



## Information Systems Research

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### A Coevolution Model of Network Structure and User Behavior: The Case of Content Generation in Online Social Networks

Prasanta Bhattacharya, Tuan Q. Phan, Xue Bai, Edoardo M. Airoidi

To cite this article:

Prasanta Bhattacharya, Tuan Q. Phan, Xue Bai, Edoardo M. Airoidi (2019) A Coevolution Model of Network Structure and User Behavior: The Case of Content Generation in Online Social Networks. Information Systems Research 30(1):117-132. <https://doi.org/10.1287/isre.2018.0790>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2019, INFORMS

Please scroll down for article—it is on subsequent pages




With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# A Coevolution Model of Network Structure and User Behavior: The Case of Content Generation in Online Social Networks

Prasanta Bhattacharya,<sup>a</sup> Tuan Q. Phan,<sup>a</sup> Xue Bai,<sup>b</sup> Edoardo M. Airolidi<sup>c, d</sup>

<sup>a</sup>Department of Information Systems and Analytics, National University of Singapore, Singapore 117418; <sup>b</sup>Department of Marketing and Supply Chain Management, Department of Management Information Systems, Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122; <sup>c</sup>Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, Pennsylvania 19122; <sup>d</sup>Department of Statistics, Harvard University, Cambridge, Massachusetts 02138

Contact: [prasanta@comp.nus.edu.sg](mailto:prasanta@comp.nus.edu.sg) (PB); [phantq@comp.nus.edu.sg](mailto:phantq@comp.nus.edu.sg) (TQP); [xue@temple.edu](mailto:xue@temple.edu),  <http://orcid.org/0000-0001-9261-4998> (XB); [airolidi@temple.edu](mailto:airolidi@temple.edu) (EMA)

Received: November 15, 2016

Revised: September 14, 2017; January 13, 2018

Accepted: February 6, 2018

Published Online in Articles in Advance:  
January 10, 2019

<https://doi.org/10.1287/isre.2018.0790>

Copyright: © 2019 INFORMS

**Abstract.** With the rapid growth of online social network sites (SNSs), it has become imperative for platform owners and online marketers to quantify what factors drive content production on these platforms. Previous research identified challenges in modeling these factors statistically using observational data, where the key difficulty is the inability of conventional methods to disentangle the effects of network formation and network influence on content generation from the subsequent feedback effect of newly generated content on network structure. In this paper, we adopt and enhance an actor-oriented continuous-time statistical model that enables the joint estimation of the coevolution of the users' social network structure and of the amount of content they produce, using a Markov chain Monte Carlo–based simulation approach. Specifically, we offer a method to analyze nonstationary and continuous-time behavioral data, typically recorded in social media ecosystems, in the presence of network effects and other observable and unobservable user-specific covariates. The proposed method can help disentangle network effects of interest from feedback effects on the network. We apply our model to social network and public posting data over six months to find that (1) users tend to connect with others that have similar posting behavior; (2) however, after doing so, these users tend to diverge in their posting behavior, and (3) peer influence effects are sensitive to the strength of the posting behavior. More broadly, the proposed method provides researchers and practitioners with a statistically rigorous approach to analyze network effects in observational data. Our results lead to insights and recommendations for SNS platform owners on how to sustain an active and viable community.

**History:** Vijay Mookerjee, Senior Editor; Yuliang Yao, Associate Editor.

**Funding:** This work was partially supported by the National Science Foundation [Grants CAREER IIS-1149662 and IIS-1409177], by the Office of Naval Research [Grants YIP N00014-14-1-0485 and N00014-17-1-2131], and by a Shutzer Fellowship to E. M. Airolidi. The second author was also supported, in part, by the National University of Singapore [Grant R-253-000-110-112].

**Supplemental Material:** The online appendix is available at <https://doi.org/10.1287/isre.2018.0790>.

**Keywords:** social network structure • content production • coevolution model • latent space model • peer influence

## 1. Introduction

With the proliferation of social network sites (SNSs), platform owners are facing increasing challenges in engaging users and, subsequently, generating revenue through advertisements (Hof 2011, Tucker 2016). Unlike other internet-based services, the unique value of SNSs lies in engaging interactions between two user roles—"content producers" who actively post, comment, and share content with their friends, and "content consumers" who view and react to such content. Content producers in particular add considerable value by generating and sharing content through the network. The content posting behavior of users as well as the users' propensity to make new connections on a SNS is influenced partly by individual-level factors (e.g., demographics, traits, etc.) and partly by

their online social network characteristics such as their number of online friends, extent of network clustering, network betweenness, and so forth (Airolidi et al. 2011, Lu et al. 2013, Newman 2010). From previous research, it remains an empirical puzzle to estimate how a user's social network, such as the number of friends or the extent of clustering in the user's network, impacts the user's content posting behavior. The key challenge lies in that the user's posting behavior and social network coevolve by affecting each other, that is, behavior shapes the network at the same time that the network shapes behavior. From a classical social network perspective, addressing this puzzle amounts to separating the effect of social influence (i.e., when network influences attitude/behavior) from any selection arising from homophily on observable or unobservable

covariates and other context effects (Borgatti and Foster 2003, McPherson et al. 2001, Shalizi and Thomas 2011). A number of earlier studies have investigated the presence of either homophily or social influence in separate contexts (Lazarsfeld et al. 1954, McPherson et al. 2001). The few that do look at their coexistence within a single context tend to focus primarily on the relative strengths of homophily and influence (Borgatti and Foster 2003, Ennett and Bauman 1994, Kirke 2004). However, it is quite plausible that both homophily and influence play an important role in different temporal stages of the individual's life cycle, and to varying extents. In addition to this temporal dependency of homophily and influence, there could also exist a dependency on the specific level of the behavior or preference in question; that is, is an individual equally susceptible to a change in friendship or behavior at all levels of magnitude of the behavior in question? These critical theoretical questions have significant practical implications for platform owners and marketers.

Within the field of information systems (IS) too, there have been recent attempts at addressing questions pertaining to homophily and influence in a multitude of contexts. The studies by Singh et al. (2011), Singh and Phelps (2013), Dewan et al. (2017), and Zeng and Wei (2013) offer some notable examples and are discussed in detail in the following section. In the present study, we add to these previous approaches by developing an actor-based and continuous-time coevolution model that operates under a set of Markovian assumptions to explicitly model and jointly estimate the evolution of an online social network and online posting behavior of users. Drawing on the framework of Snijders et al. (2007), and by leveraging prior work on latent space models (Davin et al. 2014, Hoff et al. 2002), we extend and contribute to current modeling approaches in certain key ways for SNS platforms. First, we model the coevolution in an online dynamic behavioral setting, where the behavioral traits are not limited to a dichotomous variable, as was the case in previous studies (e.g., smoking versus no smoking). Instead, we discretize the number of posts made by the user into quantiles to capture the variance in posting rate. This provides us with added information about posting behavior and increased flexibility in modeling changes in behavior over time. Second, to the best of our knowledge, this is the first study that attempts to adapt the actor-driven approach beyond slow-moving and relatively stable traits and behaviors (e.g., music tastes, smoking habits, etc.) to a dynamic and rapidly changing behavioral setting (e.g., online posting and messaging behavior, photo uploads, etc.). Third, we correct for the presence of latent homophily based on unobservable factors, which can bias the estimates for posting influence and homophily. While the presence of latent homophily has been a major

confound in studies looking to disentangle influence from homophily, we exploit our longitudinal network data set to estimate latent space positions of the actors, which can potentially control for homophily based on both observed and unobserved covariates (Davin et al. 2014, Goldsmith-Pinkham and Imbens 2013). Finally, prior applications of the coevolution model have not modeled peer effects contingent on specific levels of the traits or behavior. However, we believe that individuals are likely to display varying extents of sensitivity toward peer effects, depending on their current levels of traits or behavior. By suitably specifying the coevolution model, we show that homophily and peer influence, based on posting behavior, are sensitive to the current level of the users' posting behavior.

The estimation of network-behavior coevolution models is often nontrivial. Closed form solutions are generally not possible for the likelihood function in such models, making estimation methods such as maximum likelihood or Bayesian estimation inadequate. To overcome this hurdle, we resort to a simulation-based estimation framework based on Markov chain Monte Carlo (MCMC) estimations (Snijders 2001, Steglich et al. 2010). Specifically, we use an MCMC-based method of moments (MoM) estimator to estimate the coevolution parameters in our model. While some prior studies have used computational simulations to model endogenous evolution of network ties and individual attributes (Macy et al. 2003), our model allows for statistical inference testing and model fit assessments. Moreover, it allows for a variety of objective functions and operates under an acceptable set of assumptions (e.g., conditional independence, etc.).

Using this approach, we have made the following inferences about the nature and extent of peer influence as well as homophilous peer selection in content production on SNSs. First, we provide evidence for homophily based on similarity in content posting behavior, but not on individual-level covariates, like age or gender. Second, we report the existence of peer influence, but in a direction opposite to that of homophilous interaction. Specifically, we find opposing roles of behavioral similarity at different stages of friendship formation. Individuals befriend others who are similar in content production behavior during the friendship formation stage, but gradually diverge from these friends over time. Third, we provide evidence that the strength of homophilous friend selection as well as social influence varies as a function of the specific level of the behavior. Furthermore, we show that low content posters are more susceptible than heavy content posters to homophilous friend selection. However, once they make friends, low posters are more likely to diverge from their peers, as compared with heavy posters. Moreover, we show that these findings are robust to the presence of potential latent homophily arising from unobservable factors.

In the following section, we present a summary of previous studies that discuss peer effects in social networks and a relatively newer set of studies that have used the coevolution model in varying contexts. Next, we offer a brief summary of the coevolution model that we use in our empirical analyses. Following this, we discuss our empirical setting and demonstrate our findings. We conclude with a discussion of the key contributions of our study, the limitations, and a roadmap for future research.

