Spring 5-29-2015

Analyzing the Co-evolution of Network Structure and Content Generation in Online Social Networks

Prasanta Bhattacharya
National University of Singapore, prasanta@comp.nus.edu.sg

Tuan Q. Phan
National University of Singapore, tuan.quang.phan@gmail.com

Edoardo M. Airoldi
Harvard University, airoldi@fas.harvard.edu

Recommended Citation

Bhattacharya, Prasanta; Phan, Tuan Q.; and Airoldi, Edoardo M., "Analyzing the Co-evolution of Network Structure and Content Generation in Online Social Networks" (2015). *ECIS 2015 Completed Research Papers.* Paper 18.
ISBN 978-3-00-050284-2
http://aisel.aisnet.org/ecis2015_cr/18

ANALYZING THE CO-EVOLUTION OF NETWORK STRUCTURE AND CONTENT GENERATION IN ONLINE SOCIAL NETWORKS

Complete Research

Bhattacharya, Prasanta, National University of Singapore, Singapore 117417, prasanta@comp.nus.edu.sg

Phan, Tuan Q., National University of Singapore, Singapore 117417, phantq@comp.nus.edu.sg

Airoldi, Edoardo, Harvard University, Cambridge, MA 02138, USA, airoldi@fas.harvard.edu

Abstract

With the rapid growth of online social network sites (SNS), it has become imperative to investigate what drives content production on these platforms. We posit that the content producing behavior of users is influenced not just by their personal attributes like age and gender, but also by their social network structure. However, it is empirically challenging to estimate network structure and behavior through traditional approaches as the social network structure and the content production behavior influence the evolution of each other. In the current study, we adapt an actor-based continuous-time model to jointly estimate the co-evolution of the users' social network structure and their content production behavior using a Markov Chain Monte Carlo (MCMC) based simulation approach. We apply our model to an online social network of university students and uncover strong evidence for both social influence and homophilous friend selection. Interestingly, we find that individuals befriend others who are similar in content production during the friendship formation stage, but gradually diverge in their content production behavior from these similar others over time. We offer potential explanations for this phenomenon and emphasize the importance of these findings for platform owners and product marketers.

Keywords: Social network structure, content production, Co-evolution model, Homophily, Social influence, Markov Chain Monte Carlo, Method of Moments.

1 Introduction

With the proliferation of Social Network Sites (SNS), platform owners are facing increasing challenges to engage users, and, subsequently, sell advertisement impressions (Edwards, 2012; Holf, 2011; Tucker, 2012). Unlike other Internet-based services, the unique value of SNS lies in managing the interactions, using a sort of two-sided network between user roles – "content producers" who actively post, comment and share content with their friends, and "content consumers" who view and react to those content. Although content producers can add considerable value through network effects by sharing content for their peers to consume, the motivation and mechanisms employed by these "valuable" users to post is less understood.

We posit that the content production behavior of users on social media is influenced partly by individual level factors (e.g. demographics) and partly by their online social network characteristics viz. the number of online friends, their network clustering, their network betweenness and so forth (Newman, 2010). While it is fairly straightforward to establish the relationship between demographic features and content production due to the exogenous nature of these variables, it remains an empirical puzzle to estimate the influence of the network structure on the user's content production behavior. This is because the user's behavior and social network co-evolve by influencing each other i.e. behavior shapes network at the same time that the network shapes behavior. From a classical social networks perspective, solving this puzzle amounts to successfully separating the effect of social influence (i.e. when network influences attitude/behavior) from homophilous interactions (i.e. when attitude/behavior influences network formation) and context effects (Borgati and Foster, 2003; McPherson et al., 2001). In defining network structure variables as mentioned above, we rely on structural relationships that are defined and declared by the actors themselves on the social network platform. An example of this could be the "friendship" that is initiated and accepted by the actors on SNS (Backstrom et al., 2006) Such friendship-based networks have been used extensively in studying peer-effects in social networks, as detailed next.

We realize that it is imperative to separate the effects of homophilous friendship formation from social influence to correctly estimate the effects of the social network on an individual's propensity to produce content. In the absence of a statistically disciplined empirical approach to performing this, there remains a theoretical gap in our understanding of how network structure influences content production, particularly in online contexts. While there have been a couple of previous empirical approaches at addressing this problem, they suffer from serious shortcomings. Experimental approaches to addressing this problem usually estimate either influence or homophily while controlling for the other (Toubia and Stephen, 2013; Sacerdote, 2001). They also suffer from low ecological validity e.g. it is hard to imagine real-world situations where friendship formation would be truly exogenous. Non-experimental approaches have employed either a contingency table method (Fisher and Bauman, 1988; Kandel, 1978), an aggregated personal network based method (Yoganarasimhan, 2012; Kirke, 2004), or structural equation models (Krohn et al., 1996; Iannotti et al., 1996) to try and address this problem. However, most of these studies suffer from one of the following three limitations. First, they tend to ignore the network dependence of actors i.e. their dyadic independence assumption is often violated in real networks. Second, they fail to control for competing mechanisms like the shared contexts etc. Lastly, the studies do not take into account the possibility of errors introduced due to incomplete observations, as is common with discrete time models.

