Specification and Estimation of Network Formation and Network Interaction Models with the Exponential Probability Distribution

Chih-Sheng Hsieh[*], Lung-Fei Lee[†]

April 21, 2013

Abstract[1]

In this paper, we model network formation and network interactions under a unified framework. The key feature of our model is to allow individuals to respond to incentives stemming from interaction benefits on certain activities when they are choosing friends (network links). There are two advantages of this modeling approach: first, one can evaluate whether incentives from certain interactions are important factors for friendship formation or not. Second, possible friendship selection bias on network interactions can be corrected as the network formation is explicitly modeled. The proposed model is estimated by the Bayesian method. In the empirical study, we apply the model to American high school students' friendship networks in the Add Health dataset. From two activity variables which are considered in the paper – GPA and smoking frequency, we find a significant incentive effect from GPA, but

[*]Department of Economics, The Ohio State University, Columbus, Ohio, 43210

[†]Department of Economics, The Ohio State University, Columbus, Ohio, 43210

[1]A supplementary material for this paper is available online

not from smoking, on friendship formation. These results suggest that the benefit of interactions in academic learning is an important factor for forming friendships, while the pleasure of smoking together is not. However, from the perspective of network interactions, both GPA and smoking frequency are subject to significant positive interaction (peer) effects.

Keywords: Bayesian method, Multivariate Analysis, Spatial Analysis

1 Introduction

Economic research on social networks and social interactions has grown rapidly over the past two decades. For many economic issues, the role of social networks as a channel to disseminate information or facilitate activities is revealed.[2] Accompanying a wide application of network concepts in economics, an immediate question for both theorists and practitioners is to understand how networks are formed.[3] This question is not only interesting in its own right, but it is also important to analyze how changes of network structures alter interaction effects on economic activities. In the context of social interactions, regardless of whether the research subjects are groups of workers, students, or delinquents, one likes to know how individuals choose their friends in order to understand peer effects within a group. As a friendship net-

[2] For example, job finding and labor force participation (Calvó-Armengol and Jackson, 2004, 2007; Bayers *et al.*, 2008); social learning and knowledge diffusion (Conley and Udry, 2001, 2010); risk sharing and insurance (Fafchamps and Gubert, 2007a, 2007b); obesity transmission (Christakis and Fowler, 2007, Flower and Christakis, 2008); peer effects on students' academic achievement (Calvó-Armengol *et al.*, 2009); sport and club participation (Bramoullé *et al.*, 2009; Liu *et al.*, 2011); and juvenile delinquencies or criminal activities (Ballester *et al.*, 2010; Pattcchini and Zenou, 2008, 2012; Bayer *et al.*, 2009.)

[3] From the development of theory on network formation, the most recognized concept is strategic network formation proposed by Jackson and Wolinsky (1996). After their seminal work, theorists have widely applied this concept on building network formation models and discussed the tradeoff between network stability and efficiency (see survey in Jackson, 2008, 2009, and relevant chapters in Handbook of social economics edited by Benhabib, Bisin, and Jackson, 2011). Built on richness of theory, new empirical strategies to embed the concept of strategic network formation in real networks would be highly desirable for economic network studies. Empirical network studies are comparatively fewer than theoretical studies. Some existing examples include Fafchamps and Gubert (2007a, 2007b) and Comola (2008), which study the risk-sharing and insurance networks in rural areas of developing countries. Mayer and Puller (2008), Christakis *et al.* (2010), Currarini *et al.* (2010), Mele (2010), Goldsmith-Pinkham and Imbens (2011), and Hsieh and Lee (2011) study the friendship networks of American high school and college students.

work might be formed in order to achieve favorable economic consequences, when studying the result of network (or peer) effects on economic activities, there is a need to correct for possible endogeneity biases due to friendship selection. The choice of friendships might amplify observed peer interaction effects due to related unobserved factors behind both decisions of friendship and economic activities (Weinberg, 2008). With regard to this problem, Hsieh and Lee (2011) and Goldsmith-Pinkham and Imbens (2011) study possibly important unobserved driving factors and use them to link network formation and network interactions on economic activities.

In this paper, we propose a modeling approach for both static networks and interactions among individuals. A static network refers to a cross-sectional case in which only one observation of a network is available.[4] Our model introduces a device which allows individuals to respond to incentives stemming from interaction benefits on certain activities when making friendship decisions. This device is meaningful because in most of the empirical survey data for friendship, respondents are not asked to nominate friends for specific purposes such as "who are you studying with?" or "who are you exercising with?" Instead, respondents are asked to nominate friends in general, such as "who are your best friends?" or "who do you like to spend time with?" Therefore, it remains interesting to see from data which activities would provide significant incentives for forming friendships. The advantage of modeling both the network formation and network interactions under a unified framework is twofold: first, one can evaluate the importance of individuals' incentives as they stem from choosing their friends; second, the resulting model can correct possible friendship selection biases in activities with interactions. We apply this modeling approach to study American high school students' friendship networks

[4]We focus on a static setting because most of the available network data are cross-sectional ones without dynamics. Few students' friendship network data which have panel waves can be found in the literature of stochastic actor-based dynamic network modeling proposed by Snijders *et al.* (2010).

in the Add Health data. From two activity variables that are considered in the paper, a student's GPA and how frequently a student smokes in a usual week, we find a significant incentive effect from GPA but not from smoking, which suggests that the benefit of interactions in academic learning is a factor for building friendships, while the pleasure of smoking together is not. The estimated endogenous (peer) effect from our model is smaller than that from the network interaction model alone, with the network assumed to be exogenously given. The latter shows that our modeling approach is effective in correcting the selection bias in the network interaction model due to endogenous friendship selection.

To model static networks, one approach is to assume pairwise independence between network links. For example, Fafchamps and Gubert (2007a, 2007b) and Comola (2008) apply the pairwise independence assumption, which allows them to focus on individual and dyad-specific variables to explain network links. The estimation of these models can be done by a standard maximum likelihood approach since the likelihood of the whole network is simply the product of likelihoods from all pairwise links. However, as noted by Bramoullé and Fortin (2009), the assumption of pairwise independence is strong because it requires that the latent utility behind each pairwise link be separable.[5] Another approach to model static networks, without imposing the pairwise independence assumption, is to treat the observed network as a polychotomous choice with $2^{m(m-1)}$ alternatives made jointly by individuals, where m is the size of the network. The Exponential Random Graph Model (ERGM) proposed by Frank and Struss (1986), or more generally, the p^* model by Wasserman and Pattison (1996), are the models of this type. In either an ERGM or a p^* model, several selected network statistics, such as the number of reciprocal links, the number of k-stars, $k \geq 2$, are specified in an exponential probability distribution to capture how likely these network structures are to appear in a network. In this paper we

[5]This means that the individual utility derived from a network is equal to the sum of utilities from each of his/her link and each link utility is not affected by any other links in the network.

will follow the second approach aforementioned and motivate the model specification with economic reasoning.

The remainder of this paper is organized as follows. Section 2 presents a unified modeling approach for both network formation and network interactions. A Bayesian estimation method for the proposed model is discussed in Section 3. Section 4 includes an application of the model to high school students' friendship networks and activities in the Add Health data. Section 5 concludes the paper. We leave a simulation experiment for showing model identification in the appendix.

2 The models of network formation and network interactions

Our research subjects are individuals in a closed group setting, such as students in a school-grade or workers in a company. Let W_g be a $m_g \times m_g$ matrix (adjacency matrix; sociomatrix) representing the friendship network of m_g individuals (size) in group g, where $g = 1, \cdots, G$, with G denoting the total number of groups. The $(i,j)^{\text{th}}$ entry of W_g, denoted as $w_{ij,g}$, is a dichotomous indicator which equals one if individual i sends a link to individual j and zero, if not. The links are all directed without imposing reciprocality.[6] Diagonal elements, $w_{ii,g}$, $i = 1, \cdots, m_g$, are zeros, à priori. Let $x_{i,g}$ be a k-dimensional row vector containing individual i's exogenous characteristics and the $m_g \times k$ dimensional matrix X_g be a collection of such vectors in group g. For economic activities, we consider two types of variables – continuous and Tobit-type.[7] Let $y_{i,cg}$ ($y_{i,tg}$) denote individual i's continuous (Tobit-type) activity

[6]It is possible that individual i names j as his/her friend but j does not name i as a friend.