## 2. Related Work

### 2.1. Peer Effects in Social Networks

Social science researchers have always been interested in understanding the interdependence between the behavior of group members and the group's structure, as reflected by intermember ties within the group. For instance, sociologists and psychologists have long discussed the effect of social cohesion among group members on norm compliance and deviance (Asch 1951, Durkheim 1884, Homans 1961). Researchers have also investigated the role of individual actions on emergent social outcomes and social structures (Emirbayer and Goodwin 1994, Homans 1961, Stokman and Doreian 1997).

More recently, researchers have observed that the preference and behavior of individuals tend to be more similar when they are connected in a relationship than when they are not (Hollingshead 1949, Newcomb 1962). This phenomenon has been studied under various forms, the most common of which are homogeneity bias (Fararo and Sunshine 1964) and network autocorrelation (Doreian 1989). Over time, network autocorrelation has also been observed and studied extensively in the context of online social networks (Aral et al. 2009, 2013; Aral and Walker 2014; Lewis et al. 2012). While some propose the idea of social influence or network-driven assimilation as a potential cause of such effects (Asch 1951, Friedkin 2001, Oetting and Donnermeyer 1998, Singh and Phelps 2013), others propose selection-based mechanisms like homophily to explain why such effects might occur (Aral et al. 2009, 2013; Lazarsfeld et al. 1954; McPherson et al. 2001; Nahon and Hemsley 2014). Borgatti and Foster (2003) described these competing perspectives in terms of the temporal ordering and causal validity of network or behavioral change. Specifically, they suggest that if behavior is the consequence of network change, then this can be explained by peer influence. If, however, the network is the consequence of behavior change, then this can be explained by selection mechanisms such as homophily, but only if the temporal antecedence is causal.<sup>1</sup>

Within the field of information systems, too, the study of peer effects, particularly the study of social influence, through experimental and quasi-experimental approaches has been steadily gaining importance.

However, most prior work in this field have focused on the role of influence in either nonhuman affiliation networks, such as internal and external open source project networks (Singh et al. 2011, Singh and Phelps 2013), specialized networks such as those that operate on music or photo-sharing communities (Dewan et al. 2017, Zeng and Wei 2013), investing communities (Gu et al. 2014), or communities of users who adopt expensive items (de Matos et al. 2014). Moreover, barring a few exceptions (Aral et al. 2009, Bramoullé et al. 2009, de Matos et al. 2014), most prior studies have not explicitly modeled both homophily and influence in their networks (Bampo et al. 2008, Gu et al. 2014, Singh et al. 2011, Singh and Phelps 2013, Wang et al. 2013). In de Matos et al. (2014), the authors made intelligent use of community detection algorithms, specifically using the internal-external ratio method introduced by Krackhardt and Stern (1988), coupled with an instrumental variable approach. Their method makes strong assumptions on the presence of one-hop homophily and the absence of two-hop homophily, which can sometimes be untenable in large social networks that generally exhibit strong and varied preferential attachments. In their paper, in addition, they do recommend using “stochastic agent-based models to describe the coevolution of adoption and of the formation of social ties within and between communities” (de Matos et al. 2014, p. 1105). This is similar to our proposed approach, which does not make any assumptions on the presence or absence of specific types of homophily, and within a principled statistical modeling and inference framework. Other studies that have exploited interesting identification strategies include those by Dewan et al. (2017), who leverage the introduction of a new social cue feature to manipulate popularity influence, and Aral et al. (2009), who use propensity score matching to control for unobserved user heterogeneities. However, in our research, we are limited by the absence of any such interventions within our observation period, and by the data limitations on the number of observed covariates that limit the applicability of matching-based methods. The closest to our research context in terms of the research question as well as findings is the study by Zeng and Wei (2013), where the authors compare similarity in photo uploads to Flickr over time and the effect of this similarity on uploading behavior following tie formation. The results from their study are consistent with ours in showing that there exists homophily based on similarity in photos uploaded, as captured by the cosine similarity of photo tags. However, this similarity decreases over time following tie formation, as users try to create a unique presence on the platform. Our study generalizes this finding in the broader context of content production on a large and general-purpose SNS, while also explicitly modeling the network formation.



In the current study, we extend the above-described techniques for studying peer effects using longitudinal network data. First, our study departs from earlier attempts that investigate peer effects in specialized contexts (e.g., Dewan et al. 2017, Zeng and Wei 2013, Gu et al. 2014) by focusing on studying a commonly observed online behavior, that is, social media postings, and for a representative sample of users on a large and general-purpose social network site. Second, we improve on a number of prior studies that rely on independence of observations assumptions, by offering an approach that models the network dependence of users. Last, the continuous-time stochastic model presented in this paper is an improvement over competing methods that are problematic to use in the presence of incomplete observations, as is often the case with longitudinal discrete-time data sets where observations about the user and the network are made only at specific points in time, with little information about inter-period dynamics. However, ignoring the evolutionary dynamics between discrete time periods can significantly affect our ability to make inferences about peer effects, as pointed out by Steglich et al. (2010).

## 2.2. Content Production in Online Social Networks

Social network sites have been a subject of active research in several disciplines, including information systems, marketing, social psychology, and computer science. While a number of studies have analyzed SNS use and its effects on various user outcomes, like psychological well-being and social capital formation (Hargittai 2007, Steinfield et al. 2008, Valkenburg et al. 2006, Wellman et al. 2001), others have taken a more normative approach to discuss how user engagement increases on an SNS (Fogg and Eckles 2007) and whether such engagement has a positive impact on users (Binder et al. 2009, Livingstone 2008). Finally, there have been exemplary efforts probing how organizations use social media to engage more effectively with their target users both inside and outside the organization (Sinclair and Vogus 2011, Steinfield et al. 2009, Waters et al. 2009).

The creation and spread of user-generated content as a means of online word of mouth (WOM) has interested social network researchers for several decades, and has been extensively used by brand marketers to understand and increase brand awareness, evaluation, and sales (Bai et al. 2005, Chevalier and Mayzlin 2006, Goh et al. 2013, Reingen 1984). The common thread that emerges from this extant literature on WOM is that greater volume of WOM essentially leads to higher engagement levels on the platform. Furthermore, ensuring a persistent and critical mass of users on the platform enables advertisers to monetize by delivering more advertisements to the users in a targeted fashion (Goldfarb and Tucker 2011).

Previous work using historical WOM data, however, faces limitations by ignoring the interdependencies of the underlying network structure and the content production process. Consequently, there has been very little work that looks at objectively investigating and solving the network autocorrelation problem in online contexts (Backstrom et al. 2006, Crandall et al. 2008, Singla and Richardson 2008). The few attempts that exist focus primarily on establishing the presence of either influence or homophily, and do not provide a flexible model that is geared toward performing stronger inference testing. Two exceptions to this are the recent studies by Aral et al. (2009) and Snijders et al. (2007). Both of these models follow fundamentally different approaches to separate homophily from influence. For instance, Aral et al. (2009) uses a matched sample estimation framework that hinges on the presence of several user-specific attributes and preferences to perform suitable matching. Snijders et al. (2007), however, uses a more parsimonious random-graph-based model in an offline setting with relatively stable behaviors like smoking and alcohol consumption. In the following section, we describe and extend the approach of Snijders et al. (2007) and illustrate the utility of this model in disentangling social effects for dynamic and nonstationary behavior in an online setting.

## 3. Coevolution Model of Networks and Behavior

The coevolution approach offers a continuous-time scenario in which users simultaneously alter their network ties as well as their behavior at random instants in time, which may or may not be observed by the researcher. This class of models is an improvement over some of the earlier continuous-time Markov chain models, like the reciprocity model (Wasserman 1977, 1980a), which possess two main limitations. First, the models assume dyadic independence in the social network, which makes the analysis computationally convenient, but is untenable in most real-world contexts. Second, such models face restricted capability with parameter estimation and subsequent counterfactual analyses (Mayer 1984, Wasserman 1980b). These limitations were largely mitigated by the use of MCMC-based stochastic simulation models for sociometric data, as proposed in Snijders (1996) and later extended empirically in de Bunt et al. (1999) and Snijders et al. (2007). Consequently, a number of recent studies have used this coevolution model to investigate the effects of selection and influence on social behaviors such as substance abuse among friends (Steglich et al. 2010), diffusion of innovation (Greenan 2015), and the evolution of self-reported music and movie tastes among adolescents (Lewis et al. 2012).

We draw on these recent advances in the coevolution modeling strategy and develop an actor-based

continuous-time model for the coevolution of online network formation and content generation. Our model builds on and extends Snijders et al. (2007) and Steglich et al. (2010) in several key ways and is applied to a unique and deidentified panel data set from a large, American social networking site. Unlike previous studies that have investigated coevolution of network and behavior in an offline context with self-reported network and behavior data, the current research uses objective network and posting behavior data over a six-month period. Moreover, while previous studies have predominantly focused on stable behaviors like smoking and alcoholism, which do not change frequently over time, the current study focuses on dynamic behaviors, and specifically, content production, which has a higher frequency of change. Furthermore, we extend the previous methods to model nonbinary behaviors by discretizing online posting behavior based on several quantiles of intensity (e.g., ranging from levels 1 to 10).

We illustrate the model in Figure 1. The various blocks in the model indicate the model variables, and the directed links represent the direction of effect. The two stages of the coevolution model include the effect of posting volume on tie creation via posting volume similarity, and the effect of the tie creation on the posting volume. The bold links represent direct effects, a dotted link from A to B implies that A contributes to B in the model (i.e., A is the observed variable, while B is a derived measure). For example, the dotted link from posting volume to posting similarity indicates that the homophily in posting behavior that we model in our study, as denoted by posting similarity, is a function of the observed posting volume of the users. To illustrate in detail, our homophily hypothesis contends that users are more likely to make friends

with others from similar demographics (e.g., age, gender, and SNS tenure) as well as similar posting volume. This is largely due to (i) a need to forge social connections with similar others and (ii) the posting volume on the SNS serving as a proxy for unobserved preferences about the user (i.e., an active SNS user versus a highly discrete SNS user). However, once the friendships are created, the focal user is now exposed to a much larger set of content shared by the peers, prompting the user to feel the need to assert her own self-image and uniqueness within the social group. Thus, our influence hypothesis contends that this need for uniqueness would impact the user's rate of posting in a direction that is opposite to that of her peers. This is consistent with recent literature in economics that discusses how group contributions might crowd out contributions by the focal user, who might feel either overloaded with all the content or develop a tendency to free ride on other content (Olson 1965, Zhang and Zhu 2011). Conversely, however, users who do not post actively on the SNS might get motivated by observing the high posting rates of peers and feel a positive urge to engage in the activity as well (Andreoni 2007). This would lead to a positive peer influence.