In our study, we improve over previous approaches by leveraging an actor-based and continuous-time computational modeling approach as introduced by Snijders (2001) and Snijders et al. (2007). We adapt this approach to the online setting by jointly modelling the evolution of the online social network and the evolution of online behavior of the social network actors. The online behavior we focus on in the current study is the content production behavior of the actors. We posit that actors produce publicly-visible content on SNS that might influence their propensity to make new friends on the SNS. The interrelationship between the volume of public content produced and the associated peer-effects has been the focus of several recent studies (Zhang and Zhu, 2001; Toubia and Stephen, 2013).

The likelihood function in such models is often too complex to be computed explicitly, making maximum likelihood or Bayesian estimation methods difficult to use. However, we resort to simulation based estimation approaches supported within the general framework of Markov chain Monte Carlo (MCMC) estimations. For the present study, we use a MCMC based Method of Moments (MoM) estimator to estimate the co-evolution parameters in our model (Snijders, 2001). While some prior studies have used computational simulations to model endogenous evolution of network ties and individual attributes (Carley, 1991; Macy et al., 2003), our current model allows for statistical inference testing, model fit assessments and counterfactual simulations. Moreover, it is flexible enough to allow for different forms of objective functions and operates under an acceptable set of assumptions (e.g., conditional independence etc.).

From the subsequent analyses, we find strong evidence for homophily based on similarity in content production behaviour, but no evidence of homophily on the basis of individual covariates, like age or gender. We also observe the existence of social-influence , but in a direction opposite to that of homophilous interaction. This is an interesting insight about the opposing roles of behavioral similarity at different stages of friendship formation. We show that individuals befriend others who are similar in content production during the friendship formation stage, but gradually diverge in content production behavior from these similar others over time. We offer potential explanations for this phenomenon. Our findings contribute to several related streams of literature that looks at (i) factors that contribute to content production on social network sites (SNS), (ii) understanding the interplay between different types of peer-effects on social network sites, and (iii) understanding the co-evolution of network structure and human behavior in offline and online contexts.

In the following section, we offer a brief summary of the coevolution model that we use in our empirical analyses. Next, we discuss our empirical setting and demonstrate our findings. We conclude with a discussion on the key contributions of our study, the limitations and a roadmap for future research.

2 Co-evolution of Networks and Behavior

In this study, we leverage the actor-based co-evolution modelling approach used in Snijders et al. (2007) and Steglich et al. (2010). This network-behavior co-evolution model draws upon his past work on actor-driven pure network evolution models (Snijders, 2001). In the current co-evolution setting, the two main variables modelled are the state of the time-varying friendship network $\mathrm{X_{ijt}}$, and a time-varying integer-valued behavior vector $\mathrm{Z_{iht}}$ which comprises the values of actor *i* , alter *j* on behavioral attribute *h* at time *t*. The evolution of both the network as well as the behavior is assumed to follow a first-order Markov process, using very small time-increments, called “micro-steps.” The network evolves in continuous-time but is observed at discrete moments. At a given micro-step, the network or the behavior in constrained to make only a unit change i.e. a tie forms or dissolves, or the behavioral attribute gains or loses strength by 1 unit. The instances when any given actor *i* gets the opportunity to make a decision to change the vector of outgoing tie variables ($\mathrm{X_{i1}}, \ldots, \mathrm{X_{in}}$) or a behavior variable Z_{ih} are randomly determined and follow a Poisson distribution. Consequently, the waiting times for network decisions and behavioral decisions between time periods t_m and t_{m+1} are modelled by exponential distributions with parameters decided by rate functions as given in Eqs. 1 and 2. Here, the variable *Y* denotes the joint configuration of the user’s network *X* and her behavior *Z*.

$$\lambda_i^{[X]}(Y,m) = \rho_m^{[X]} exp^{(\Sigma_k \alpha_k^{[X]} a[X]_{ki}(Y(t)))} \quad \text{(network decisions)} \tag{1}$$

$$\lambda_i^{[Z_h]}(Y,m) = \rho_m^{[Z_h]} exp^{(\Sigma_k \alpha_k^{[Z_h]} a[Z_h]_{ki}(Y(t)))} \quad \text{(behavioral decisions)} \tag{2}$$

where, parameters ρ_{m} indicate period-dependence and α indicates dependence on the statistics $\alpha_{\mathrm{ki}}(\mathrm{Y(t)})$.