[7]We do not consider the case of binary variables in this paper because it might involve the issue of multiple equilibria if network interactions are based on observed binary variables (Krauth, 2006; Soetevent and Kooreman, 2007). The modeling of binary variables is of interest and challenging. We will leave it for future research.

variable in group g; $Y_{cg} = (y_{1,cg}, y_{2,cg}, \cdots, y_{m_g,cg})'$ and $Y_{tg} = (y_{1,tg}, y_{2,tg}, \cdots, y_{m_g,tg})'$ be the m_g-dimensional column vectors for all members' continuous and Tobit-type activity variables in group g.

2.1 Network interactions on continuous and Tobit-type activities

To model the process of network interactions on economic activities, we use the Spatial Autoregressive (SAR) model (Bramoullé *et al.*, 2009; Lee *et al.*, 2010; Lin, 2010). The SAR model for a continuous activity variable is specified as

$$Y_{cg} = \lambda_c W_g Y_{cg} + X_g \beta_{1c} + W_g X_g \beta_{2c} + l_g \alpha_{cg} + \epsilon_{cg}, \quad \epsilon_{cg} \sim \mathcal{N}_{m_g}(0, \sigma^2_{\epsilon_c} I_{m_g}), \tag{1}$$

$g = 1, \cdots, G$, where l_g is the m_g-dimensional vector of ones; α_{cg} represents the unobserved group fixed-effect for group g; $\mathcal{N}_{m_g}$ represents a multivariate normal distribution of dimension m_g and I_{m_g} is the identity matrix of dimension m_g. The coefficient λ_c in Eq. (1) is the endogenous (peer) effect, which is the key parameter of interest to us. As each entry of W_g takes the value of either 0 or 1 and is not row-normalized, λ should be interpreted as an influence from aggregated friendship links. The vector of coefficients, $\beta_c = (\beta'_{1c}, \beta'_{2c})'$, will capture influences from individuals' own and friends' exogenous characteristics, i.e., own and contextual effects, on Y_{cg}. Specifying group fixed effects in Eq. (1) will help to handle the identification problem caused by correlated effects. Moffitt (2001) argues that correlated unobservables in a group may contribute to correlations of Y_{cg} and cause an identification problem by confounding the endogenous effect. Here, group fixed effects not only refer to effects from unobserved environmental factors shared by all members in the same group, but also self-selection into groups if group choices depend on group characteristics.

The use of the SAR model to study network interactions can be motivated by individual utility maximization. Ballester *et al.* (2006) and Calvó-Armengol *et al.* (2009) consider that, given the network W_g, an individual chooses $y_{i,cg}$ to maximize a quadratic utility function

$$u_{i,cg}(Y_{cg}, W_g) = \left(x_{i,g}\beta_{1c} + \sum_{j=1}^{m_g} w_{ij,g}x_{j,g}\beta_{2c} + \alpha_{cg} + \epsilon_{i,cg} \right) y_{i,cg} - \frac{1}{2}y_{i,cg}^2 + \lambda_c y_{i,cg} \sum_{j=1}^{m_g} w_{ij,g}y_{j,cg}, \quad (2)$$

for $i = 1, \cdots, m_g$, in a simultaneous non-cooperative game. The first and second terms show that the utility is concave in individuals' own choice of activity, which guides them to choose their optimum activities. The third term reflects a complementary effect (competitive effect) from peers' choices of activities if $\lambda \geq 0$ ($\lambda \leq 0$). By the theorem of Ballester *et al.* (2006), as long as $|\lambda_c|$ is less than the largest eigenvalue of W_g, the unique interior Nash equilibrium outcome vector will take the form as in Eq. (1) and, correspondingly, an individual optimum utility from network interactions, given the network W_g, will be $u_{i,cg}(y_{i,cg}(W_g)) = \frac{1}{2}y_{i,cg}^2(W_g)$.

In certain cases, activity variables might be continuous, but nonnegative, i.e., a Tobit-type variable which is left-censored at the value zero. Such a variable may occur when individual i maximizes the utility in Eq. (2) with $y_{i,cg}$ subject to a nonnegative constraint. To distinguish a Tobit-type activity variable from a continuous one, we replace $y_{i,cg}$ with $y_{i,tg}$ and impose a constraint, $y_{i,tg} \geq 0$, on the individual utility of Eq (2). Under this constraint, the Nash equilibrium vector can be summarized by the following equation:

$$Y_{tg}(W_g) = \max\left(0, \ddot{Y}_{tg}\right) \quad \text{with} \quad \ddot{Y}_{tg} = \lambda_t W_g Y_{tg} + \mathbf{X}_g\beta_t + l_g\alpha_{tg} + \epsilon_{tg}, \quad (3)$$

where $\ddot{Y}_{tg}$ represents a vector of latent variables, $\mathbf{X}_g = (X_g, W_g X_g)$, and $\beta_t = (\beta_{1t}', \beta_{2t}')'$. The solution $Y_{tg}(W_g)$ must satisfy $Y_{tg} \geq \lambda_t W_g Y_{tg} + \mathbf{X}_g\beta_t + l_g\alpha_{tg} + \epsilon_{tg}$, such that $Y_{tg}(W_g) \geq 0$, and $y_{i,tg} = \lambda_t \sum_{j=1}^{m_g} w_{ij,g}y_{j,tg} + x_{i,g}\beta_{1t} + \sum_{j=1}^{m_g} w_{ij,g}x_{j,g}\beta_{2t} + \alpha_{tg} + \epsilon_{i,tg}$ whenever $y_{i,tg}(W_g) > 0$ for $i = 1, \cdots, m_g$. Under the conditions as in Amemiya (1974) for a general simultaneous Tobit equation system,[8] the solution

[8] A sufficient condition for a unique solution of this quadratic programming problem is that the

$Y_{tg}(W_g)$ is unique and can be obtained from a constrained quadratic programming problem

$$Y_{tg}(W_g) = \min_{Y_{tg}} \left\{ \begin{array}{l} Y'_{tg}[(I_{m_g} - \lambda_t W_g)Y_{tg} - \mathbf{X}_g\beta_t - l_g\alpha_{tg} - \epsilon_{tg}] : \\ Y_{tg} \geq 0, (I_{m_g} - \lambda_t W_g)Y_{tg} - \mathbf{X}_g\beta_t - l_g\alpha_{tg} - \epsilon_{tg} \geq 0 \end{array} \right\}. \quad (4)$$

As an alternative, we show that with proper restricted parameter space on λ_t, namely, $||\lambda W_g||_\infty < 1$, where $||.||_\infty$ is the maximum row sum norm, the solution can be conveniently obtained via a contraction mapping algorithm provided in Appendix A.