### 3.1. Model Specification

We observe a network with  $N$  users for a total of  $T$  months and model two main variables, namely, the state of the time-varying friendship network, an  $N \times N$  matrix  $A_t$ , and the number of public posts contributed by users at time  $t$ , denoted by an  $N \times 1$  time-varying, integer-valued posting behavior vector  $P_t$ .

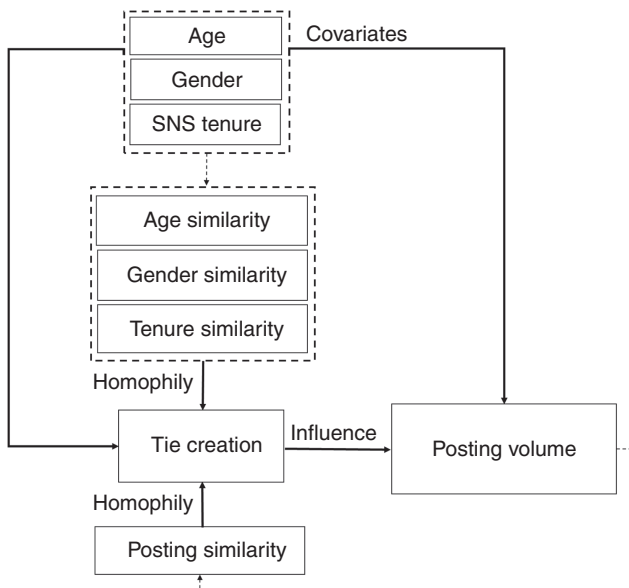
**3.1.1. Timing of Decision.** We assume that the evolution of both the network and the behavior follows a first-order Markov process, using very small time increments, often referred to as “microsteps,” that occur at random instants in time. In other words, the network evolves in continuous time but is observed at discrete moments. At any given microstep, we constrain the network or the behavior to allow only a single unit change, that is, a tie forms or dissolves, or the posting volume increases or decreases by one unit. Using a Poisson process, we model these specific points in time when any given user  $i$  gets the opportunity to make a decision to change the vector of her outgoing tie variables  $a_{ijt} = [A_t]_{ij}$ ,  $j = 1, \dots, N - 1$ , or her behavior variable  $p_{it} = [P_t]_i$ .

The Poisson rates at which the users make network decisions ( $\lambda_i^{[A]}$ ) and behavioral decisions ( $\lambda_i^{[P]}$ ) between time periods  $t$  and  $t + 1$  are controlled by rate functions as described in Equations (1) and (2):

$$\lambda_i^{[A]}(A_t, P_t) = \rho_m^{[A]} \exp(h_i^{[A]}(\alpha^{[A]}, A_t, P_t)) \quad (\text{network decisions}), \quad (1)$$

$$\lambda_i^{[P]}(A_t, P_t) = \rho_m^{[P]} \exp(h_i^{[P]}(\alpha^{[P]}, A_t, P_t)) \quad (\text{behavioral decisions}), \quad (2)$$

Figure 1. Feedback Based Two-Stage Coevolution Model



where the parameters  $\rho_m^{[A]}$  and  $\rho_m^{[P]}$  are dependent on the observed discrete time period and capture periodic variations in either network or posting behavior, and the functions  $h_i^{[A]}(\cdot)$  and  $h_i^{[P]}(\cdot)$  model dependence of the rates on the current state of the network and the posting behavior. The exact functional forms of  $h_i^{[A]}(\cdot)$  and  $h_i^{[P]}(\cdot)$  depend on the network and behavioral effects that the analyst chooses to model in the context of an application, and we fully specify these functions in Section 3.2 for the purpose of our analysis. In the current model specification, we assume that the rate functions are constant across the actors and are dependent only on the specific discrete observation periods  $m$ .

**3.1.2. Objective Function.** While the rate functions control the timing of the users' decisions (i.e., to change network or behavior), the objective functions model the specific changes that are made. A user  $i$  optimizes an objective function in the current time period over the set of feasible microsteps she can take (Steglich et al. 2010). This objective function is composed of three parts: the evaluation functions  $f_i^{[A]}$  and  $f_i^{[P]}$ , the endowment functions  $g_i^{[A]}$  and  $g_i^{[P]}$ , and random disturbances  $\epsilon_i^{[A]}$  and  $\epsilon_i^{[P]}$ , capturing residual noise.

The evaluation functions are parameterized by the vectors  $\beta^{[A]}$  and  $\beta^{[P]}$ ; the endowment functions are parameterized by the vectors  $\gamma^{[A]}$  and  $\gamma^{[P]}$ , as shown in Equations (3) and (4):

$$f_i^{[A]}(\beta^{[A]}, A_t, P_t) + g_i^{[A]}(\gamma^{[A]}, A_t, P_t | A_{t-1}, P_{t-1}) + \epsilon_i^{[A]}(A_t, P_t) \quad (\text{network decisions}), \quad (3)$$

$$f_i^{[P]}(\beta^{[P]}, A_t, P_t) + g_i^{[P]}(\gamma^{[P]}, A_t, P_t | A_{t-1}, P_{t-1}) + \epsilon_i^{[P]}(A_t, P_t) \quad (\text{behavioral decisions}), \quad (4)$$

The evaluation functions capture the utility obtained by a user  $i$  from her network-behavior configuration. The functions  $f_i^{[A]}(\cdot)$  and  $f_i^{[P]}(\cdot)$  in (3) and (4) provide a measure of fitness of the state of the network and posting behavior, as perceived by the users. This implies that users constantly strive to make specific changes to their friendship network and posting behavior to maximize the value of this evaluation function.

The endowment functions  $g_i^{[A]}(\cdot)$  and  $g_i^{[P]}(\cdot)$ , from (3) and (4), capture the part of utility that is lost when either the network ties or the posting behavior is changed by a single unit, but that was obtained without any "cost" when this unit was gained earlier. In other words, such endowment functions are useful to model situations where the creation and dissolution of ties, or the increase and decrease in posting behavior, are asymmetric in terms of utility gained or lost. However, since in the context of our study we do not model deletion of friends on the platform or the deletion of content, we do not include such endowment functions in our model.

**3.1.3. Choice Probabilities and Intensity Matrix.** The final terms in the objective function described in (3) and (4) are the set of random and i.i.d. residuals  $\epsilon_i^{[A]}$  and  $\epsilon_i^{[P]}$ . As is the case with random utility models, if we assume that these residuals follow a type 1 extreme value distribution, we can write the resulting choice probabilities for the network and posting microstep decisions as a multinomial logit (Maddala 1986). For the network microstep decision, the resulting choice probability is as illustrated in Equation (5):

$$\Pr(a_{ijt+1} = a_{ijt} + \delta | a_t, p_t, \beta^{[A]}) = \frac{\exp(f_i^{[A]}(\beta^{[A]}, a_{ijt} + \delta, p_t))}{\sum_{\varphi} \exp(f_i^{[A]}(\beta^{[A]}, a_{ijt} + \varphi, p_t))}, \quad (5)$$

where  $a_{t+1}$  is the resulting network at  $t+1$  when user  $i$  at microstep  $t$  alters the value of her tie variables by  $\delta$  (or  $\varphi$ ), where  $\delta, \varphi \in \{0, 1\}$ ; that is, user  $i$  either creates a new tie or makes no change to her network.<sup>2</sup> Also note that the tie formation function  $f_i^{[A]}$  depends on the alters' posting behavior, through the vector  $p_t$ , which runs over all actors,  $j = 1 \dots n$ , and thus may be different for different groups of alters. Similarly, for the posting microstep decision, the resulting choice probability is as illustrated in Equation (6):

$$\Pr(p_{it+1} = p_{it} + \delta | a_t, p_t, \beta^{[P]}) = \frac{\exp(f_i^{[P]}(\beta^{[P]}, a_t, p_{it} + \delta))}{\sum_{\varphi} \exp(f_i^{[P]}(\beta^{[P]}, a_t, p_{it} + \varphi))}, \quad (6)$$

where  $p_{t+1}$  denotes the resulting state of posting behavior in  $t+1$  when user  $i$  changes her posting volume at microstep  $t$  by a factor of  $\delta$  (or  $\varphi$ ), where  $\delta, \varphi \in \{-1, 0, 1\}$ ; that is, the user increases her posting volume by one unit, decreases it by one unit, or makes no new posts.

Once we have formulated the choice probabilities, the subsequent transition matrix  $Q$ , also called as the intensity matrix, models the transition from state  $(a_t, p_t)$  at microstep  $t$  to a new state  $(a_{t+1}, p_{t+1})$  at microstep  $t+1$ , and can be specified by the following entries:

$$Q(a_{t+1}, p_{t+1}) = \begin{cases} \lambda_i^{[A]} \Pr(a_{ijt+1} = a_{ijt} + \delta | a_t, p_t), & \text{if } (a_{ijt+1}, p_{it+1}) = (a_{ijt} + \delta, p_t); \\ \lambda_i^{[P]} \Pr(p_{it+1} = p_t + \delta | a_t, p_t), & \text{if } (a_{ijt+1}, p_{it+1}) = (a_{ijt}, p_t + \delta); \\ -\sum_i \left\{ \sum_{\delta \in \{-1, 1\}} Q(a_{ijt} + \delta, p_t) + \sum_{\delta \in \{-1, 1\}} Q(a_{ijt}, p_t + \delta) \right\}, & \text{if } (a_{t+1}, p_{t+1}) = (a_t, p_t); \text{ and} \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

**3.1.4. Model Estimation.** Because of the complexity of explicitly computing the likelihood function, we employ the use of simulation-based estimators. Specifically, we use an MCMC-based MoM estimator to recover the parameters of these rate and evaluation functions. The MoM estimator for our data and the parameters is based on the set of network and behavior-related statistics that are specified in the following section. The MCMC implementation of the MoM estimator uses a stochastic approximation algorithm that is a variant of the Robbins–Monro algorithm (Robbins and Monro 1951), as detailed in the Online Appendix 2. This method uses a score function method (Schweinberger and Snijders 2007) to estimate the standard errors, an improvement over previous approaches that used finite-difference-based estimators. In general, the standard errors are computed by taking the square root of the variance–covariance matrix  $D'_\theta \Sigma_\theta D_\theta^{-1}$ , where  $D_\theta$  is the partial derivatives matrix, and  $\Sigma_\theta$  is the covariance matrix. Other modeling and inference strategies are available (Azari and Airolidi 2012, Han et al. 2015, Yang et al. 2013).