While the rate functions model the timing of the actors’ decisions (i.e. to change network or behavior), the objective functions model the specific changes that are made. It is assumed that actors *i* myopically optimize an objective function over the set of possible micro-steps they can take. This objective function is composed of three parts: the evaluation function *f*, the endowment function *g*, and a random term ϵ capturing residual noise:

$$f_i^{[X]}(\beta^{[X]}, y) + g_i^{[X]}\left(\gamma^{[X]}, y \middle| Y(t)\right) + \epsilon_i^{[X]}(y) \quad \text{(network decisions)} \tag{3}$$

$$f_i^{[z_h]}(\beta^{[z_h]}, y) + g_i^{[z_h]}\left(\gamma^{[z_h]}, y \middle| Y(t)\right) + \epsilon_i^{[z_h]}(y) \quad \text{(behavioral decisions)} \tag{4}$$

The network evaluation function can be modeled as a weighted sum of various network characteristic (e.g. degree, transitivity etc.) and behavioral characteristics (e.g. similarity measure, non-linear behavior trends etc.). These are illustrated through the following expressions for $f_i^{[X]}(\beta^{[X]}, y)$ and $f_i^{[z_h]}(\beta^{[z_h]}, y)$.

$$f_i^{[X]}(\beta^{[X]}, y) = \sum_k \beta_k^{[X]} s_{ik}^{[X]}(y) \quad \text{(network evaluations)} \tag{5}$$

$$f_i^{[z_h]}(\beta^{[z_h]}, y) = \sum_k \beta_k^{[z_h]} s_{ik}^{[z_h]}(y) \quad \text{(behavior evaluations)} \tag{6}$$

In addition to the above evaluation functions, the network endowment effects $g_i^{[X]}$ and $g_i^{[z_h]}$ are entered into the objective function to assess systematic differences between the creation and the dissolution of ties that cannot be captured by the evaluation function alone. For instance, the cost of losing a reciprocal friendship tie is greater than the gain in establishing such a tie. Psychological theories of loss aversion support the existence of such an endowment effect (Kahneman et al., 1991; Thaler, 1980). For the present analyses, however, we assume that there are no endowment effects at play, and the evaluation functions are sufficient to explain user decisions on network and behaviour. As mentioned earlier in the paper, the likelihood function for more complex model specifications is extremely difficult to compute explicitly which mandates the use of simulation based estimators. We use a Markov Chain Monte Carlo (MCMC) based Method-of-Moments (MoM) estimator to recover the parameters of these rate and evaluation functions. The MCMC implementation of the MoM estimator uses a stochastic approximation algorithm which is a variant of the Robbins-Monro (1951) algorithm.

3 Empirical Analyses

We leverage the co-evolution approach as detailed in the previous section to investigate the role of the network structure on an individual's content production behavior. The posting behavior of an individual in a future period would be assumed to follow a continuous-time Markov process, as a function of time and the present "state", where a *state* is defined to be a combination of her sociometric variables (e.g. number of friends, network transitivity, closeness etc.) and her behavioral attributes (e.g. measure of content produced e.g. number public of posts produced by the actor etc.). Additionally, we also plan to accommodate time varying and time-invariant covariates in our model (e.g. age, gender etc.) to control for some competing motivations to form friendships and produce content.

3.1 Data Context

We use complete online network data from a large and popular social network site for 2507 undergraduate students belonging to a North American University for the months from September 2008 till February 2009. Additionally, we record the number of public posts made by these users on the social media platform during the said period. It is important to note here that the public posts contributed by the users are visible to all other user on the platform, and not just to the user's friends. The network and behavior descriptive summaries are detailed in Appendices 1 and 2. Within our observation period, the students produced a substantial amount of content on the social media platform, and also established several new friendships. By exploiting the temporal sequentiality of these actions (e.g. friendship -> behaviour or behaviour -> friendship) from our longitudinal data, we are able to make restricted claims about casuality of peer effects.

