coefficient of interest. $b_5 > 0$ means the impact of network effects is larger for attention grabbing videos. Column 1 of Table 8 presents the regression results. We find that b_5 is significantly positive, supporting Hypothesis 4(a). Our result suggests that videos will be more likely to go viral through network effects if they provoke controversy and stir heated discussion. This is consistent with some experimental evidence: Content that evokes high-arousal emotions (i.e., awe, anger, and anxiety) is more viral (Berger and Milkman 2012). This finding can help YouTube providers craft contagious content and produce viral videos.

To study social learning, we first estimate regression model (8) without using an instrument variable. In Column 2 of Table 9, b'_5 is the coefficient on $t \times D_j \times quality_j$. We find that b'_5 is significantly positive, which suggests that social contagion is more pronounced for high-quality videos. However, it does not provide sufficient evidence for social learning because social contagion can be driven by network effects as well. Therefore, we re-estimate model (8) using ads_j as an instrument variable. The result is presented in Column 3 of Table 9. We find that $b'_5 = -0.0208$ and is not significant, which indicates that high-quality videos do not have higher network effects. Combining the results shown in Column 2 and 3, we can conclude that social learning is more pronounced for high-quality videos. In other words, Hypothesis 4(b) is also supported. Our empirical results of hypothesis testing provide supports for the strategic use of attention grabbers (Eliaz and Spiegler 2011).

7. Conclusions

Our paper uses a combination of analytical modeling and empirical estimation to study social learning and network effects in the context of social media. We find that social media content consumption is affected by both social learning network effects. A straightforward implication of our study is that YouTube should take social learning and network effects into account when fostering the growth of video views. Our results suggest that social contagion on YouTube is driven by both social learning and by network effects. As the influence of YouTube on our society, education, entertainment, and lifestyle increases, more and more organizations, including government agencies, TV networks, commercial companies, universities, etc., are all seeking for their presence and influence in social media. Our findings provide valuable insights as how to achieve this objective with videos on YouTube.

Considering that the amount of a traditional marketing campaign of YouTube content is limited, consumers rely heavily on advice from others to make decisions about watching videos. Social learning and network effects differ from traditional marketing activity in their social multiplier effect (Trusov et al. 2009). From a managerial perspective, YouTube can play a much greater role in encouraging the creation of original content because, as it nurtures and subsidizes individual content creators, the multiplier effect of both social learning and network effects extends their reach.[12] We also find the influence of social learning is stronger for unpopular videos, for videos in unpopular categories, and for videos with less precise priors.

Previous literature has focused on distinguishing social contagion from homophily, but only provides limited insights into how to disentangle social learning and network effects in the context of UGC. In this paper, we develop an empirical framework that allows us to make a causal inference about the presence of social learning and network effects on YouTube. More specifically, by applying a theoretical model, we examine the existence of social learning and network effects using a unique data set from YouTube. Our paper also makes methodological contributions for consistent and efficient estimation using noisy social media data.

Although in this study we only focus on social learning and network effects on UGC sites, our tests are relatively generalizable and can be practically carried out by practitioners in social media. We categorize the most popular videos on YouTube into quality-oriented videos and attention grabbing videos, and find that videos with attention grabbing content initiate higher network effects than quality-oriented productions. These findings provide a nuanced view of how YouTube providers can produce viral videos.

While this study has highlighted the importance of social learning and network effects, our work does not consider the effect of network characteristics and network topological structure on social contagion (Ghose et al. 2012). Further work could incorporate network data to examine the effect of network structure and tie strength on social learning and network effects. Another limitation of this study is that we only use data on top providers. Whether and to what extent the same patterns apply to unpopular providers need further examinations.

[12]In fact, YouTube is providing creators with resources and opportunities to improve their skills, build larger audiences, and make more money through its partnership program. As a New York Times article reported (Miller, 2011), a sketch comedy show called “AsKassem” is financed by grants from YouTube. The amount of content on YouTube covered by partnership agreements has risen steadily, to 10% of the total videos.