The following section describes the empirical context for testing (1) the proposed coevolution model to investigate the presence of peer effects and (2) the dependence of these peer effects on the state of the posting behavior.

### 3.2. Model Parameterization in Context of SNSs

In our context, the functions  $h_i^{[A]}(\cdot)$ ,  $h_i^{[P]}(\cdot)$ ,  $f_i^{[A]}(\cdot)$ , and  $f_i^{[P]}(\cdot)$  from (1)–(4) can be modeled as weighted sums of various network characteristics (e.g., degree, transitivity, homophily based on user covariates, etc.) and behavioral characteristics (e.g., behavior trends, similarity measure, effect of user covariates on behavior, etc.). We denote the matrices of network and behavior statistics computed in each time period  $t$  by  $S_t^{[A]}$  and  $S_t^{[P]}$ , which are  $N \times K_1$  and  $N \times K_2$  matrices of  $K_1$  network and  $K_2$  behavioral characteristics, respectively. The functions  $h_i^{[A]}(\cdot)$  and  $h_i^{[P]}(\cdot)$  from the rate functions are specified as follows:

$$h_i^{[A]}(\alpha^{[A]}, A_t, P_t) = \sum_q \alpha_q^{[A]} s_{iq_t}^{[A]}(A, P), \quad (8)$$

$$h_i^{[P]}(\alpha^{[P]}, A_t, P_t) = \sum_r \alpha_r^{[P]} s_{ir_t}^{[P]}(A, P). \quad (9)$$

Here,  $\alpha_q$  indicates dependence on the statistics  $s_{iq_t}^{[A]}(A, P)$ , and  $q \in K_1$ . Similarly, coefficient  $\alpha_r$  indicates dependence on the statistics  $s_{ir_t}^{[P]}(A, P)$ , and  $r \in K_2$ , where  $s_{iq_t}^{[A]}(A, P)$  and  $s_{ir_t}^{[P]}(A, P)$  are vectors of one-dimensional statistics defined for each user  $i$  and used to capture the rate dependence on the user's network characteristics (e.g., out-degree) and behavioral characteristics (e.g., SNS tenure), respectively. For the current set of analyses, however, we hold both sets of rate functions to be constant across all actors and model only the

dependence on the time period, that is, parameters  $\rho_m^{[A]}$  and  $\rho_m^{[P]}$  in (1) and (2). Similarly, the functions  $f_i^{[A]}(\cdot)$  and  $f_i^{[P]}(\cdot)$  can be specified as follows:

$$f_i^{[A]}(\beta^{[A]}, A_t, P_t) = \sum_{k_1} \beta_{k_1}^{[A]} s_{ik_1}^{[A]}(A_t, P_t) \quad (\text{network evaluations}), \quad (10)$$

$$f_i^{[P]}(\beta^{[P]}, A_t, P_t) = \sum_{k_2} \beta_{k_2}^{[P]} s_{ik_2}^{[P]}(A_t, P_t) \quad (\text{behavior evaluations}), \quad (11)$$

where  $s_{ik_1}^{[A]} = [S^{[A]}]_{ik_1}$  is the  $k_1^{\text{th}}$  network statistic of user  $i$ , and, similarly,  $s_{ik_2}^{[P]} = [S^{[P]}]_{ik_2}$  is the  $k_2^{\text{th}}$  behavioral statistic of user  $i$ .

We parameterize the objective function based on our current research context, that of online posting behavior among a student population on a large and popular SNS. Specifically, we seek to investigate the presence of homophilous friendship formation based on similarities in posting behavior, as well as the role of peer influence in regulating content generation over time. Furthermore, we also analyze the dependency of peer effects on the specific state of the posting behavior to investigate whether active content posters react differently to peer effects compared with less active posters.

#### 3.2.1. The Presence of Homophily and Peer Influence.

In this section, we define and specify key estimation statistics for both the network and the posting behavior effects that we model in our study.

**Social network effects.** The network effects from  $S_t^{[A]}$  that we model are the user  $i$ 's out-degree ( $s_{i1t}^{[A]}$ ), transitivity ( $s_{i2t}^{[A]}$ ), homophily effects based on posting behavior ( $s_{i3t}^{[A]}$ ), and homophily based on the covariates, gender ( $s_{i4t}^{[A]}$ ), age ( $s_{i5t}^{[A]}$ ), and SNS tenure ( $s_{i6t}^{[A]}$ ). We also include effects that model the influence of individual covariates, that is, gender ( $Gender_i^{[A]}$ ), age ( $Age_i^{[A]}$ ), and SNS tenure ( $SNSTenure_i^{[A]}$ ), on the propensity to form new friends. The mathematical illustrations are provided in Equations (12) through (17):

(i) Degree ( $s_{i1t}^{[A]}$ ) and Transitivity ( $s_{i2t}^{[A]}$ ):

$$s_{i1t}^{[A]}(a) = \sum_j a_{ijt}, \quad (12)$$

$$s_{i2t}^{[A]}(a) = \sum_{j,h} a_{ijt} \cdot a_{jht} \cdot a_{iht}. \quad (13)$$

(ii) Homophily based on posting behavior and covariates (gender, age, SNS tenure):

$$s_{i3t}^{[A]}(a, p) = a_{i3t}^{-1} \sum_j a_{ijt} \left( 1 - \frac{|p_{it} - p_{jt}|}{R_{pt}} \right), \quad (14)$$

where  $R_{pt}$  is the range of the posting variable  $P$  at step  $t$ . Variable  $s_{i3t}^{[A]}$  represents the effect of homophily,



based on posting behavior, such that  $s_{i3t}^{[A]}$  takes a higher value for those users whose posting volume is closer to that of their peers (i.e., the value of  $|p_{it} - p_{jt}|$  is small). Thus, a drive toward a higher value of  $s_{i3t}^{[A]}$  can be seen as an increased propensity toward creating homophilous friendships based on similarity in posting behavior.

For covariates  $X = \{Gender, Age, SNSTenure\}$ , we have similar expressions for  $s_{i4t}^{[A]}$ ,  $s_{i5t}^{[A]}$ , and  $s_{i6t}^{[A]}$ , respectively:

(iii) *Covariate ( $X_j$ ) on the degree effects (i.e., effect of user's gender, age, and SNS tenure on her degree):*

$$Gender_{it}^{[A]}(a, x) = \sum_j a_{ijt} \cdot x_{1i}, \quad (15)$$

$$Age_{it}^{[A]}(a, x) = \sum_j a_{ijt} \cdot x_{2it}, \quad (16)$$

$$SNSTenure_{it}^{[A]}(a, x) = \sum_j a_{ijt} \cdot x_{3it}. \quad (17)$$

$Gender_{it}^{[A]}$  represents the effect of the user  $i$ 's gender ( $x_1$ ) on her propensity to make new friends during step  $t$ , such that a positive and significant estimate on the statistic would imply that females ( $Gender = 2$ ) make more friends than males ( $Gender = 1$ ). We have similar expressions for  $Age_{it}^{[A]}$  and  $SNSTenure_{it}^{[A]}$ , respectively. In all the above equations,  $a_{ijt} = 1$  if a tie exists between  $i$  and  $j$  in step  $t$  and 0 otherwise.

**Posting behavior effects.** Next, we specify the rate and evaluation functions as defined for the posting behavior. In (11), the behavior effects that we model are the user's behavior tendency effect ( $s_{i1t}^{[P]}$ ); the peer influence effect, that is, social influence  $s_{i2t}^{[P]}$ ; and effects that capture the influence of individual covariates like gender ( $Gender_{it}^{[P]}$ ), age ( $Age_{it}^{[P]}$ ), and SNS tenure ( $SNSTenure_{it}^{[P]}$ ) on the posting behavior,  $P$ . We provide the mathematical illustrations in (18) through (22):

(i) *Behavioral tendency effect (this captures the natural tendency of users to increase or decrease behavior over time):*

$$s_{i1t}^{[P]}(a, p) = p_{it}. \quad (18)$$

(ii) *Peer influence effect (the propensity of users to assimilate in behavior toward their peers):*

$$s_{i2t}^{[P]}(a, p) = a_{i+}^{-1} \sum_j a_{ijt} \left( 1 - \frac{|p_{it} - p_{jt}|}{R_{pt}} \right), \quad (19)$$

where  $R_{pt}$  is the range of the posting variable  $P$ . The effect of peer influence based on posting behavior is represented by  $s_{i2t}^{[P]}$ , which would have a higher value for those users whose posting volume is closer to that of their peers (i.e., the value of  $|p_{it} - p_{jt}|$  is smaller). Thus, a positive and significant estimate on this statistic would indicate that users regulate their posting behavior to assimilate with their peers, that is, match the posting rate of their peers, and vice versa.