One concern of using the SAR model in studying network interactions is the possible endogeneity of the weight matrix, W_g. If W_g is endogenous and it correlates with the disturbance term ϵ_g of the model, estimating the SAR model will result in biases on the estimated endogenous effect, as well as other effects. A standard Instrumental Variable (IV) approach suggests finding instruments for the endogenous weight matrix. However, without utilizing the information provided by the structure of W_g or its formation process, an effective instrument may be difficult to find. As a result, we do not pursue an IV approach, but instead propose a modeling approach which unifies the SAR model with a network formation model. The network formation model is represented by the ERGM with the specified network statistics motivated from economic reasoning. The key to combining these two models, which will be illustrated in the next section, is to allow individuals to consider potential benefits, which they can earn from the interaction process, when choosing friends.

quadratic objective function is strictly concave, which will be guaranteed if $I_{m_g} - \frac{\lambda_t}{2}(W_g + W'_g)$ is positive definite. A necessary and sufficient condition is every principle minor of $(I_{m_g} - \lambda_t W_g)$ is positive. Another sufficient condition is that $(I_{m_g} - \lambda_t W_g)$ has positive dominant diagonals, i.e., there exists positive d_i, $i = 1, \cdots, m_g$ such that $d_i > |\lambda_t| \sum_{j \neq i}^{m_g} |W_{ij,g}| d_j$ for all $i = 1, \cdots, m_g$.

2.2 Network formation with incentives from interaction benefits

To model the formation of static networks, we consider the exponential distribution framework used in the ERGM due to its capacity in explaining link dependencies within a network as well as its computational tractability in empirical applications. A network space, Ω_g, for group g consists of any possible network pattern, W, for that group. An exponential distribution for W_g has a probability specification in the following formula:

$$P(W_g) = \frac{\exp(Q(W_g))}{\sum_{W \in \Omega_g} \exp(Q(W))}. \tag{5}$$

A specification of the function, $Q(W_g)$, can accommodate various network statistics, and thus Eq. (5) allows for an arbitrarily general probability specification. But in order to apply this type of a model in empirical studies, researchers should specify the dependencies sparingly so that the resulting probability is simple and practical (Jackson, 2010). In standard ERGMs (Frank and Strauss, 1986; Wasserman and Pattison, 1996; Snijders, 2002), network statistics such as the number of k-stars, $k \geq 2$, and triangles are used in $Q(W_g)$ to measure how likely those network structures are to appear in observed networks. However, the coefficients of those network statistics do not represent causal relationships. To handle this drawback, we propose including network statistics in $Q(W_g)$ with economic reasoning.[9]

We consider that each individual i, $i = 1, \cdots, m_g$, obtains the following utility from network links in W_g:

$$v_{i,g}(W_g) = \underbrace{\sum_{j=1}^{m_g} w_{ij,g}\psi_{ij,g}}_{\text{Exogenous Effects}} + \underbrace{\varpi_{i,g}(w_{i.,g}, W_{-i.,g})\eta}_{\text{Network Structure Effects}} + \underbrace{\sum_{d=1}^{\bar{d}} \frac{\delta_d}{2} y_{i,dg}^2(W_g)}_{\text{Incentive Effects}}. \tag{6}$$

[9] A similar idea is used in Snijders *et al.* (2010) and Mele (2010) for their dynamic network formation models.

In Eq. (6), the exogenous effects capture influences from individual-specific and dyad-specific exogenous characteristics on the link utility. The function, $\psi_{ij,g}$, has the expression,

$$\psi_{ij,g} = c_{i,g}\gamma_1 + c_{j,g}\gamma_2 + c_{ij,g}\gamma_3. \tag{7}$$

The variables, $c_{i,g}$ and $c_{j,g}$, in Eq. (7) are $\bar{s}$-dimensional row vectors of individual-specific characteristics and the variable, $c_{ij,g}$, is a $\bar{q}$-dimensional row vector of dyad-specific characteristics, such as the same age, sex, or race shared by each pair of individuals (i, j) in group g to capture the utility from homophily of observed characteristics in friendship formation. The idea of using $C_g = [(c_{i,g}, c_{j,g}, c_{ij,g})'; i = 1, \cdots, m_g, j = 1, \cdots, m_g, i \neq j]'$ and coefficients, $\gamma = (\gamma_1', \gamma_2', \gamma_3')'$, in explaining link decisions is general (see e.g., Fafchamps and Gubert (2007a, 2007b) on the study of risk-sharing network formation). The network structure effects in Eq. (6) capture influences from some patterns of link dependence on the link utility, where $\varpi_{i,g}(w_{i\cdot,g}, W_{-i\cdot,g})$ represents a $\bar{h}$-dimensional row vector of summary statistics constructed from W_g which are relevant to individual i's utility and η is a corresponding vector of coefficients. The empirical specification of the network structure effects used in this paper will be discussed later in Section 4.1.

The incentive effects are innovative in this paper, which represent the benefits from network interactions, i.e., utilities obtained from the interaction process. For example, students may want to make friends with someone who is doing well in school in order to learn from him or her to improve their own performance. These incentive effects are the key to linking the network formation and network interaction models. We implicitly assume that individuals make their decisions on friendship links and economic activities sequentially in two stages. In the first stage, individuals choose their friends to maximize their utilities from link formation. In the second stage, individuals choose economic activities to maximize their utilities from network interactions. This two-stage process is characterized as a two-stage static game. Individuals adopt strategies on choosing friends and economic activities and obtain utilities as payoffs of the game. There is perfect information between the two stages. The equilibrium of this two-stage game satisfies the principle of sequential rationality, i.e., a player's strategy should specify optimal actions at every point in the game tree (Mas-Colell *et al.*, 1995). Hence, one can solve the equilibrium of this game by backward induction: first, determine equilibrium activities in the second stage and calculate corresponding optimum utilities from network interactions for each possible network pattern. Second, by incorporating the optimum utilities from network interactions into the utilities of links in the first stage, solve for the equilibrium network.

There may be several ($\bar{d}$) economic activities which provide incentives for forming friendships. For simplicity, these incentive effects are assumed to be separable, i.e., none of them affect each other. As noted by Ballester *et al.* (2006), utilities from network interactions always increase with the number of links in the network if interactions provide complementary effects on activities. Since the utility from network links contains incentive effects, individuals might choose to add as many links as possible if there were no cost on link formation. This is also related to the problem of network degeneracy as discussed in Snijders *et al.* (2006). To mitigate such a strong incentive to form links, we rely on nontrivial negative effects from some exogenous or network structural effects to represent possible costs of forming friendship links.[10]

[10]The activity variables, $\{y_{i,dg}\}_{d=1}^{\bar{d}}$, enter into the utility of linking through incentive effects. It is possible to allow $y_{i,dg}$, $y_{j,dg}$ (or $|y_{j,dg} - y_{i,dg}|$) to directly appear in $\psi_{ij,g}$ for capturing individual-specific (or dyad-specific) effects. Such an extension will emphasize that activity variables (or the absolute difference of activity variables) directly affect the utility from network links. For example, one may consider activities which are usually done by individuals alone, e.g., watching TV and playing video games. The more time students spend on those activities, the less time they can spend on associating with friends. Hence, $y_{i,dg}$ and $y_{j,dg}$, which denote the time individual i and j spend on the activity, should be specified in the function, $\psi_{ij,g}$, to capture the influences on the

To motivate the specification of $Q(W_g)$ in Eq. (5) from the individual utility function $v_{i,g}$ in Eq. (6), we assume that a network is formed through a cooperative game. Cooperative behavior in friendship formation is argued by economists and biologists with the theory from iterated prisoner's dilemma and cooperative strategies (i.e., tit-for-tat) (See Peck, 1993; Hruschka and Henrich, 2006; Majolo *et al.*, 2006; Ule 2008; Fu *et al.*, 2008; and Fosco and Mengel, 2011). Also, Jackson (2010) indicates that studying the allocation rules behind the cooperative network formation game is rational in many economic applications, such as students' friendship network in schools, where the favors can be exchanged between students. We can define the Transferable Utility (TU) of this cooperative network formation game as the sum of individual utility function, which is

$$\begin{aligned}\mathrm{TU}(W_g) &= \sum_{i=1}^{m_g} v_{i,g}(W_g) \\ &= \sum_{i=1}^{m_g}\sum_{j=1}^{m_g} w_{ij,g}\psi_{ij,g} + \sum_{i=1}^{m_g} \varpi_{i,g}(w_{i\cdot,g}, W_{-i\cdot,g})\eta + \sum_{d=1}^{\bar{d}} \frac{\delta_d}{2} Y_{dg}(W_g)' Y_{dg}(W_g). \quad (8)\end{aligned}$$