3.2 Results

In this section, we report some initial findings from adapting the co-evolution model to our data context. For the current study, we ignore any endowment effects in our model and assume a constant rate function across all the 6 periods. A constant rate function implies that all users receive equal number of opportunities to make a network or behavior change. However, the number of actual changes made by the users would still differ. The model specifications for the network rate and evaluation functions are illustrated in Eq. 7 and 8 below.

$$\lambda_i^{[X]}(Y,m) = \rho_m^{[X]} exp^{(\sum_k \alpha_k^{[X]} a[X]_{ki}(Y(t)))} \tag{7}$$

$$f_i^{[X]}(\beta^{[X]},x,z) = \sum_{k=1}^{9} \beta_k^{[X]} s_{ik}^{[X]}(x,z) \tag{8}$$

In (8) above, the nine network effects that we model in our study are the actor's out-degree, the transitivity, homophily effects based on gender, age, social network tenure, homophily effects based on posting behavior, and the influence of individual covariates (e.g. gender, age, social network site (SNS) tenure) and posting behavior on the propensity to form new friends. The mathematical illustrations are provided in Eqs. (9) through (12).

(i) Degree(s_{i1}) and Transitivity(s_{i2})

$$s_{i1}^{[X]}(x) = \sum_j x_{ij} \tag{9}$$

$$s_{i2}^{[X]}(x) = \sum_{j,h} x_{ij} * x_{jh} * x_{ih} \tag{10}$$

(ii) Homophily based on attributes (age, gender, social network tenure) and posting behaviour

$$s_{i3}^{[X]}(x,z) = x_{i+}^{-1} \sum_j x_{ij} \left(1 - \frac{|z_{ih} - z_{jh}|}{R_h}\right) where, R_h \text{ is the range of the variable } Z_h \tag{11}$$

s_{i3} represents the effect for age related homophily. For Z_h = {gender, SNS tenure, posting behavior}, we have exactly similar expressions for s_{i5}, s_{i7} and s_{i9} respectively.

(iii) Behavior/Covariate (V_j) on Degree effects (effect of age, gender, SNS tenure, posting on Degree)

$$s_{i4}^{[X]}(x,z) = \sum_j x_{ij} * v_j \tag{12}$$

s_{i4} represents the effect of age on degree. For V_j = {gender, SNS tenure}, we have exactly similar expressions for s_{i6} and s_{i8} respectively. In all the above equations, $x_{ij} = 1$ if a tie exists between i and j, and 0 otherwise. Similarly, the model specifications for the behavior rate and evaluation functions are illustrated in Eq. 13 and 14 below.

$$\lambda_i^{[Z_h]}(Y,m) = \rho_m^{[Z_h]} exp^{(\sum_k \alpha_k^{[Z_h]} a[Z_h]_{ki}(Y(t)))} \tag{13}$$

$$f_i^{[Z_h]}\left(\beta^{[Z_h]}, x, z\right) = \sum_{k=1}^{5} \beta_k^{[Z_h]} s_{ik}^{[Z_h]}(x, z) \tag{14}$$

In (14) above, the five behavior effects that we model are the actor's behavior tendency effect, the network assimilation effect (i.e. social influence), and effects that capture the influence of individual covariates like age, gender and SNS tenure on the posting behavior, Z_i. The mathematical illustrations are provided in (15) through (17).

(i) Behavioral tendency effect (This captures the tendency of users to increase or decrease behavior over time)

$$s_{i1}^{[Z_h]}(x, z) = z_i \tag{15}$$

(ii) Network Assimilation Effect (The propensity of users to assimilate in behavior towards their peers)

$$s_{i2}^{[X]}(x, z) = x_{i+}^{-1} \sum_j x_{ij} \left(1 - \frac{|z_{ih} - z_{jh}|}{R_h}\right) \text{ where, } R_h \text{ is the range of the variable } Z_h \tag{16}$$

(iii) Influence of covariates on behavior

$$s_{i3}^{[Z_h]}(x, z) = z_i * v_i \tag{17}$$

Here, s_{i3} represents the effect of gender on posting behaviour. Similar expressions can be constructed for s_{i4} and s_{i5} to represent the effects of age and SNS tenure on posting behaviour respectively. In all the above expressions, z_i denotes the number of posts for the user *i*.

It is clear from the above formulation of effects, that the mathematical illustration for the network and behaviour effects to compute homophily (Eq. 11) and social influence (Eq. 16) are identical. This point lies at the core of the problem that is separating the effect of homophilous selection from social influence. However, we exploit the longitudinal nature of our dataset to successfully identify temporal sequentiality across the periods. In other words, we use dyads of users who first become friends and then assimilate in behavior, to identify assimilation. Similarly, we use dyads of users who show similarity in behaviour before becoming friends, to identify homophily. While there might be other latent confounds that we do not capture in our modelling, our present approach makes an attempt at demonstrating a restricted form of causality. This view is consistent with several recent studies investigating related topics on homophily and influence among student populations (Lewis et al., 2012, Steglich et al., 2010).