(iii) *Influence of covariates (i.e., gender, age, SNS tenure) on behavior:*

$$Gender_{it}^{[P]}(p, x) = p_{it} \cdot x_{1i}, \quad (20)$$

$$Age_{it}^{[P]}(p, x) = p_{it} \cdot x_{2it}, \quad (21)$$

$$SNSTenure_{it}^{[P]}(p, x) = p_{it} \cdot x_{3it}. \quad (22)$$

Here,  $Gender_{it}^{[P]}$  represents the effect of gender ( $x_{1i}$ ) on posting behavior such that a significant and positive estimate on this statistic would indicate that females ( $Gender = 2$ ) post more than males ( $Gender = 1$ ). Similarly, (21) and (22) represent the effects of age and SNS tenure on posting behavior, respectively.

It is clear from the above formulation of effects that the mathematical illustration for the network and behavior effects to compute homophily (14) and peer influence (19) are identical. This point lies at the core of the problem that is separating the effect of homophilous selection from peer influence. However, we exploit the longitudinal nature of our data set to successfully identify temporal sequentiality across the periods. In other words, we use dyads of users who first become friends and then converge in behavior to identify influence. Similarly, we use dyads of users who show similarity in behavior before becoming friends to identify homophily. While there might be other latent confounds that we do not capture in our modeling, our 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).

**3.2.2. Behavioral Dependency of Homophily and Peer Influence.** While homophilous or assortative relationships among individuals have been reported extensively in previous research on the subject (Aral et al. 2009, McPherson et al. 2001, Park and Barabási 2007), what remains to be investigated is whether such homophilous selection effects vary in strength depending on the current state of the observable attribute or behavior. For instance, consider how an individual who smokes cigarettes is more likely to make friends with a fellow smoker (Christakis and Fowler 2008, Pearson and West 2003). However, would his affinity to make friends with a similar smoker be any higher or lower depending on how many cigarettes he smokes each day at the present moment? An analogous problem arises in studying influence. It has been widely observed that peer influence plays an important role in the onset and sustenance of various addictive behaviors, including smoking (Christakis and Fowler 2008, Ennett and Bauman 1994). However, little is understood about whether such peer influence effects are particularly stronger or weaker for different levels of a behavior itself.

In our study, we investigate whether SNS users show varying strengths of selection bias due to homophily

and susceptibility to peer influence depending on their current levels of posting behavior. To achieve this, we cluster all users depending on their levels of posting behavior into three major categories. Based on the volume of content generated, we categorize the top 10 percentile of individuals in each time period as most active posters (MAPs), the bottom 10 percentile of individuals as least active posters (LAPs). All other users are categorized as moderately active posters (MoAPs). We introduce dummy variables for the MAP and LAP groups in our model, keeping the MoAP group for comparison. This is shown in (23) and (24). The estimates from the interaction between these dummy variables and our homophily and peer influence variables will help us address our question at hand:

$$s_{i7t}^{[A]}(a, p) = MAP_i \cdot a_{i+t}^{-1} \sum_j a_{ijt} \left( 1 - \frac{|p_{it} - p_{jt}|}{R_{pt}} \right), \quad (23)$$

$$s_{i8t}^{[A]}(a, p) = LAP_i \cdot a_{i+t}^{-1} \sum_j a_{ijt} \left( 1 - \frac{|p_{it} - p_{jt}|}{R_{pt}} \right), \quad (24)$$

where  $R_{pt}$  is the range of the variable  $P$ . In the above equations, the  $MAP_i$  and  $LAP_i$  dummy variables denote whether a user  $i$  is a heavy poster or low poster. The middle group ( $MoAP_i$ ) is held as the baseline group for comparison of estimates. Similar effects are constructed for the interaction of these activity dummies and the behavioral homophily effect ( $s_{i3t}^{[P]}$  and  $s_{i4t}^{[P]}$ ).

## 4. Estimation Results

### 4.1. Data Context

We obtained online network data from a large, American social networking site for 2,507 deidentified undergraduate students attending a North American university for the months from September 2008 till February 2009. Additionally, we recorded the aggregate number of monthly public posts made by these users on the social media platform during the same period. The descriptive statistics of the key variables are illustrated in Table 1.

The key variables for the coevolution model are  $P_{it}$  and  $D_{it}$ , which depict the total number of monthly public posts and new friends added on the SNS, respectively. The covariates include  $Gender_i$ , the gender of

the user;  $Age_i$ , the biological age of the user; and  $SNSTenure_i$ , the total number of days spent by the user on the SNS at the time of recording the data.

The network and behavior descriptive summaries are detailed in Online Appendix 3 (see Tables A1–A4). Within our observation period, the individuals produced a substantial amount of content on the social media platform and established several new friendships. This provides us with sufficient variability in our data to test our proposed models. Figure A1 in the online appendix demonstrates the trend in key network metrics like number of new friends added, network density, and average degree. Since, within our study period, we do not observe any event of unfriending on the platform, all the three metrics display a positive trend. Figures A2 and A3 in the online appendix illustrate the dynamics of the posting variable, which is the focal behavior modeled in this study.

### 4.2. The Evolution of Homophily and Peer Influence

We estimate the rate and evaluation functions from the coevolution model as specified in Sections 3.1 and 3.2.1 using a method of moments estimator and present the results in Table 2. The MoM estimator essentially tries to recover parameter estimates by matching the observed network data with the simulated network data. The convergence  $t$ -statistics for these simulations were reasonably low for a large network such as the one we are considering; this assessment is based suggestions by Ripley et al. (2018), who, based on empirical simulations on small networks, suggest the convergence statistics should be as close to zero as possible. Specifically, we provide information about the deviation of our simulated network and behavioral statistics from the observed data. Table 2 highlights the estimation results for rate parameters  $\rho_m^{[A]}$  and  $\rho_m^{[P]}$ , for a total of five months (i.e., one less than the total number of time periods since the first among six periods is conditioned upon during estimation), and estimates for  $\beta_p^{[A]}$ , where  $p$  ranges from 1 to 9, and for  $\beta_q^{[P]}$ , where  $q$  ranges from 1 to 5.

**4.2.1. Results on Networks.** For the network structure variables, as shown from the results in Table 2, we observe that the estimate for the out-degree of the

**Table 1.** Descriptive Summary of Model Variables

	Min	Max	Mean	Std. dev.
Dependent variable				
Total Monthly Public Posts ( $P_{it}$ )	0.000	556.000	12.493	25.264
Independent variables				
Biological Age ( $Age_i$ ) (years)	20.000	26.000	22.312	1.531
SNS Tenure ( $SNSTenure_i$ ) (days)	712.000	2,591.000	1,771.59	382.650
Number of Friends on SNS ( $D_{it}$ )	1.000	722.000	79.980	74.329
Total Monthly Public Posts by Friends ( $\sum_j P_{jt-1}$ )	0.000	19,080.000	1,630.18	1,702.356

**Table 2.** Estimation Results for Network and Behavior Effects

Network parameters	Estimate	Behavior parameters	Estimate
Friendship rate (period 1)	7.767*** (0.100)	Posting rate (period 1)	4.337*** (0.145)
Friendship rate (period 2)	6.267*** (0.088)	Posting rate (period 2)	3.889*** (0.143)
Friendship rate (period 3)	3.930*** (0.082)	Posting rate (period 3)	5.229*** (0.205)
Friendship rate (period 4)	4.547*** (0.075)	Posting rate (period 4)	4.995*** (0.195)
Friendship rate (period 5)	5.353*** (0.084)	Posting rate (period 5)	3.681*** (0.119)
Out-degree	−9.536*** (0.011)	Posting tendency (linear shape)	−0.196*** (0.007)
Transitivity	0.109*** (0.001)	Influence	−2.995*** (0.134)
Gender homophily	0.068 (0.057)	Gender on posting	0.007 (0.012)
Gender on degree	0.031 (0.020)	Age on posting	0.003 (0.006)
Age homophily	0.024 (0.034)	Tenure on posting	0.003 (0.010)
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)		

Note.  $N = 2,507$ .

\*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

users is negative and statistically significant ( $-9.536$ ;  $p < 0.01$ ). Since we can think of the evaluation function as a measure of the “fitness” or “attractiveness” of the state of the network, this estimate indicates that users in our network show a lower propensity over time to establish new social connections. This can be attributed to the cost of forming social connections or constrained resources (Dunbar 1992, Phan and Airolidi 2015). Furthermore, we observe that the estimate for network transitivity is positive ( $0.109$ ;  $p < 0.01$ ). This indicates that there is an increased drive toward network closure in our observed network. For instance, if users  $i$  and  $j$  are friends and users  $j$  and  $h$  are friends as well, then the user  $i$  has a stronger motivation to befriend user  $h$  over any other user in the network, as this increases the overall attractiveness of the new network state for  $i$ . We also find strong evidence for friendship formation among those with a similar level of posting behavior ( $0.127$ ;  $p < 0.01$ ). Thus, the more active posters prefer to befriend other active posters, while the less active posters prefer other less active ones. Interestingly, even though we contend that homophily on gender and age do influence the propensity to make

friends, we found that in our sample these factors were not significant. This could be due to (i) the fairly homogeneous sample and (ii) presence of other unobserved attributes that might have stronger roles, for example, ethnicity. Here, we address the latter through our latent space models, as described later. Moreover, based on the findings of Kossinets and Watts (2006), we also assert that, despite the presence of observable and unobservable homophily, some amount of random matching does happen during network formation. Thus, conditional on the network being already formed, homophily based on age or gender might not be significant anymore.

**4.2.2. Results on Behavior.** Among the behavior variables, we observe that the estimate for the linear tendency parameter is negative and statistically significant ( $-0.196$ ;  $p < 0.01$ ). As mentioned earlier, the tendency effect represents a drive toward high posting volume. A zero value on this parameter indicates the user’s preference for the average posting volume. Since we obtain a negative estimate on this parameter, it indicates that as time goes by, users prefer to post less. We also find strong evidence of peer influence among the individuals, with a significantly negative parameter for the influence effect ( $-2.995$ ;  $p < 0.01$ ). This implies that individuals tend to correct their posting behavior over time in a direction away from their peers. Specifically, this negative influence indicates the propensity of users to diverge in behavior from the average behavior across peers of the focal user. This could be a result of free-riding behavior in case the peers are contributing more, or could also be representative of an increased drive to behave in a nonconformist manner (e.g., “If everyone else is posting more, I should do something different”). Note that in the current study, we do not specifically classify peers into high- or low-volume posters because it is not clear how the content delivery system on the SNS displays content from the peers to the focal user. For instance, the content filtering and ranking system on the SNS might filter or restrict display of contents from the high-volume peer, to provide a more equitable presentation of content from multiple peers. Thus, we feel the average volume of content produced by the peers is an appropriate metric to model at the current stage.