To relate $\mathrm{TU}(W_g)$ to the specification of $Q(W_g)$ in the exponential distribution framework, we introduce a disturbance ξ_W for each network pattern W in Ω_g additively to $\mathrm{TU}(W)$. Thus, W_g is the formed network if and only if $\mathrm{TU}(W_g) + \xi_{W_g} = \max_{W\in\Omega_g}\{\mathrm{TU}(W) + \xi_W\}$. The disturbances can be regarded as some stochastic elements in the network formation process. By assuming that ξ_W's are from i.i.d. type I extreme value distribution, we have the exponential probability in Eq. (5)

link utility. However, those activities can be still subject to friendship interactions, as friends may share information of new games or TV programs and therefore individuals may spend even more time on those activities. In another example, we can consider the frequency of delinquent behaviors as the activity variable. Students care about $|y_{j,dg} - y_{i,dg}|$ in forming friendships as differences in levels of their delinquent behaviors could create negative effects on their probability of linking. In terms of estimation, having $y_{i,dg}$, $y_{j,dg}$ or $|y_{j,dg} - y_{i,dg}|$ directly affect the probability of linking will not cause substantial changes in the Bayesain approach and corresponding MCMC algorithms proposed in this paper.

with the function $Q(W_g)$ being replaced by $\text{TU}(W_g)$ and thus it relates our model specification to the ERGM framework.

Modeling the endogenous network formation and activity variables jointly under a unified framework has two advantages. First, it allows us to study how individuals respond to incentives from network interactions when choosing their friends, which are revealed by the coefficients δ_d's. Second, it handles the problem of friendship selection bias on the interaction effects. The disturbance terms ϵ_{dg}'s appear in both the network formation and network interaction processes. Hence, they capture unobserved factors which contribute to both friendship and economic activity decisions. In the following section, we will discuss how to estimate this model.

3 Model estimation

3.1 The likelihood function of the model

To illustrate how the likelihood function of our model is calculated, we first consider models with a single incentive effect from either a continuous or a Tobit-type activity variable. Then, we introduce a correlation between disturbances of activity variables in these two types and incorporate them into the model. The joint likelihood function based on this bivariate case will be used for the posterior analysis in Section 3.2.

Continuous activity variable

We assume an individual idiosyncratic shock $\epsilon_{i,cg}$ in Eq. (2) is i.i.d. distributed from $\mathcal{N}(0, \sigma^2_{\epsilon_c})$. With the incentive effect from the continuous variable Y_{cg}, the joint probability of the activity variable Y_{cg} and the network W_g is

$$\begin{aligned} P(W_g, Y_{cg}|\theta_c, \alpha_{cg}) &= P(Y_{cg}|W_g, \theta_c, \alpha_{cg}) \cdot P(W_g|\theta_c, \alpha_{cg}) \\ &= |S_{cg}(W_g)| \cdot f(\epsilon_{cg}|W_g, \theta_c, \alpha_{cg}) \cdot P(W_g|\theta_c, \alpha_{cg}) \\ &= |S_{cg}(W_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot P(W_g|\epsilon_{cg}, \theta_c, \alpha_{cg}) \\ &= |S_{cg}(W_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot \frac{\exp(\mathrm{TU}(W_g, \epsilon_{cg}, \theta_c, \alpha_{cg}))}{\sum_W \exp\left(\mathrm{TU}(W, \epsilon_{cg}, \theta_c, \alpha_{cg})\right)}, \end{aligned} \tag{9}$$

where

$$f(\epsilon_{cg}|\theta_c, \alpha_{cg}) = (2\pi)^{-\frac{m_g}{2}} \left(\sigma^2_{\epsilon_c}\right)^{-\frac{m_g}{2}} \exp\left(-\frac{1}{2\sigma^2_{\epsilon_c}} \epsilon'_{cg}\epsilon_{cg}\right),$$

with $\epsilon_{cg} = S_{cg}(W_g)Y_{cg} - \mathbf{X}_g\beta_c - l_g\alpha_{cg}$ and $\theta_c = (\gamma', \eta', \delta_c, \lambda_c, \beta'_c, \sigma^2_{\epsilon_c})$ being the parameter vector.

Tobit-type activity variable

For the Tobit-type activity variable Y_{tg}, we can divide the m_g individuals in group g into two blocks, such that the first m_{g1} individuals have activity variables equal to zero and the remaining individuals who are arranged from $m_{g1} + 1$ to m_g have positive activity variables. Eq. (3) of the activity variable Y_{tg} and the network W_g can be conformably decomposed into

$$\begin{aligned} \begin{pmatrix} \ddot{Y}_{tg1} \\ Y_{tg2} \end{pmatrix} = \lambda_t \begin{pmatrix} W_{11,g} & W_{12,g} \\ W_{21,g} & W_{22,g} \end{pmatrix} \begin{pmatrix} Y_{tg1} \\ Y_{tg2} \end{pmatrix} + \begin{pmatrix} X_{1g} \\ X_{2g} \end{pmatrix} \beta_{1t} \\ + \begin{pmatrix} W_{11,g} & W_{12,g} \\ W_{21,g} & W_{22,g} \end{pmatrix} \begin{pmatrix} X_{1g} \\ X_{2g} \end{pmatrix} \beta_{2t} + \begin{pmatrix} l_{g1} \\ l_{g2} \end{pmatrix} \alpha_{tg} + \begin{pmatrix} \epsilon_{tg1} \\ \epsilon_{tg2} \end{pmatrix}, \end{aligned} \tag{10}$$

where $Y_{tg2} > 0$ and $Y_{tg1} = 0$ with the corresponding latent variables, $\ddot{Y}_{tg1} \leq 0$. Individual idiosyncratic shocks, $\epsilon_{i,tg}$'s, are assumed from i.i.d. $\mathcal{N}(0, \sigma^2_{\epsilon_t})$. Based on

Eq. (10), the probability function of Y_{tg} and W_g can be written as

$$
\begin{aligned}
&P(Y_{tg}, W_g|\theta_t, \alpha_{tg}) \\
&= P(Y_{tg1} = 0, Y_{tg2}, W_g|\theta_t, \alpha_{tg}) \\
&= \int I(Y_{tg1} = 0, \ddot{Y}_{tg1}) \cdot P(\ddot{Y}_{tg1}, Y_{tg2}, W_g|\theta_t, \alpha_{tg}) \cdot d\ddot{Y}_{tg1} \\
&= \int_{-\infty}^{-(\lambda_t W_{12,g} Y_{tg2} + X_{1g}\beta_{1t} + (W_{11,g}X_{1g} + W_{12,g}X_{2g})\beta_{2t})} \left|I_{m_g - m_{g1}} - \lambda_t W_{22,g}\right| \cdot f(\epsilon_{tg1}, \epsilon_{tg2}|\theta_t, \alpha_{tg}) \cdot \\
&\qquad P(W_g|\epsilon_{tg1}, \epsilon_{tg2}, \theta_t, \alpha_{tg}) \cdot d\epsilon_{tg1} \\
&= \int_{-\infty}^{-(\lambda_t W_{12,g} Y_{tg2} + X_{1g}\beta_{1t} + (W_{11,g}X_{1g} + W_{12,g}X_{2g})\beta_{2t})} \left|I_{m_g - m_{g1}} - \lambda_t W_{22,g}\right| \cdot f\left(\epsilon_{tg1}, \epsilon_{tg2}|\theta_t, \alpha_{tg}\right) \cdot \\
&\qquad \frac{\exp\left(\mathrm{TU}(W_g, \epsilon_{tg1}, \epsilon_{tg2}; \theta_t, \alpha_{tg})\right)}{\sum_W \exp\left(\mathrm{TU}(W, \epsilon_{tg1}, \epsilon_{tg2}; \theta_t, \alpha_{tg})\right)} \cdot d\epsilon_{tg1}, \qquad (11)
\end{aligned}
$$

where $I(Y_{tg1} = 0, \ddot{Y}_{tg1})$ is a dichotomous indicator which is equal to 1 when $\ddot{Y}_{tg1}$ is negative, and equal to 0, otherwise; $\epsilon_{tg2} = \left(I_{m_g - m_{g1}} - \lambda_t W_{22,g}\right) Y_{tg2} - X_{2g}\beta_{1t} - (W_{21,g}X_{1g} + W_{22,g}X_{2g})\beta_{2t} - l_{2g}\alpha_{tg}$ and $\theta_t = (\gamma', \eta', \delta_t, \lambda_t, \beta_t', \sigma_{\epsilon_t}^2)$.