We estimate the above rate and evaluation functions using a Method of Moments (MoM) estimator and the results are presented in Table 1 in the following page. The MoM estimator essentially tries to recover parameter estimates by matching the observed network data with the simulated network data. Appendices 3 and 4 provide details on the convergence descriptives for these simulations. Specifically, we provide information about the deviation of our simulated network and behavioral statistics from the observed data. Tables 1(a) and 1(b) highlight the estimation results for rate parameters ρ_m^{net} and ρ_m^{beh} ,where *m* ranges from periods 1 to 5 (i.e. the first among six periods is conditioned upon during the estimation), and estimates for $\beta_p^{[X]}$ where p ranges from 1 to 9 and for $\beta_q^{[Z_h]}$ where q ranges from 1 to 5.

Network Parameters	**Estimate**
Friendship rate (Period 1)	7.767*** (0.100)
Friendship rate (Period 2)	6.267*** (0.088)
Friendship rate (Period 3)	3.930*** (0.082)
Friendship rate (Period 4)	4.547*** (0.075)
Friendship rate (Period 5)	5.353*** (0.084)
Out-Degree	-9.536*** (0.011)
Transitivity	0.109*** (0.001)
Gender homophily	0.068 (0.057)
Gender on Degree	0.031 (0.020)
Age homophily	0.024 (0.034)
Age on degree	0.011 (0.009)
Tenure homophily	0.003 (0.022)
Tenure on degree	-0.032** (0.016)
Posting homophily	**0.127*** (0.034)**

Behavior Parameters	**Estimate***
Posting rate (Period 1)	4.337*** (0.145)
Posting rate (Period 2)	3.889*** (0.143)
Posting rate (Period 3)	5.229*** (0.205)
Posting rate (Period 4)	4.995*** (0.195)
Posting rate (Period 5)	3.681*** (0.119)
Posting Tendency (Linear Shape)	-0.196*** (0.007)
Assimilation	**-2.995*** (0.134)**
Gender on Posting	0.007 (0.012)
Age on Posting	0.003 (0.006)
Tenure on Posting	0.003 (0.010)

*** <0.01, ** <0.05, *<0.1

Table 1(a), 1(b). Estimation results for network and behaviour parameters

From the results in Tables 1(a) and 1(b), we observe that the estimate for the out-degree of the actor is significantly negative. Since, the evaluation function can be thought of as a measure of the "fitness" or "attractiveness" of the state of the network, this estimate indicates that actors in our network prefer to not establish more social connections. This can be explained by the fact that forming social connections involve a certain social cost, and rational actors would refrain from undertaking this cost unless it is over-compensated by the social benefits accrued (e.g. reciprocity or heightened social status). A recent study has found that the out-degree does not significantly change amongst Facebook users who face traumatic experience (in their case, a natural disaster from Hurricane Ike) as compared to a control group (Phan and Airoldi, 2015). This provides supporting evidence that online networks may be stable and exhibit an, albeit, small social cost to maintain relationships and create new ones. In addition to this, it is important to recall that our sample comprises University students, who follow a certain pattern in forming friendships i.e. more towards the beginning of an academic year, and less afterwards. This might further contribute to the negative estimate. Further, we observe that the estimate for transitivity is positive. This indicates that there is an increased drive towards network closure in our network. For instance, if actors *i* and *j* are friends, and actors *j* and *h* are friends as well, then the actor *i* has a stronger motivation to befriend actor *h* than any other actor in the network, as this increases the overall attractiveness of the new network state for *i*. We also find strong evidence for homophilous friendship formation based on similarity in posting rates. This is consistent with past research that emphasizes the role of homophily based on stable human-attributes or tastes (Lewis et al., 2012; McPherson et al., 2001). However, unlike these prior studies that look at relatively stable human attributes (e.g. smoking, music tastes etc.), the current work focuses on a dynamic user behavior and demonstrates evidence of homophily in such type of dynamic behaviors too. Our results show that the more active posters prefer to befriend other active posters, while the less active posters prefer other less active ones. Interestingly, none of the other covariates were found to contribute to homophilous friendship formation.