We test the robustness of our estimation results to distributional assumptions by implementing two additional model specifications. The first model uses log-transformed posting variable to account for any skewness in the variable distribution that might be affecting our results. The second model uses *total* similarity between the focal user and her peers as a measure of our focal homophily and influence effect, as opposed to average similarity that we have used in our base models. While average similarity measure is defined by the average of centered similarity scores between

the user  $i$  and her peers, total similarity is defined by taking the sum of centered scores between the user  $i$  and her peers. This helps to check the robustness of our homophily and influence metric to changes in operationalization of our focal network effects. The results from these two robustness tests are detailed in Online Appendix 6 (Tables A7 and A8) and show consistency with results from our base model specifications, that is, a significantly positive homophily estimate and significantly negative influence estimate. We also baseline our results against regression-based models, specifically a panel fixed effects regression and a panel fixed effects Poisson model. The model specifications and results are illustrated in Online Appendix 4.

While it is hard to uncover the specific reasons for the peer effects we find, interpreting the parameters for homophily and peer influence together leads to an increased understanding of the interplay between friendship formation and content production behavior in online networks. 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 Section 5.

**4.2.3. Results on Behavioral Dependency of Homophily and Peer Influence.** In addition to the above, we also find strong evidence for the behavioral dependency of homophily and peer influence. Table 3 illustrates the estimation results for rate parameters  $\rho_m^{[A]}$  and  $\rho_m^{[P]}$ , for periods 2 to 6, and estimates for  $\beta_p^{[A]}$ , where  $p$  ranges from 1 to 11, and for  $\beta_q^{[P]}$ , where  $q$  ranges from 1 to 7. The results from the estimation show that the users in our sample demonstrate varying propensities to create homophilous relationships and varying susceptibility to peer influence, depending on the current state of their posting behavior. Specifically, compared to MoAPs, MAPs were less likely to form friendships with other MAPs ( $-0.375$ ;  $p < 0.01$ ), while LAPs were more likely to form friendships with other LAPs ( $0.293$ ;  $p < 0.01$ ). Furthermore, compared to MoAPs, both MAPs and LAPs were found to be more susceptible to peer influence. However, while MAPs showed positive influence (i.e., converge in behavior with peers;  $1.183$ ,  $p < 0.01$ ), the LAPs showed negative influence (i.e., diverge in behavior from peers;  $-6.437$ ,  $p < 0.01$ ).

### 4.3. Sensitivity to Latent Homophily

While our analysis conditions on observable behavioral (e.g., posting) and individual-level covariates,

**Table 3.** Estimation Results with Behavioral Dependency

Network parameters	Estimate	Behavior parameters	Estimate
Friendship rate (period 1)	7.750*** (0.100)	Posting rate (period 1)	4.948*** (0.274)
Friendship rate (period 2)	6.368*** (0.089)	Posting rate (period 2)	4.137*** (0.166)
Friendship rate (period 3)	3.997*** (0.075)	Posting rate (period 3)	5.477*** (0.317)
Friendship rate (period 4)	4.543*** (0.075)	Posting rate (period 4)	5.879*** (0.255)
Friendship rate (period 5)	5.431*** (0.084)	Posting rate (period 5)	4.583*** (0.252)
Out-degree	-9.562*** (0.011)	Posting tendency (linear shape)	-0.182*** (0.011)
Transitivity	0.107*** (0.001)	Influence	-2.951*** (0.268)
Gender homophily	0.076 (0.058)	MAP influence	1.183*** (0.096)
Gender on degree	0.023 (0.019)	LAP influence	-6.437*** (0.763)
Age homophily	0.023 (0.034)	Gender on posting	0.009 (0.012)
Age on degree	0.013 (0.009)	Age on posting	0.004 (0.005)
Tenure homophily	-0.001 (0.022)	Tenure on posting	0.004 (0.010)
Tenure on degree	-0.019 (0.016)		
Posting homophily	0.112*** (0.042)		
MAP homophily	-0.375*** (0.021)		
LAP homophily	0.293*** (0.023)		

Note.  $N = 2,507$ .

\*\*\* $p < 0.01$ .

(e.g., age and gender), there is a possibility that the network formation might be driven by homophily based on latent factors, such as personality traits and similarity in tastes or preferences. The presence of such latent homophily has been cited as an important confound in the estimation of social influence (Shalizi and Thomas 2011). We look to test the sensitivity of our modeling approach to the presence of such latent homophily using a latent space modeling approach, similar to what has been described in Davin et al. (2014). Latent space models are well known in social networks literature and have been traditionally employed in identifying and visualizing communities within networks. For our analysis, we use two-dimensional latent space positions as proxy variables to control for potential latent homophily. The intuition behind this approach is that if two actors are close to each other in a latent social space, then this similarity is driven by both observed as well as unobserved factors. Thus, adding latent space coordinates as model covariates would serve to reduce



the bias associated with the influence estimate by controlling for some latent homophily. There have been some prior work that have used latent space models to address similar questions in economics and marketing (Ansari et al. 2011, Braun and Bonfrer 2011). A summary of how the latent space models for our current context were specified and estimated is illustrated in Online Appendix 5.

We estimated the rate and evaluation functions from the coevolution model as specified in Sections 3.1 and 3.2.1, using the latent space positions as covariates. Specifically, we first estimated a latent space model corresponding to the network at the onset of each period, and then use these estimated latent space coordinates as model covariates for our coevolution model in each period. In other words, the latent space model was applied to a sequence of static networks, one for each period, thereby essentially creating time-varying covariates for the coevolution model. The estimation results reported in Table 4 for both homophily based on posting behavior and peer influence are consistent with our previous results. As expected, after controlling for homophily based on latent space coordinates, the estimate for posting homophily (0.104;  $p < 0.01$ ) reduces in strength, but continues to be statistically significant. This shows that there does exist evidence of homophily based on latent factors beyond the observable factors of age, gender, and SNS tenure. However, our proposed effect of posting homophily exists even after controlling for possible latent confounders. Similarly, the estimate for peer influence is weaker ( $-0.015$ ;  $p < 0.01$ ) than in our earlier models that do not account for latent homophily. We also validate the robustness of our latent space estimation model to changes in the number of spatial dimensions. Specifically, we generated per-period latent space coordinates of the social network based on a three-dimensional latent space, compared to the two-dimensional space for our current model, and reestimated our coevolution model using the three-dimensional coordinates as covariates. We provide the summary of estimation results in Online Appendix 6 (Table A9) and show that the estimates for homophily and influence are consistent with all our previous models.

In summary, we leverage latent space positions of actors in our network to account for possible latent homophily and show that our results for homophily and peer influence based on posting behavior are valid even after controlling for these latent positions in both two and three dimensional latent spaces.

## 5. Discussion and Conclusion

In the current study, we develop and estimate a model for analyzing the coevolution of content production and social network structure using real-world data from a large social network site. Extending prior

**Table 4.** Latent Homophily–Corrected Estimation Results for Network and Behavior

Network parameters	Estimate	Behavior parameters	Estimate
Friendship rate (period 1)	7.529*** (0.098)	Posting rate (period 1)	3.496*** (0.182)
Friendship rate (period 2)	6.204*** (0.141)	Posting rate (period 2)	3.590*** (0.132)
Friendship rate (period 3)	3.949*** (0.075)	Posting rate (period 3)	4.223*** (0.114)
Friendship rate (period 4)	4.586*** (0.085)	Posting rate (period 4)	4.603*** (0.122)
Friendship rate (period 5)	5.487*** (0.091)	Posting rate (period 5)	3.104*** (0.104)
Out-degree	−9.913*** (0.014)	Posting tendency (linear shape)	−0.191*** (0.008)
Transitivity	0.098*** (0.001)	Influence	−0.015*** (0.001)
Gender homophily	0.048 (0.077)	Gender on posting	0.007 (0.014)
Gender on degree	0.011 (0.021)	Age on posting	0.008 (0.012)
Age homophily	0.011 (0.032)	Tenure on posting	−0.002 (0.012)
Age on degree	0.005 (0.024)		
Tenure homophily	−0.010 (0.030)		
Tenure on degree	−0.012 (0.018)		
Posting homophily	0.104** (0.047)		
Latent posting (X) homophily	1.030*** (0.126)		
Latent posting (Y) homophily	1.110*** (0.112)		

Note.  $N = 2,507$ .

\*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

research in IS that has focused on specialized networks (e.g., open source project networks, photo-sharing communities, and financial investor networks) and has tended to underplay the network dependence of the focal actors in favor of regression-based models, our study proposes a stochastic structural model for jointly estimating the evolution of a large social network and associated content posting behavior. Our results demonstrate the role of social network structure and user characteristics in influencing content production on SNSs. We adopt an actor-driven and coevolution-based MCMC modeling approach to jointly estimate the evolution of the user's social network and posting behavior. We contend that this approach is more statistically disciplined than alternate methods, which tend to violate some key assumptions of network-based modeling. Furthermore, we depart from previous instances of the actor-driven models whose applicability is restricted to stable dichotomous behaviors,

like smoking and substance abuse. In the current study, we adapt the coevolution model to a dynamic behavior (i.e., online public posting), which often changes rapidly over successive time periods. We avoid convergence-related difficulties with MCMC estimations of such continuous behavioral variables by discretizing our behavioral variable into several quantiles to represent the intensity of behavior. We contend that by using this quantile-based binning strategy, we are able to achieve adequate convergence in estimations without much loss of information. Furthermore, we account for homophilous friend selection based on unobserved covariates, that is, latent homophily, by including latent space positions of the actors as covariates in our model. We also perform a number of robustness analyses to test for sensitivity to distributional variations, alternate operationalizations of our posting similarity measure, and dimensionality of the latent space model. The estimation results as reported in the Online Appendix 6, Tables A7–A9, are consistent with our base results. The findings from our analysis uncover important insights about how users make friends on SNSs, and how the network, in turn, influences their content production behavior. Specifically, we show that users are more likely to make friends with users who show a similar level of posting behavior, as observed by the number of public posts. However, this homophilous behavior is short lived, and the users are found to diverge in their content production rates from their peers over time. Furthermore, our analysis shows that the propensity to form friendships based on homophily and the susceptibility to peer influence after forming the friendships are dependent on the current state of the behavior. Thus, users who are very active contributors on SNSs show very different peer effects compared with users who are less active on the SNSs.