Incentive effects can be from a total of $\bar{d}$ activity variables which consist of mixed continuous and Tobit-type ones. For simplicity, we consider a model with one continuous and one Tobit-type activity variable, where the disturbances, $\epsilon_{i,tg}$ and $\epsilon_{i,cg}$, follow a bivariate normal distribution,

$$
(\epsilon_{i,tg}, \epsilon_{i,cg}) \sim i.i.d.\ \mathscr{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\epsilon_t}^2 & \sigma_{\epsilon_{tc}} \\ \sigma_{\epsilon_{ct}} & \sigma_{\epsilon_c}^2 \end{pmatrix} \right), \quad i = 1, \cdots, m_g. \qquad (12)
$$

From Eq. (12), one has

$$
\epsilon_{tg} = \sigma_{\epsilon_{tc}} \sigma_{\epsilon_c}^{-2} \epsilon_{cg} + u_g, \quad u_g \sim \mathscr{N}_{m_g}(0, \sigma_u^2 I_{m_g}), \qquad (13)
$$

where $\sigma_u^2 = (\sigma_{\epsilon_t}^2 - \sigma_{\epsilon_{tc}} \sigma_{\epsilon_c}^{-2} \sigma_{\epsilon_{ct}})$. Let $\theta_{ct} = (\gamma', \eta', \delta_c, \delta_t, \lambda_c, \lambda_t, \beta_c', \beta_t', \sigma_{\epsilon_c}^2, \sigma_{\epsilon_t}^2, \sigma_{\epsilon_{tc}})$, the joint probability function of Y_{tg}, Y_{cg}, and W_g is

$$P(Y_{tg}, Y_{cg}, W_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg})$$
$$= \int_{-\infty}^{-(\lambda_t W_{12,g} Y_{t2g} + X_{1g}\beta_{1t} + (W_{11,g}X_{1g} + W_{12,g}X_{2g})\beta_{2t})} \left| I_{m_g - m_{g1}} - \lambda_t W_{22,g} \right| \cdot f\left(u_g|\epsilon_{cg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}\right) \cdot$$
$$|S_{cg}(W_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot \frac{\exp(\text{TU}(W_g, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}))}{\sum_W \exp\left(\text{TU}(W, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg})\right)} \cdot d\epsilon_{tg1}. \quad (14)$$

If ϵ_{tg} and ϵ_{cg} are uncorrelated, i.e., $\sigma_{\epsilon_{tc}} = \sigma_{\epsilon_{ct}} = 0$, then

$$P(Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg})$$
$$= \int_{-\infty}^{-(\lambda_t W_{12,g} Y_{t2g} + X_{1g}\beta_{1t} + (W_{11,g}X_{1g} + W_{12,g}X_{2g})\beta_{2t})} \left| I_{m_g - m_{g1}} - \lambda_t W_{22,g} \right| \cdot f\left(\epsilon_{tg}|\theta_t, \alpha_{tg}\right) \cdot$$
$$|S_{cg}(W_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot \frac{\exp(\text{TU}(W_g, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}))}{\sum_W \exp\left(\text{TU}(W, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg})\right)} \cdot d\epsilon_{tg1}. \quad (15)$$

The main issue we will encounter during the estimation is to calculate the likelihood function of the exponential distribution for the network. When the network size is large, its calculation is almost impossible since it requires evaluating all network patterns in Ω_g for the denominator of the exponential distribution function.[11] Hence, the standard maximum likelihood estimation approach (without simulation) would be unfeasible. This problem applies to all ERGMs for networks and can be traced back to the spatial analysis in Besag (1974). To deal with this problem, we turn to the Bayesian estimation with an effective MCMC technique (discussed in Section 3.3) developed to handle an intractable normalizing term in the posterior density function.[12]

[11]For example, even in a network with just 5 individuals, it needs to evaluate $2^{4\times5} = 2^{20}$ possible network realizations for the denominator.

[12]For the classical approach, several estimation methods have been proposed. The first is the maximum pseudo-likelihood approach (MPL). This approach was first mentioned in Besag (1974) and later applied to the network study in Strauss and Ikeda (1990). A pseudo-likelihood simply uses the product of conditional probabilities for estimation. The estimates from the MPL would not be the MLE. One may use the estimates from the MPL as initial values for other estimation

Regarding identification of parameters in our models, one may focus on the coefficients of incentive effects, δ_d's, in the network formation model, which link to the activity variables and the endogenous effects, λ_d's, in the network interaction models.[13] To show that these two sets of parameters are identified in Bayesian theory, i.e., data brings information to update the posterior distributions of parameters which distinguish them from the prior distributions (see Kadane, 1974; Hsiao, 1983; Poirier and Tobias, 2003), we conduct a simulation experiment which shows that the posterior distributions of these parameters collapse to the true values when the sample size increases. The details of this experiment are left in Appendix B.

3.2 Posterior distributions of parameters and the MCMC

The posterior distributions of parameters considered here are based on the model with both continuous and Tobit-type activity variables. To deal with Tobit-type activity variables in the Bayesian approach, it is natural to include the sampling of latent variables, $(\ddot{Y}_{t11}, \cdots, \ddot{Y}_{tG1})$, during the MCMC procedure along with other unobservables as an augmentation (Albert and Chib, 1993). By Bayes' theorem, the

approaches. Another approach is the Monte Carlo maximum likelihood (MCML) estimation approach which simulates auxiliary networks for approximating the denominator of the exponential distribution function with its simulated counterpart (Geyer and Thompson, 1992). One shortcoming of the MCML approach is that the choice of initial values during the optimization algorithm plays a critical role. They have to be close enough to the true parameter values, otherwise, the convergence of the algorithm might not be attained (Bartz *et al.*, 2008; Caimo and Friel, 2010). The Robbins-Monro approach used in Snijders (2002) to simulate auxiliary networks for constructing simulated moments usually accepts a wide range of initial values which will lead to a convergent algorithm.