References

Acemoglu, D., Dahleh, M. A., Lobel, I., and Ozdaglar, A. (2011). "Bayesian Learning in Social Networks," *Review of Economic Studies*, (78:4), pp. 1201-1236.

Amemiya, T. 1973. "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica* (41:6), pp. 997–1016.

Anderson, S. P., and Gans J. S. 2011. "Platform Siphoning: Ad-Avoidance and Media Content," *American Economic Journal: Microeconomics* (3:4), pp. 1–34.

Aral, S., Muchnik, L. and Sundararajan A. 2009. "Distinguish Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences* (106:51), pp. 21544-21549.

Aral, S., and Walker D. 2011. "Creating Social Contagion through Viral ProductDesign: A Randomized Trial of Peer Influence in Networks," *Management Science* (57:9), pp. 1623-1639.

Banerjee, A. V. 1992. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, (107:3), pp. 797-817.

Bapna, R., and Umyarov A. 2012. "Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network," NBER working paper.

Barro, R. J. 1977. "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review*, (67:2), pp. 101-115.

Baillie, R. T. (1987). "Inference in Dynamic Models Containing Surprise Variables," *Journal of Econometrics*, (35:1), pp. 101-117.

Berger, J., and Milkman, K. L. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* (49:2), pp. 192-205.

Cai, H., Chen Y., and Fang H. 2009. "Observational Learning: Evidence from a Randomized Natural Field Experiment," *American Economic Review* (99:3), pp. 864-882.

Canay, I. A. 2011. "A Simple Approach to Quantile Regression for Panel Data," *Econometrics Journal* (14:3), pp. 368-386.

Chevalier J., and Mayzlin D. 2006. "The Effect of Word of Mouth on Sales: Online Book Review," *Journal of Marketing Research* (43:3), pp. 345-354.

Cameron, A. C., and Trivedi, P. K. 2005. *Microeconometrics: Methods and Applications*, Cambridge University Press.

Duan, W., Gu B., and Whinston A. B. 2009, "The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry," *Journal of Retailing* (84:2), pp. 233-242.

Eliaz, K., and Spiegler, R. 2011, "On the Strategic Use of Attention Grabbers," *Theoretical Economics* (6:1), pp. 127-155.

Godes, D. and Mayzlin D. 2009. "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science* (28:4), pp. 721-739.

Goldenberg, J., Han, S., Lehmann, D. R., and Hong, J. W. 2009. "The Role of Hubs in the Adoption Process," *Journal of Marketing* (73:2), pp. 1-13.

Goldenberg, J., Libai, B., and Muller, E. 2010. "The Chilling Effects of Network Externalities," *International Journal of Research in Marketing* (27:1), pp. 4-15.

Goldenberg, J., Oestreicher-Singer, G., and Reichman, S. 2012. "The Quest for Content: The Integration of Product Networks and Social Networks in Online Content Exploration," *Journal of Marketing Research* (49:4), pp. 452-468.

Ghose, A., Han, S., and Iyengar, R. 2012. "An Empirical Analysis of Network Characteristics and Social Contagion on the Mobile Internet," working paper.

Ho, T., Li, S., Park, S. E., and Shen, M. Z. 2012. "Customer Influence Value and Purchase Acceleration in New Product Diffusion," *Marketing Science* (31:2), pp. 236-256.

Hirshleifer, D., Lim, S. S., and Teoh, S. H. 2009. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News," *Journal of Finance*, (64:5), pp. 2289-2325.

Iyengar, R., Van den Bulte, C., and Choi, J. 2012. "Distinguishing among Mechanisms of Social Contagion in New Product Adoption: Framework and Illustration," working Paper.

Iyengar, R., Van den Bulte, C., and Valente, T. W. 2011. "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science* (30:2), pp. 195-212.