### 5.1. Theoretical Implications

Using our coevolution perspective, we address two important theoretical gaps in the extant research on the evolution of online social networks and social behavior. First, we show that homophilous peer selection and peer influence might have varying strengths depending on the stage of network evolution. We find strong evidence of selection bias on the basis of homophily in content production; that is, individuals 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 influence effect. This suggests that the individuals might actively try to distinguish themselves from their friends in terms of their content production behavior. This is an interesting phenomenon, which would indicate that dynamic behaviors such as content production can influence network evolution in competing ways. While

this result is consistent with related findings in recent IS literature, like those of Zeng and Wei (2013), who show that there exists homophily based on similarity in photos uploaded to a large photo-sharing platform, as captured by the cosine similarity of photo tags, we generalize this result to a large general-purpose SNS and a more prevalent online behavior, that is, posting content on an SNS. This is among the first studies in IS to explicitly model both the degree of influence and homophily in a social network (e.g., Bramoullé et al. 2009, Aral et al. 2009, de Matos et al. 2014), and also builds on earlier attempts in the domain that focus on specialized contexts such as music sharing, photo sharing, and investor communities (Dewan et al. 2017, Zeng and Wei 2013, Gu et al. 2014) to offer a flexible framework for modeling the coevolution of network and behavior in a general online context with dynamic networks and fast-changing behaviors. Moreover, like Aral et al. (2009), who use a propensity score matching technique to control for unobserved heterogeneity, we address the same problem with the use of a latent space modeling approach. In the absence of a large set of available covariates, the latent space coordinates of the actors on a  $d$ -dimensional social space offers a good approximation of any latent homophily.

Second, we find that network effects such as homophilous selection and peer influence can increase or decrease in strength as a function of the magnitude of an individual's behavior. For instance, our results show that when compared with MoAPs, MAPs were less likely to form connections with other MAP users, while LAPs were more likely to form connections with other LAPs. Moreover, compared with MoAPs, both MAPs and LAPs were more susceptible to peer influence. Interestingly, while MAPs exhibited a positive influence (i.e., tendency to converge in behavior with peers), LAPs exhibited a negative influence (i.e., tendency to diverge in behavior from peers). Taken together, these results reveal an interesting pattern of how online social networks coevolve with the content produced on the SNS. These findings on the user heterogeneity in peer effects are consistent with a number of recent works in IS that emphasize how peer effects are often strengthened or weakened based on various individual-level differences. For instance, in Gu et al. (2014), the authors found that while investors showed greater interactions with other investors exhibiting similar sentiment in virtual investor communities (VICs), the strength of this homophily was significantly attenuated for investors with greater experience in the VICs. Similarly, in Singh et al. (2011), the authors highlight that while internal and external cohesion as well as technology diversity are key factors driving the success of open source projects, the degree centrality of project members (e.g., direct and indirect ties to other project

contacts) was also correlated with project success, since members with a higher number of direct and indirect ties could likely tap into external knowledge bases. In yet another study, Singh and Phelps (2013) highlight that the degree of prior experience of project managers attenuates their susceptibility to social influence in a new open source software project.

Third, and on the point of methodology, we draw attention to the concerns about latent homophily in inferring network influence, and demonstrate how we can draw on recent works in the area of latent space models to potentially control for homophily based on unobserved attributes. Our results using a two- and three-dimensional latent space modeling strategy show consistent estimates of homophily and influence. Thus, we contend that our stochastic structural model, together with the addition of latent-space covariates, can prove to be an effective empirical strategy for the study of peer influence and homophily in many other observational contexts where experiments are infeasible but large amounts of fine-grained user data might be available. This is consistent with recent studies that seek to quantify the extent of peer influence using large-scale observational data (e.g., Dewan et al. 2017, de Matos et al. 2014, Zeng and Wei 2013, Aral et al. 2009).

## 5.2. Practical Contributions

Understanding the nature of peer effects on SNSs has clear practical implications for several stakeholders. First, and most importantly, we offer a framework within which online user contributions can be studied as a function of the underlying network. Our research provides prescriptive guidance to platform owners and marketers on the dynamics and interplay of content production in connected communities. Second, our results provide intelligence to marketers to identify and better target valuable users on SNSs. For example, platforms like Facebook and Twitter can help improve friend recommendations and personalized content through customized “newsfeeds.” Specifically, our results suggest that it might not be a good idea to recommend heavy content posters as friends to other heavy posters, as such friendships tend to be detrimental to the content production of either of the friends; that is, high posters prefer other high posters in making friends, but reduce their posting rate over time after the friendship is created. Third, 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 social networks. Such predictive models for user behavior on dynamic networks can be invaluable not just to the platform owners, but also to advertisers and third-party marketers who wish to leverage social media for their own businesses.

## 5.3. Limitations and Future Work

As an initial attempt to model and analyze the coevolution of network structure and user behavior in online social networks, this study is prone to several limitations that offer opportunities for future research. First, and as mentioned earlier, the current paper focuses on providing a statistically sound method to uncover the dynamic peer effects in a social network. However, additional analyses are required to further separate out the specific rationale behind why individuals show such effects. Second, our current modeling approach requires computational resources to simulate the networks in each stage of the estimation procedure. This might be a concern, in terms of model specification and quality of model convergence, for extremely large networks of users and networks with high sparsity. Our model imposes a standard Markovian assumption on the data, which is reasonable in most cases but assumes there are no external factors that might influence the social network or the user behavior, such as natural disasters or other exogenous shocks. Last, we consider all friendships to be bidirectional or symmetric ties. While this is not a limitation in the present study, it could be useful to identify the directionality of friendship, that is, separate out in-degree from out-degree. While in-degree can be considered to be a measure of popularity, out-degree provides a better indication of SNS activity. Thus, by separating out the two effects, we will be able to investigate more complex social constructs in future studies.

## Acknowledgments

The authors are grateful to Sinan Aral, Dean Eckles, Donald B. Rubin, Tom Snijders, the editor, and the referees for comments and suggestions, and to a large, American social networking site for providing anonymized data for this study.

## Endnotes

<sup>1</sup>We list a summary of prior experimental and nonexperimental approaches to estimate influence and homophily in Online Appendix 1.

<sup>2</sup>There is no observed case of friendship dissolution (i.e., 1 to 0) in our data context.

## References

- Airoldi EM, Bai X, Carley KM (2011) Network sampling and classification: An investigation of network model representations. *Decision Support Systems* 51(3):506–518.
- Andreoni J (2007) Giving gifts to groups: How altruism depends on the number of recipients. *J. Public Econom.* 91(9):1731–1749.
- Ansari A, Koengsberg O, Stahl F (2011) Modeling multiple relationships in social networks. *J. Marketing Res.* 48(4):713–728.
- Aral S, Walker D (2014) Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Sci.* 60(6):1352–1370.
- Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* 106(51):21544–21549.
- Aral S, Muchnik L, Sundararajan A (2013) Engineering social contagions: Optimal network seeding in the presence of homophily. *Network Sci.* 1(2):125–153.