[13]The rest of the parameters in the network formation model will be identified as long as corresponding regressors are not linearly dependent. So are the coefficients of regressors in the network interaction models.

joint posterior distribution of the parameters and unobservables in the model is[14]

$$
\begin{aligned}
& P\left(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\}, \{\ddot{Y}_{tg1}\} | \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}\right) \\
& \propto \pi(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\}) \cdot \\
& \quad \prod_{g=1}^{G} \left\{ \left(\prod_{i=1}^{m_{g1}} I(y_{i,tg} = 0) \cdot I(\ddot{y}_{i,tg} \leq 0) \right) \cdot P\left(Y_{tg}, Y_{cg}, W_g, \ddot{Y}_{tg1} | \theta_{ct}, \alpha_{cg}, \alpha_{tg}\right) \right\}, \quad (16)
\end{aligned}
$$

where $\pi(\cdot)$ represents the density function of the prior distribution. The exogenous variables, $\{X_g\}$ and $\{C_g\}$, are suppressed from the above expression for simplicity. We assume independence between prior distributions of common parameters and group effects, namely, $\pi(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\}) = \pi_1(\theta_{ct})\pi_2(\{\alpha_{cg}\})\pi_3(\{\alpha_{tg}\})$. It is not easy to directly simulate draws from the joint posterior density in Eq. (16). But one can use the Gibbs sampling algorithm and work on the conditional posterior densities of parameters. By properly blocking parameters in θ_{ct} into subgroups, we define prior distributions for parameters and group effects in the model as follows:

(1) Coefficients of network formation model,

$$\phi = (\gamma', \eta', \delta_c, \delta_t) \sim \mathscr{TN}_{2\bar{s}+\bar{q}+\bar{h}+2}(\phi_0, \Phi_0).$$

(2) Endogenous interaction parameters in network interaction models,

$$\lambda_c, \lambda_t \sim U[-1/\tau_G, 1/\tau_G].$$

(3) Coefficients of own and contextual effects in network interaction models,

$$\beta_c, \beta_t \sim \mathscr{N}_{2k}(\beta_0, B_0).$$

(4) Variances and covariance of disturbance in network interaction models,

$$\sigma = (\sigma^2_{\epsilon_c}, \sigma^2_{\epsilon_t}, \sigma_{\epsilon_{ct}}) \sim \mathscr{TN}_3(\sigma_0, \Sigma_0),$$

(5) Group fixed effects in network interaction models,

$$\alpha_{cg}, \alpha_{tg} \sim \mathscr{N}(\alpha_0, A_0), \quad g = 1, \cdots, G.$$

[14] We use the notation $\{A_g\}$ to represent the collection of A_g across G groups, i.e., $\{A_g\} := (A_1, \cdots, A_G)$.

In the prior distributions of (1) and (4) above, $\mathscr{T}\mathscr{N}_q$ represents a truncated multivariate normal distribution of dimension q. These prior distributions, except for λ_c and λ_t, are conjugate priors commonly used in the Bayesian literature. We assign γ, η, δ_c and δ_t into the group, ϕ, since they are all (linear) coefficients in the function, $\mathrm{TU}(W_g)$. We require the incentive effects, δ_c and δ_t, to be nonnegative. This constraint helps us to rule out the case of negative incentive effects, which is not reasonable for the utility specification. Thus, the prior distribution of ϕ is a truncated normal, which is defined on the convex area, $O = \{\phi \in \mathbb{R}^{2\bar{s}+\bar{q}+\bar{h}+2} | \delta_c \geq 0, \delta_t \geq 0\}$, with ϕ_0 and Φ_0 being the prior mean vector and the variance matrix before truncation, respectively. For λ_c and λ_t, their prior distributions are independent and we employ a uniform prior for each as suggested in Smith and LeSage (2002). We restrict the valid values of λ_c and λ_t between $-1/\tau_G$ and $1/\tau_G$, where $\tau_G = \max\{\tau_1^*, \cdots, \tau_G^*\}$ and $\tau_g^* = \min\{\max_{1\leq i\leq m_g} \sum_{j=1}^{m_g} |w_{ij,g}|, \max_{1\leq j\leq m_g} \sum_{i=1}^{m_g} |w_{ij,g}|\}$.[15] We put $\sigma_{\epsilon_c}^2$, $\sigma_{\epsilon_t}^2$, and $\sigma_{\epsilon_{ct}}$ into a group, σ, and specify a truncated distribution for σ to the area, $T = \{\sigma \in \mathbb{R}^3 | \sigma_{\epsilon_c}^2 > 0, \sigma_{\epsilon_t}^2 > 0, \sigma_{\epsilon_c}^2 \sigma_{\epsilon_t}^2 - \sigma_{\epsilon_{ct}}^2 \geq 0\}$, so that $\sigma_{\epsilon_c}^2$, $\sigma_{\epsilon_t}^2$, and $\sigma_{\epsilon_{ct}}$ can form a proper covariance matrix. The group effects, α_{cg} and α_{tg}, are treated as fixed effects and therefore the hyperparameters, α_0 and A_0, are fixed in their prior distributions.[16] Within the Gibbs sampling steps, random draws can be simulated from the conditional posterior distribution for each of the parameter groups. Here we list the set of conditional posterior distributions required by the Gibbs sampler:

(i) $P\left(\ddot{Y}_{tg1} \middle| \theta_{ct}, \alpha_{cg}, \alpha_{tg}, Y_{cg}, Y_{tg}, W_g\right)$, $g = 1, \cdots, G$.

[15] This interval is suggested by Kelejian and Prucha (2010) in which $I_{m_g} - \lambda W_g$ is nonsingular for all values of λ in this interval.

[16] If α_{cg} and α_{tg} are treated as random group effects, we should assign them with hierarchical priors, which means that the parameters, α_0 and A_0, in their prior distributions also have their own priors and should be updated with data.

By applying Bayes' theorem, we have

$$P\left(\ddot{Y}_{tg1}\middle|\,\theta_{ct},\alpha_{cg},\alpha_{tg},Y_{cg},Y_{tg},W_g\right)$$
$$\propto\left(\prod_{i=1}^{m_{g1}}I(y_{i,g}=0)I(\ddot{y}_{i,g}\leq 0)\right)P(\ddot{Y}_{tg1},Y_{cg},Y_{tg},W_g|\theta_{ct},\alpha_{cg},\alpha_{tg}),\quad(17)$$

for $g=1,\cdots,G$.

(ii) $P(\phi|\{\ddot{Y}_{tg1}\},\{Y_{cg}\},\{Y_{tg}\},\{W_g\},\theta_{ct}\backslash\phi,\{\alpha_{cg}\},\{\alpha_{tg}\})$, where $\theta_{ct}\backslash\phi$ stands for θ_{ct} excluding ϕ.

By applying Bayes' theorem, we have

$$P(\phi|\{\ddot{Y}_{tg1}\},\{Y_{cg}\},\{Y_{tg}\},\{W_g\},\theta_{ct}\backslash\phi,\{\alpha_{cg}\},\{\alpha_{tg}\})$$
$$\propto\mathscr{T}\mathscr{N}_{2\bar{s}+\bar{q}+\bar{h}+2}(\phi;\phi_0,\Phi_0)\cdot\prod_{g=1}^{G}P(\ddot{Y}_{tg1},Y_{cg},Y_{tg},W_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\quad(18)$$

where $\mathscr{T}\mathscr{N}_{2\bar{s}+\bar{q}+\bar{h}+2}(\phi;\phi_0,\Phi_0)$ is the prior truncated density function of ϕ.

(iii) $P(\lambda_c|\{\ddot{Y}_{tg1}\},\{Y_{cg}\},\{Y_{tg}\},\{W_g\},\theta_{ct}\backslash\lambda_c,\{\alpha_{cg}\},\{\alpha_{tg}\})$.

By applying Bayes' theorem, we have

$$P(\lambda_c|\{\ddot{Y}_{tg1}\},\{Y_{cg}\},\{Y_{tg}\},\{W_g\},\theta_{ct}\backslash\lambda_c,\{\alpha_{cg}\},\{\alpha_{tg}\})$$
$$\propto\prod_{g=1}^{G}P(\ddot{Y}_{tg1},Y_{cg},Y_{tg},W_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\quad(19)$$

where $\lambda_c\in A=[-1/\tau_G,1/\tau_G]$.

(iv) $P(\lambda_t|\{\ddot{Y}_{tg1}\},\{Y_{cg}\},\{Y_{tg}\},\{W_g\},\theta_{ct}\backslash\lambda_t,\{\alpha_{cg}\},\{\alpha_{tg}\})$.

By applying Bayes' theorem, we have

$$P(\lambda_t|\{\ddot{Y}_{tg1}\},\{Y_{cg}\},\{Y_{tg}\},\{W_g\},\theta_{ct}\backslash\lambda_t,\{\alpha_{cg}\},\{\alpha_{tg}\})$$
$$\propto\prod_{g=1}^{G}P(\ddot{Y}_{tg1},Y_{cg},Y_{tg},W_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\quad(20)$$

where $\lambda_t\in A=[-1/\tau_G,1/\tau_G]$.