Jegadeesh, N. and Titman, S. 1993. "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance* (48:1), pp. 65-91.

Katona, Z., Zubcsek, P. P., and Sarvary M. 2011. "Network Effects and Personal Influences: The Diffusion of an Online Social Network," *Journal of Marketing Research* (48:3), pp. 425-443.

Katz, M. L., and Shapiro, C. 1985. "Network Externalities, Competition and Compatibility," *American Economic Review* (75:3), pp. 424–440.

Koenker, R., and Hallock, K. 2001. "Quantile Regression: An Introduction," *Journal of Economic Perspectives* (15:4), pp. 43-56.

Manski, C. F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* (60:3), pp. 531-542.

Moretti, E. 2011, "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales," *Review of Economic Studies*, (78:1), pp. 356-393.

Miller, C. 2011. "Actors in Smaller Studios, Making Pictures for the Smaller Screen," *New York Times*, (accessed April 10, 2011), [available at http://www.nytimes.com/2011/04/11/business/media/11youtube.html?_r=1&pagewanted=all].

Pagan, A. 1984. "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, (25:1), pp. 221-247.

Qiu, L., and Whinston A. B. 2012. "Naive Herding in Location-Based Networks: Induced Social Learning and Optimal Dynamic Pricing," working paper, Available at SSRN: http://ssrn.com/abstract=2090093.

Sargent, T. J. 1976. "A Classical Macroeconometric Model for the United States," *Journal of Political Economy* (84:2), pp. 207-237.

Shi, Z., and Whinston, A. B. 2013. "Network Structure and Observational Learning: Evidence from a Location-Based Social Network," Forthcoming at *Journal of Management Information Systems*.

Shi, Z., Rui, H., and Whinston, A. B. 2013. "Content Sharing in a Social Broadcasting Environment: Evidences from Twitter," Forthcoming at *MIS Quarterly*.

Simon, H. A. 1982. "Designing organizations for an information-rich world," in *Computers, Communications, and the Public Interest*, M. Greenberger (ed.), John Hopkins Press, Baltimore, 1971, pp. 38–52.

Stock, J. H., Wright, J. H., and Yogo, M. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics* (20:4), pp. 518-529

Susarla, A., Oh, J. and Tan, Y. 2012. "Social Networks and the Diffusion of User-Generated

Content: Evidence from YouTube," *Information Systems Research* (23:1), pp. 23-41.
Tang, Q., Gu, B., and Whinston, A. B. 2012. "Content Contribution for Revenue Sharing and Reputation in Social Media: A Dynamic Structural Model," *Journal of Management Information Systems* (29:2), pp. 41-75.
Trusov, M., Bodapati, A. V. and Bucklin, R. E. (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research* (47:4), pp. 643-658.
Trusov, M., Bucklin, R. E. and Pauwels K. (2009), "Estimating the Dynamic Effects of Online Word-of-Mouth on Member Growth of a Social Network Site," *Journal of Marketing* (73:5), pp. 90–102.
Tucker, C., and Zhang, J. 2010. "Growing Two-Sided Networks by Advertising the User Base: A field Experiment," *Marketing Science* (29:5), pp. 805-814.
Tucker, C., and Zhang, J. 2011. "How Does Popularity Information Affect Choices? A field Experiment" *Management Science* (57:5), pp. 828-842.
Van Den Bulte, C. and Stremersch S. 2004. "Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test," *Marketing Science* (23:4), pp. 530-544.
Wilbur, K. C. 2008. "A Two-Sided Empirical Model of Television Advertising and Viewing Markets," *Marketing Science* (27:3), pp. 356-378.
Wooldridge, J. M. 2007. "What's New in Econometrics? Lecture 10 Difference-in-Differences Estimation," NBER Summer Institute, mimeo.
Wooldridge, J.M. 2002. *Econometrics Analysis of Cross Section and Panel Data*, MIT Press: Cambridge, MA.