- Asch S (1951) Effects of group pressure upon the modification and distortion of judgments. Gueztzow H, ed. *Groups, Leadership and Men: Research in Human Relations* (Carnegie Press, Oxford, UK), 222–236.
- Azari SH, Airolidi EM (2012) Graphlet decomposition of a weighted network. Gordon G, Dunson D, Dudík M, eds. *Proc. 15th Internat. Conf. Artificial Intelligence Statistics, April 11–13, Fort Lauderdale, FL*, Vol. 22, 54–63.
- Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks. *Proc. 12th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM Press, New York), 44–54.
- Bai X, Padman R, Airolidi EM (2005) On learning parsimonious models for extracting consumer opinions. *Proc. 38th Annual Hawaii Internat. Conf. System Sciences*, Vol. 9 (IEEE Computer Society, Washington, DC), 1530–1605.
- Bampo M, Ewing MT, Mather DR, Stewart D, Wallace M (2008) The effects of the social structure of digital networks on viral marketing performance. *Inform. Systems Res.* 19(3):273–290.
- Binder J, Howes A, Sutcliffe A (2009) The problem of conflicting social spheres. *Proc. 27th Internat. Conf. Human Factors Comput. Systems* (ACM Press, New York), 965–974.
- Borgatti SP, Foster PC (2003) The network paradigm in organizational research: A review and typology. *J. Management* 29(6):991–1013.
- Bramoullé Y, Djebbari H, Fortin B (2009) Identification of peer effects through social networks. *J. Econometrics* 150(1):41–55.
- Braun M, Bonfrer A (2011) Scalable inference of customer similarities from interactions data using Dirichlet processes. *Marketing Sci.* 30(3):513–531.
- Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.
- Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *New England J. Medicine* 358(21):2249–2258.
- Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. *Proc. 14th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (ACM Press, New York), 160–168.
- Davin JP, Davin JP, Gupta S, Piskorski MJ (2014) Separating homophily and peer influence with latent space. Working Paper 14-053, Harvard Business School, Boston.
- de Bunt GG, Van Duijn MAJ, Snijders TAB (1999) Friendship networks through time: An actor-oriented dynamic statistical network model. *Comput. Math. Organ. Theory* 5(2):167–192.
- de Matos MG, Ferreira P, Krackhardt D (2014) Peer influence in the diffusion of iPhone 3G over a large social network. *MIS Quart.* 38(4):1103–1133.
- Dewan S, Ho YJ, Ramaprasad J (2017) Popularity or proximity: Characterizing the nature of social influence in an online music community. *Inform. Systems Res.* 28(1):117–136.
- Doreian P (1989) Network autocorrelation models: Problems and prospects. Griffith DA, ed. *Spatial Statistics: Past, Present, and Future* (Michigan Document Services, Ann Arbor), 369–389.
- Dunbar RIM (1992) Neocortex size as a constraint on group size in primates. *J. Human Evolution* 22(6):469–493.
- Durkheim E (1884) *The Division of Labor in Society* (Free Press, New York).
- Emirbayer M, Goodwin J (1994) Network analysis, culture, and the problem of agency. *Amer. J. Sociol.* 99(6):1411–1454.
- Ennett ST, Bauman KE (1994) The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. *J. Personality Soc. Psych.* 67(4):653–663.
- Fararo TJ, Sunshine MH (1964) A study of a biased friendship net. Technical report, Youth Development Center, Syracuse University, Syracuse, NY.
- Fogg BJ, Eckles D (2007) The behavior chain for online participation: How successful web services structure persuasion. *Proc. Second Internat. Conf. Persuasive Tech.*, Lecture Notes in Computer Science, Vol. 4744 (Springer-Verlag, Berlin), 199–209.
- Friedlkin NE (2001) Norm formation in social influence networks. *Soc. Networks* 23(3):167–189.
- Goh KY, Heng CS, Lin Z (2013) Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Inform. Systems Res.* 24(1):88–107.
- Goldfarb A, Tucker CE (2011) Privacy regulation and online advertising. *Management Sci.* 57(1):57–71.
- Goldsmith-Pinkham P, Imbens GW (2013) Social networks and the identification of peer effects. *J. Bus. Econom. Statist.* 31(3):253–264.
- Greenan CC (2015) Diffusion of innovations in dynamic networks. *J. Roy. Statist. Soc.: Ser. A (Statist. Soc.)* 178(1):147–166.
- Gu B, Konana P, Raghunathan R, Chen HM (2014) Research note—The allure of homophily in social media: Evidence from investor responses on virtual communities. *Inform. Systems Res.* 25(3):604–617.
- Han QC, Xu KS, Airolidi EM (2015) Consistent estimation of dynamic and multi-layer networks. Bach F, Blei D, eds. *Proc. 32nd Internat. Conf. Machine Learning, July 6–11, Lille, France*, Vol. 37, 1511–1520.
- Hargittai E (2007) Whose space? Differences among users and non-users of social network sites. *J. Comput.-Mediated Comm.* 13(1):276–297.
- Hof RD (2011) Advertisers flock to social networks. *MIT Tech. Rev.* (June 21), <http://www.technologyreview.com/article/424409/advertisers-flock-to-social-networks/>.
- Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* 97(460):1090–1098.
- Hollingshead AdB (1949) *Elmtown's Youth: The Impact of Social Classes on Adolescents* (Wiley, New York).
- Homans GC (1961) *Social Behavior: Its Elementary Forms* (Harcourt Brace & World, New York).
- Kirke DM (2004) Chain reactions in adolescents' cigarette, alcohol and drug use: Similarity through peer influence or the patterning of ties in peer networks? *Soc. Networks* 26(1):3–28.
- Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311(5757):88–90.
- Krackhardt D, Stern RN (1988) Informal networks and organizational crises: An experimental simulation. *Soc. Psych. Quart.* 51(2):123–140.
- Lazarsfeld PF, Merton RK, others (1954) Friendship as a social process: A substantive and methodological analysis. *Freedom Control Modern Soc.* 18(1):18–66.
- Lewis K, Gonzalez M, Kaufman J (2012) Social selection and peer influence in an online social network. *Proc. Natl. Acad. Sci. USA* 109(1):68–72.
- Livingstone S (2008) Taking risky opportunities in youthful content creation: Teenagers' use of social networking sites for intimacy, privacy and self-expression. *New Media Soc.* 10(3):393–411.
- Lu Y, Jerath K, Singh PV (2013) The emergence of opinion leaders in a networked online community: A dyadic model with time dynamics and a heuristic for fast estimation. *Management Sci.* 59(8):1783–1799.
- Macy MW, Kitts JA, Flache A, Benard S (2003) Polarization in dynamic networks: A Hopfield model of emergent structure. Breiger R, Carley K, Pattison P, eds. *Dynamic Social Network Modeling and Analysis* (National Academies Press, Washington, DC), 162–173.
- Maddala GS (1986) *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge University Press, Cambridge, UK).
- Mayer TF (1984) Parties and networks: Stochastic models for relationship networks. *J. Math. Sociol.* 10(1):51–103.
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annual Rev. Sociol.* 27(1):415–444.
- Nahon K, Hemsley J (2014) Homophily in the guise of cross-linking political blogs and content. *Amer. Behavioral Scientist* 58(10):1294–1313.
- Newcomb TM (1962) Student peer-group influence. Sanford N, ed. *The American College: A Psychological and Social Interpretation of the Higher Learning* (John Wiley & Sons Inc., Hoboken, NJ), 469–488.
- Newman M (2010) *Networks: An Introduction* (Oxford University Press, Oxford, UK).
- Oetting ER, Donnermeyer JF (1998) Primary socialization theory: The etiology of drug use and deviance. I. *Substance Use Misuse* 33(4):995–1026.



- Olson M (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard University Press, Cambridge, MA).
- Park J, Barabási AL (2007) Distribution of node characteristics in complex networks. *Proc. Natl. Acad. Sci. USA* 104(46): 17916–17920.
- Pearson M, West P (2003) Drifting smoke rings. *Connections* 25(2): 59–76.
- Phan TQ, Airolidi EM (2015) A natural experiment of social network formation and dynamics. *Proc. Natl. Acad. Sci. USA* 112(21): 6595–6600.
- Reingen PH, et al. (1984) Brand congruence in interpersonal relations: A social network analysis. *J. Consumer Res.* 11(3):771–83.
- Ripley RM, Snijders TAB, Boda Z, Vörös A, Preciado P (2018) Manual for RSiena. Department of Statistics, Nuffield College, University of Oxford, Oxford, UK.
- Robbins H, Monro S (1951) A stochastic approximation method. *Ann. Math. Statist.* 22(3):400–407.
- Schweinberger M, Snijders TAB (2007) Markov models for digraph panel data: Monte Carlo-based derivative estimation. *Comput. Statist. Data Anal.* 51(9):4465–4483.
- Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* 40(2):211–239.
- Sinclair JK, Vogus CE (2011) Adoption of social networking sites: An exploratory adaptive structuration perspective for global organizations. *Inform. Tech. Management* 12(4):293–314.
- Singh PV, Phelps C (2013) Networks, social influence, and the choice among competing innovations? Insights from open source software licenses networks, social influence, and the choice among competing innovations? Insights from open source software licenses. *Management Sci.* 24(3):539–560.
- Singh PV, Tan Y, Mookerjee V (2011) Network effects: The influence of structural capital on open source project success. *MIS Quart.* 35(4):813–829.
- Singla P, Richardson M (2008) Yes, there is a correlation—From social networks to personal behavior on the web. *Proc. 17th Internat. Conf. World Wide Web* (ACM Press, New York), 655–664.
- Snijders T, Steglich C, Schweinberger M (2007) Modeling the coevolution of networks and behavior. van Montfort K, Oud H, Satorra A, eds. *Longitudinal Models in the Behavioral and Related Sciences* (Lawrence Erlbaum, Mahwah, NJ), 41–71.
- Snijders TAB (1996) Stochastic actor-oriented models for network change. *J. Math. Sociol.* 21(1–2):149–172.
- Snijders TAB (2001) The statistical evaluation of social network dynamics. *Sociol. Methodology* 31(1):361–395.
- Steglich C, Snijders TAB, Pearson M (2010) Dynamic networks and behavior: Separating selection from influence. *Sociol. Methodology* 40(1):329–393.
- Steinfeld C, Ellison NB, Lampe C (2008) Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *J. Appl. Developmental Psych.* 29(6):434–445.
- Steinfeld C, DiMicco JM, Ellison NB, Lampe C (2009) Bowling online. *Proc. Fourth Internat. Conf. Communities Tech.* (ACM Press, New York), 245–254.
- Stokman FN, Doreian P, eds. (1997) Evolution of social networks: Processes and principles. *Evolution of Social Networks* (Gordon and Breach, Amsterdam), 233–250.
- Tucker C (2016) Social advertising: How advertising that explicitly promotes social influence can backfire. Working paper, Massachusetts Institute of Technology, Cambridge, MA.
- Valkenburg PM, Peter J, Schouten AP (2006) Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *Cyberpsych. Behav.* 9(5):584–90.
- Wang Y, Meister DB, Gray PH (2013) Social influence and knowledge management systems use: Evidence from panel data. *MIS Quart.* 37(1):299–313.
- Wasserman S (1977) Stochastic models for directed graphs. PhD dissertation, Harvard University, Cambridge, MA.
- Wasserman S (1980a) A stochastic model for directed graphs with transition rates determined by reciprocity. *Sociol. Methodology* 11:392–412.
- Wasserman S (1980b) Analyzing social networks as stochastic processes. *J. Amer. Statist. Assoc.* 75(370):280–294.
- Waters RD, Burnett E, Lamm A, Lucas J (2009) Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. *Public Relations Rev.* 35(2): 102–106.
- Wellman B, Haase AQ, Witte J, Hampton K (2001) Does the internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment. *Amer. Behavioral Scientist* 45(3):436–455.
- Yang J, Han QC, Airolidi EM (2013) Nonparametric estimation and testing of exchangeable graph models. Kaski S, Corander J, eds. *Proc. 17th Internat. Conf. Artificial Intelligence Statistics, April 22–25, Reykjavik, Iceland*, Vol. 33, 1060–1067.
- Zeng X, Wei L (2013) Social ties and user content generation: Evidence from Flickr. *Inform. Systems Res.* 24(1):71–87.
- Zhang X(M), Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.