(v) $P(\beta_c|\{\ddot{Y}_{tg1}\}, \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}, \theta_{ct}\backslash\beta_c, \{\alpha_{cg}\}, \{\alpha_{tg}\})$.

By applying Bayes' theorem, we have

$$P(\beta_c|\{\ddot{Y}_{tg1}\}, \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}, \theta_{ct}\backslash\beta_c, \{\alpha_{cg}\}, \{\alpha_{tg}\}) \propto \mathscr{N}_{2k}(\beta_c; \beta_0, B_0) \cdot \prod_{g=1}^{G} P(\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{tg}, \alpha_{cg}), \tag{21}$$

where $\mathscr{N}_{2k}(\beta_c; \beta_0, B_0)$ is the prior normal density function of β_c.

(vi) $P(\beta_t|\{\ddot{Y}_{tg1}\}, \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}, \theta_{ct}\backslash\beta_t, \{\alpha_{cg}\}, \{\alpha_{tg}\})$.

By applying Bayes' theorem, we have

$$P(\beta_t|\{\ddot{Y}_{tg1}\}, \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}, \theta_{ct}\backslash\beta_t, \{\alpha_{cg}\}, \{\alpha_{tg}\}) \propto \mathscr{N}_{2k}(\beta_t; \beta_0, B_0) \cdot \prod_{g=1}^{G} P(\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{tg}, \alpha_{cg}), \tag{22}$$

where $\mathscr{N}_{2k}(\beta_t; \beta_0, B_0)$ is the prior normal density function of β_t.

(vii) $P(\sigma|\{\ddot{Y}_{tg1}\}, \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}, \theta_{ct}\backslash\sigma, \{\alpha_{cg}\}, \{\alpha_{tg}\})$.

By applying Bayes' theorem, we have

$$P(\sigma|\{\ddot{Y}_{tg1}\}, \{Y_{cg}\}, \{Y_{tg}\}, \{W_g\}, \theta_{ct}\backslash\sigma, \{\alpha_{cg}\}, \{\alpha_{tg}\}) \propto \mathscr{T}\mathscr{N}_3(\sigma; \sigma_0, \Sigma_0) \cdot \prod_{g=1}^{G} P(\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{tg}, \alpha_{cg}), \tag{23}$$

where $\mathscr{T}\mathscr{N}_3(\sigma; \sigma_0, \Sigma_0)$ is the prior truncated normal density function of σ.

(viii) $P(\alpha_{cg}|\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g, \theta_{ct}, \alpha_{tg})$, $g = 1, \cdots, G$.

By applying Bayes' theorem, we have

$$P(\alpha_{cg}|\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g, \theta_{ct}, \alpha_{tg}) \propto \mathscr{N}(\alpha_{cg}; \alpha_0, A_0) \cdot P(\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{tg}, \alpha_{cg}), \quad g = 1, \cdots, G, \tag{24}$$

where $\mathscr{N}(\alpha_{cg}; \alpha_0, A_0)$ is the prior normal density function of α_{cg}.

(ix) $P(\alpha_{tg}|\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g, \theta_{ct}, \alpha_{cg}),\ g = 1, \cdots, G.$

By applying Bayes' theorem, we have

$$P(\alpha_{tg}|\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g, \theta_{ct}, \alpha_{cg}) \propto \mathscr{N}(\alpha_{tg}; \alpha_0, A_0) \cdot P(\ddot{Y}_{tg1}, Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{tg}, \alpha_{cg}), \quad g = 1, \cdots, G, \tag{25}$$

where $\mathscr{N}(\alpha_{tg}; \alpha_0, A_0)$ is the prior normal density function of α_{tg}.

All of the conditional posterior distributions from (i) to (ix) are not available in a closed form. However, we may use the Metropolis-Hastings (M-H) algorithm to draw from these distributions. Tierney (1994) and Chib and Greenberg (1996) have shown that the combination of Markov chains (Metropolis-within-Gibbs) is still a Markov chain with the invariant distribution being the correct objective distribution. The procedure of the MCMC sampling starts with arbitrary initial values for $\{\alpha_{cg}^{(0)}\}$, $\{\alpha_{tg}^{(0)}\}$, and $\theta_{ct}^{(0)}$, and then the sampling proceeds sequentially from the above set of conditional posterior distributions.[17] In the following section, we will discuss a relative new version of the M-H algorithm which could be used when the likelihood function of the model contains an intractable normalizing term.

3.3 The double M-H algorithm

From Section 3.1, the likelihood function of $y = (\{Y_{cg}\}, \{Y_{tg}\}, \{W_g\})$, given the parameter θ, takes the form $P(y|\theta) = f(y; \theta)/D(\theta)$, where $D(\theta)$ is an intractable normalizing term.[18] The standard M-H algorithm to simulate random draws of θ runs as follows: given the old draw, θ_{old}, one proposes a new draw, θ_{new}, from a proposal distribution, $q(\cdot|\theta_{old})$, and then updates the old draw to the new draw with

[17] A further detail about implementing the MCMC sampling based on steps (i) to (ix) is available online.

[18] Here $P(y|\theta) = \prod_{g=1}^{G} P(Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg})$ with $P(Y_{cg}, Y_{tg}, W_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg})$ from Eq (15) and θ would refer to the vector $(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\})$.

an acceptance probability, $\alpha(\theta_{new}|\theta_{old})$. Denoting $\pi(\theta)$ as the prior probability of θ, the acceptance probability is computed as

$$\begin{aligned}\alpha(\theta_{new}|\theta_{old}) &= \min\left\{1, \frac{P(\theta_{new}|y)q(\theta_{old}|\theta_{new})}{P(\theta_{old}|y)q(\theta_{new}|\theta_{old})}\right\} \\ &= \min\left\{1, \frac{\pi(\theta_{new})f(y;\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})f(y;\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{D(\theta_{old})}{D(\theta_{new})}\right\}. \end{aligned} \tag{26}$$

One can see that in Eq. (26), the normalizing terms, $D(\theta_{old})$ and $D(\theta_{new})$, are left in both the numerator and denominator and will not cancel out, so the evaluation of the acceptance-rejection criterion with α in Eq. (26) would be intractable. Murray *et al.* (2006) consider to include auxiliary variables, $\widetilde{y} = (\{\widetilde{Y}_{cg}\}, \{\widetilde{Y}_{tg}\}, \{\widetilde{W}_g\})$, into the acceptance probability, i.e., the acceptance probability conditional on $\widetilde{y}$ can be written as

$$\begin{aligned}\alpha(\theta_{new}|\theta_{old}, \widetilde{y}) &= \min\left\{1, \frac{\pi(\theta_{new})P(y|\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})P(y|\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{P(\widetilde{y}|\theta_{old})}{P(\widetilde{y}|\theta_{new})}\right\} \\ &= \min\left\{1, \frac{\pi(\theta_{new})f(y;\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})f(y;\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{f(\widetilde{y};\theta_{old})}{f(\widetilde{y};\theta_{new})}\right\}, \end{aligned} \tag{27}$$

where $\widetilde{y}$ are simulated from the likelihood function, $P(\widetilde{y}|\theta_{new}) = f(\widetilde{y};\theta_{new})/D(\theta_{new})$ with the exact sampling (Propp and Wilson, 1996). In the conditional acceptance probability of Eq (27), all normalizing terms cancel out and the other terms left are computable. This algorithm bypasses evaluating the normalizing terms. However, implementing the exact sampling is time consuming. In order to save time on the computation, Liang (2010) proposes a 'double M-H algorithm' which utilizes the reversibility condition and shows that when $\widetilde{y}$ are simulated by the M-H algorithm starting from y with m iterations, the conditional acceptance probability in Eq. (27) can be obtained regardless of the value of m. This gives the double M-H algorithm an advantage, as a small value of m can be used, removing the need of the exact sampling. Due to this computational efficiency, we adopt the double M-H algorithm in this study.