Among the behavior variables, we observe that the estimate for the linear tendency parameter is significantly negative. As mentioned earlier, the tendency effect represents a drive towards high posting volume. A zero value on this parameter indicates actor's preference for the average posting volume. Since we obtain a negative estimate on this parameter, it indicates that users prefer to post lesser number of posts as time goes on. We also find strong evidence of social influence among the students, with a significantly negative parameter for the network assimilation effect. This implies that individuals tend to correct their posting behaviour over time in a direction away from their peers. This could be a result of free-riding behaviour in case the peers are contributing more, or could also be representative of an increased drive to behave in a non-conformist manner (e.g. "*If everyone else is posting more, I should do something different*"). This presence of social influence is consistent with previous studies that have argued in favor of social influence in online and offline social networks (Toubia and Stephen, 2013; Sacerdote, 2001).

While it is hard to uncover the specific reasons for the peer-effects we find, interpreting the parameters for homophily and network assimilation together leads to an interesting observation. Taken together, the two parameters suggest that while students prefer to befriend other students who are similar to themselves in posting behavior, they tend to move apart over time after becoming friends. Thus, behavioral similarity could play the role of a facilitator during the early days of friendship formation, but act as a deterrent in the longer run. We contend that this insight is not only theoretically important to uncover but has very strong practical implications as well, which we shall discuss in the following section.

4 Study Contributions

In the current study, we demonstrate the role of social network structure and user-characteristics in influencing content production on social media platforms. We adopt a co-evolution based modeling approach to jointly estimate the evolution of the user's social network and behavior. We contend that this approach is statistically more disciplined than several of the prior methods that tend to violate some key assumptions of network-based modelling. Our analyses suggest an interesting interplay between the two mechanisms of homophilous peer selection and peer-influence. Even though previous studies have established the existence of peer-influence and homophily in social networks, our study goes a step further to illustrate that both these effects can co-exist and lead to different network dynamics. For instance, we find no evidence that students from a North American University make friendships based on age and gender based similarities. However, we find strong evidence of homophilous selection on the basis of content production i.e. the students make friends with others who are similar in their content production behavior. Once they become friends, however, our findings show that they exhibit a negative assimilation effect. This means that the students actively try to distinguish themselves from their peers in terms of their content production behavior. This is an interesting phenomenon as it shows that dynamic behaviors like content production can influence the network evolution in competing ways. While content producers seek to make friends with similar others, they move apart in similarity once they become friends. There could be several theoretical explanations for the reverse assimilation effect. If content produced on social media can be thought of as a public good (Zhang and Zhu, 2011; Andreoni, 1988), this could be a result of a free-riding behavior on part of the users. With an increase in content production in the group, users reduce their own contribution and vice versa. Alternately, this result could also represent an increased drive on part of the social network users to carve out their own identity, and exhibit a non-conformist behavior (e.g. "..if everyone else is posting, it is not cool anymore"). While it is tenuous to try and separate out the specific mechanisms using a non-experimental setting such as ours, we hope to rule out the major confounds in future versions of this work. In the current study, we have already controlled for common covariates that might influence friendship formation and subsequent behavior i.e. user's age, gender and SNS tenure.

In addition to uncovering an interesting interplay between homophily and social influence, we contend that ours is the first study that looks at the effect of peer effects on dynamic behaviors like content production on online platforms. Most prior work in the area of offline and online peer-effects have restricted themselves to relatively stable or slow-moving behavioral attributes like smoking behavior, substance abuse or taste in music and movies. However, the results from the current study show that such peer effects also exist and influence the evolution of dynamic human behaviors. Using the insights from our study, and the modeling approach in general, researchers can now potentially investigate a wide array of dynamic behaviors on online social networks (e.g. ad clicking behavior, self-presentation and impression management, privacy-consciousness etc.).

Uncovering peer-effects on social networks has strong practical implications. Our results can enable platform owners to identify and better target valuable users on their sites. Moreover, by understanding how friendships are made and maintained over time, social network sites like Facebook and Twitter can help improve friend recommendations and personalize content through customized "newsfeeds". Our model also allows for predictive analysis of posting behavior on these platforms, such that managers and researchers can effectively seed content, and forecast the diffusion of this content through the user's social network. This is invaluable not just for the platform owners, but also to advertisers and third-party marketers who leverage the social media users and data for their own businesses. Thus, we believe that the findings from our study, and the methodology in general can be used to improve user-satisfaction on online platforms, and increase retention and value-creation.

5 Limitations and Future Work

Our study in its current form has a number of limitations that we are working towards mitigating. Firstly, and as mentioned earlier, the current paper focuses on providing a statistically sound method to uncover the dynamic peer-effects in an university social network. However, additional analyses are required to further separate out the specific rationale behind why individuals show such effects. Secondly, our current modelling approach is assumption-intensive and requires a significantly large investment of computational resources to simulate the networks in each stage of the estimation procedure. This might be a concern for extremely large networks of users, and networks with high sparsity. In such cases, we might have to resort to bootstrapping approaches which introduces concerns about network-based sampling, another non-trivial research area. Thirdly, the model imposes a standard Markovian assumption on the data, which is reasonable in most cases. However, this assumption implies that there are no latent confounding factors that might influence the social network or the user behavior. Even though we have controlled for the most frequently reported covariates that have been used in past social network studies (e.g. age, gender and experience), we cannot completely rule out the possibility of unobserved confounds that might play a role. However, we contend that even with the existence of such potential confounds, our method provides a more theoretically and statistically disciplined approach to studying peer effects, than what has been offered by previous studies. However, it needs to be mentioned here that all claims about causality in the current work is essentially restricted in light of such potential confounds.

We plan to extend this current study in two major directions. First, we wish to investigate the observed peer-effects deeper by analysing whether these effects differ based on the posting behavior of the user. For example, we wish to understand if low posters are more or less susceptible to peer influence than high posters. If generalized, the results from such analyses can pose strong implications for policy interventions (e.g. “at what level of smoking addiction should we expose a person to a non-smoking ad?”). Second, in the present study, we consider all friendships to be bi-directional or symmetric ties. However, it will be useful to identify the directionality of friendship i.e. separate out in-degree from out-degree. While in-degree can be considered to be a measure of popularity, out-degree provides a measure of activeness. Thus, by separating out the two effects, we will be able to investigate more complex social constructs.

Appendix 1

1(a) : Descriptive summary for social network data

	Time Period					
Observation	**1**	**2**	**3**	**4**	**5**	**6**
Density	0.025	0.027	0.028	0.029	0.030	0.031
Average Degree *	63.276	67.165	70.377	72.42	74.844	77.747
Number of Ties	79317	84191	88217	90778	93817	97456
Missing Fraction	0	0	0	0	0	0

*** Average degree across all periods = 70.971**

1(b) : Social network evolution summary

	Change in Ties					
Period	**0 => 0**	**0 => 1**	**1 => 0**	**1 => 1**	**Jaccard ***	**Missing**
1==>2	3057080	4874	0	79317	0.942	0 (0%)
2==>3	3053054	4026	0	84191	0.954	0 (0%)
3==>4	3050493	2561	0	88217	0.972	0 (0%)
4==>5	3047454	3039	0	90778	0.968	0 (0%)
5==>6	3043815	3639	0	93817	0.963	0 (0%)

* Jaccard Index = $\frac{N_{11}}{N_{01}+N_{10}+N_{11}}$, where N_{hk} is the number of tie variables with value h in one wave, or observation from our dataset, and the value k in the next wave.

Appendix 2

2(a) Descriptive summary for behavior data

	Time Period					
Posting quantile	**1**	**2**	**3**	**4**	**5**	**6**
1 (lowest)	630	763	787	644	806	774
2	945	1027	1019	903	1015	1009
3	364	317	324	326	325	336
4	193	177	157	223	147	147
5	122	86	76	118	85	88
6	77	47	41	84	53	54
7	33	24	37	50	18	30
8 (highest)	26	18	19	45	18	20

Note: The figures in the cells indicate the number of users who have posted in that time period. Row 1 indicates the total number of first-quantile posters (i.e. low posters) in each of the 6 time periods. Similarly, Column 1 indicates the number of posters in each of the 8 posting quantiles for the first time period.

2(b) Behavior evolution summary

	Number of users			
Period	**Decrease Posting Behavior**	**Increase Posting Behavior**	**Constant**	**Missing**
1 => 2	1009	427	1071	0
2 => 3	674	653	1180	0
3 => 4	378	1057	1072	0
4 => 5	1066	367	1074	0
5 => 6	625	711	1171	0

Appendix 3

Convergence Assessment for Network Variables

Network Variables	Observed Value for Target Statistics	Av. Deviation of simulated statistic from target statistic (SD Deviation)
Friendship rate (Period 1)	9748.000	-370.044 (139.914)
Friendship rate (Period 2)	8052.000	-141.393 (126.777)
Friendship rate (Period 3)	5122.000	24.940 (97.883)
Friendship rate (Period 4)	6078.000	63.709 (108.249)
Friendship rate (Period 5)	7278.000	254.976 (120.326)
Out-Degree	454459.000	-83.906 (131.380)
Transitivity (No. of triads)	4445064.000	-2548.047 (3834.512)
Gender on Degree	11547.242	-76.155 (108.814)
Gender homophily	-2109.296	-37.071 (37.992)
Age on degree	-29955.212	-209.697 (276.408)
Age homophily	-6381.205	-45.223 (68.114)
Tenure on degree	19104.656	186.366 (157.918)
Tenure homophily	-1968.574	12.749 (98.977)
Posting homophily	22210.253	-199.796 (80.153)

Appendix 4

Convergence Assessment for Behavior Variables

Behavior Variables	**Observed Value for Target Statistics**	**Av. Deviation of simulated statistic from target statistic (SD Deviation)**
Posting rate (Period 1)	2150.000	-33.206 (47.848)
Posting rate (Period 2)	1663.000	-25.970 (46.780)
Posting rate (Period 3)	2359.000	-89.171 (51.167)
Posting rate (Period 4)	2310.000	-67.194 (50.710)
Posting rate (Period 5)	1614.000	-77.020 (46.518)
Posting Tendency (Linear Shape)	1801.000	-5.009 (129.923)
Assimilation	1798.000	7.562 (15.943)
Gender on Posting	1825.000	3.782 (78.867)
Age on Posting	1755.000	-8.259 (220.323)
Tenure on Posting	1900.000	-8.098 (119.231)

References

Andreoni, J. (1988). Privately provided public goods in a large economy: the limits of altruism. *Journal of Public Economics, 35*(1), 57-73.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 44-54). ACM.

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, *29*(6), 991-1013.

Carley, K. (1991). A theory of group stability. *American Sociological Review*, 56, 331–354.

Edwards, J. (2012). "*Meet The 30 Biggest Social Media Advertisers Of 2012*." At: http://www.businessinsider.com/the-30-biggest-social-media-advertisers-of-2012-2012-9?op=1#ixzz2S8ayIUHf

Fisher, L. A., & Bauman, K. E. (1988). Influence and Selection In the Friend-adolescent Relationship: Findings from Studies of Adolescent Smoking and Drinking1. *Journal of Applied Social Psychology*, *18*(4), 289-314.

Holf, R. (2011) "Advertisers Flock to Social Networks." At: http://www.technologyreview.com/article/424409/advertisers-flock-to-social-networks/

Iannotti, R. J., Bush, P. J., & Weinfurt, K. P. (1996). Perception of friends' use of alcohol, cigarettes, and marijuana among urban schoolchildren: A longitudinal analysis. *Addictive Behaviors*, *21*(5), 615-632.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, 84(2), 427-436.

Kahneman, D., J. Knetsch, and R. Thaler (1991). Anomalies: The endowment effect, loss aversion and status quo bias. *Journal of Economic Perspectives* 5, 193–206.

Kirke, D. M. (2004). Chain reactions in adolescents' cigarette, alcohol and drug use: similarity through peer influence or the patterning of ties in peer networks?. *Social Networks*, *26*(1), 3-28.

Krohn, M. D., Lizotte, A. J., Thornberry, T. P., Smith, C., & McDowall, D. (1996). Reciprocal causal relationships among drug use, peers, and beliefs: A five-wave panel model. *Journal of Drug Issues*, *26*, 405-428.

Lewis, K., Gonzalez, M., & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, *109*(1), 68-72

Macy, M., J. Kitts, A. Flache, and S. Benard (2003). *A Hopfield model of emergent structure. In R. Breiger, K. Carley, and P. Pattison (Eds.), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pp. 162–173. Washington, DC: National Academies Press.

McPherson, J. M., L. Smith-Lovin, and J. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27, 415–444.

Newman, M. (2010). *Networks: an introduction*. Oxford University Press.

Phan, T., & Airoldi, E.M. (2015). The Effects of a Natural Disaster on Social Network Dynamics. *Proceedings of the National Academy of Sciences*, (forthcoming).

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 400-407

Sacerdote, B. I. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, *116*(2), 681-704.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*(1), 361-395.

Snijders, T. A., Steglich, C. E., & Schweinberger, M. (2007). Modeling the co-evolution of networks and behavior. *Longitudinal models in the behavioral and related sciences*, 41-71.

Steglich, C., Snijders, T. A., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40(1), 329-393.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1(1), 39–60.

Toubia, O., & Stephen, A. T. (2013). Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?. *Marketing Science*, *32*(3), 368-392.

Tucker, C. (2012). Social advertising. *Mimeo*, MIT.

Yoganarasimhan, H. (2012). Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics*, *10*(1), 111-150.

Zhang, X., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *The American Economic Review*, *101*(4), 1601-1615.