One thing worth mentioning is that in this paper we have provided a technical modification on the double M-H algorithm to make it simplify the simulation and better fit into our application. Using the double M-H algorithm to update θ from $P(\theta|y)$ requires simulating auxiliary variables, $\widetilde{y}$. However, the auxiliary activity variables, $\{\widetilde{Y}_{cg}\}$ and $\{\widetilde{Y}_{dg}\}$, in $\widetilde{y}$ are redundant as they can be fully replaced by a function of auxiliary networks, $\widetilde{w} = \{\widetilde{W}_g\}$. Therefore, we modify the conditional acceptance probability in Eq. (27) to

$$\begin{aligned}\alpha(\theta_{new}|\theta_{old}, \widetilde{w}) &= \min\left\{1, \frac{\pi(\theta_{new})P(y|\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})P(y|\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{P(\widetilde{w}|\theta_{old})}{P(\widetilde{w}|\theta_{new})}\right\} \\ &= \min\left\{1, \frac{\pi(\theta_{new})f(y;\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})f(y;\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{f(\widetilde{w};\theta_{old})}{f(\widetilde{w};\theta_{new})}\right\}. \qquad (28)\end{aligned}$$

To evaluate $\alpha(\theta_{new}|\theta_{old}, \widetilde{w})$ in Eq. (28), we only simulate the auxiliary networks, $\widetilde{w}$, from the probability density function, $P(\widetilde{w}|\theta_{new}) = f(\widetilde{w};\theta_{new})/D(\theta_{new})$, which shares the same normalizing term, $D(\theta_{new})$, with $P(\widetilde{y}|\theta_{new})$.[19]

To show that one can successfully draw from a target density $P(\theta|y)$ (assuming θ is continuous for simplicity) by using the double M-H algorithm with the acceptance probability of Eq. (28), we need to show that the Markov chain based on the transition density, $p(\theta_{new}|\theta_{old}) = \alpha(\theta_{new}|\theta_{old})q(\theta_{new}|\theta_{old})$, is reversible, i.e.,

$$P(\theta_{old}|y)\alpha(\theta_{new}|\theta_{old})q(\theta_{new}|\theta_{old}) = P(\theta_{new}|y)\alpha(\theta_{old}|\theta_{new})q(\theta_{old}|\theta_{new}), \qquad (29)$$

and therefore, $P(\theta|y)$ is an invariant distribution. To check Eq. (29), we need the

[19]Here, $P(\widetilde{w}|\theta_{new})$ denotes the joint probability density function of $(W_1, \cdots, W_G)$. In practice, each auxiliary network, $\widetilde{W}_g$, is simulated by the M-H algorithm from W_g based on $P(W_g|\theta_{new})$ in Eq. (5). The following step is implemented iteratively: one randomly picks up an entry of W_g, $w_{ij,g}$, $i \neq j$, and proposes $\widetilde{w}_{ij,g} = 1 - w_{ij,g}$, with the acceptance probability

$$\alpha(\widetilde{w}_{ij,g}|w_{ij,g}) = \min\left\{\frac{\exp\left(\mathrm{TU}(\tilde{w}_{ij,g}, W_{-ij,g}, \theta_{new})\right)}{\exp\left(\mathrm{TU}(w_{ij,g}, W_{-ij,g}, \theta_{new})\right)}, 1\right\},$$

updating $w_{ij,g}$ to $\widetilde{w}_{ij,g}$. Note that the denominators of $P(\tilde{w}_{ij,g}, W_{-ij,g}|\theta_{new})$ and $P(w_{ij,g}, W_{-ij,g}|\theta_{new})$ are canceled out because the two probabilities are evaluated at the same θ_{new}.

unconditional acceptance probability, that is

$$
\begin{aligned}
\alpha(\theta_{new}|\theta_{old}) &= \int \alpha(\theta_{new}|\theta_{old}, \widetilde{w}) P(\widetilde{w}|\theta_{new}) d\widetilde{w} \\
&= \int \min\left\{ \frac{\pi(\theta_{new})P(y|\theta_{new})q(\theta_{old}|\theta_{new})}{\pi(\theta_{old})P(y|\theta_{old})q(\theta_{new}|\theta_{old})} \cdot \frac{P(\widetilde{w}|\theta_{old})}{P(\widetilde{w}|\theta_{new})}, 1 \right\} P(\widetilde{w}|\theta_{new}) d\widetilde{w} \\
&= \int \min\left\{ \frac{P(\theta_{new}|y)q(\theta_{old}|\theta_{new})}{P(\theta_{old}|y)q(\theta_{new}|\theta_{old})} P(\widetilde{w}|\theta_{old}), P(\widetilde{w}|\theta_{new}) \right\} d\widetilde{w}. \qquad (30)
\end{aligned}
$$

With Eq. (30), the left hand side of Eq. (29) equals to

$$
\begin{aligned}
& P(\theta_{old}|y)\alpha(\theta_{new}|\theta_{old})q(\theta_{new}|\theta_{old}) \\
&= \int \min\left\{ P(\theta_{new}|y)q(\theta_{old}|\theta_{new})P(\widetilde{w}|\theta_{old}), P(\theta_{old}|y)q(\theta_{new}|\theta_{old})P(\widetilde{w}|\theta_{new}) \right\} d\widetilde{w}.
\end{aligned}
$$

The right hand side of Eq. (29) equals to

$$
\begin{aligned}
& P(\theta_{new}|y)\alpha(\theta_{old}|\theta_{new})q(\theta_{old}|\theta_{new}) \\
&= \int \min\left\{ P(\theta_{old}|y)q(\theta_{new}|\theta_{old})P(\widetilde{w}|\theta_{new}), P(\theta_{new}|y)q(\theta_{old}|\theta_{new})P(\widetilde{w}|\theta_{old}) \right\} d\widetilde{w}.
\end{aligned}
$$

Since these two are equal, the reversibility condition in Eq. (29) is satisfied.

4 Empirical Study

We apply our model to study American high school students' friendship networks in the Add Health data, which is a national survey based on grades 7 through 12 in 132 schools.[20] Four waves of surveys were conducted between 1994 and 2008. In the

[20] This is a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). No direct support was received from grant P01-HD31921 for this analysis.

wave I in-school survey, a total of 90,182 students were interviewed. Each respondent answered questions about their demographic backgrounds, academic performances, and health-related behaviors. Most uniquely, students were asked to nominate up to five male and five female friends. This provides information about their friendship networks. In the following waves of in-home surveys, more information about students' families and neighborhoods in which they live are available for a subset of the total sample. To accommodate most of the students' nominated friends into our framework, the sample used in this study is constructed from the wave I in-school survey. We consider two activity variables which may be relevant for friendship formation. One is a student's academic performance (measured by GPA), which is represented by a continuous variable.[21] The other is how frequently a student smokes in a week, which is represented by a Tobit-type variable.

In the context of social interactions, students' academic performance and smoking behavior are extensively studied as they have important long-term consequences on students' future lives and health.[22] To obtain interaction effects on these two objects, researchers face difficulty identifying correlated effects from group-level unobservables and endogenous selection into groups (Moffitt, 2001), and separating the endogenous interaction effect from contextual effects in a linear model (the reflection problem by Manski, 1993). With various approaches to solving these difficulties, researchers generally provide evidence for the existence of peer effects. Hsieh and Lee (2011) further consider the problem of endogenous friendship selection on peer effects by modeling unobservables in both the network interaction and network for-

[21]GPA is regarded as a proxy for studying activities.

[22]One can find studies of peer effects on students' academic performance in Hoxby (2000), Sacerdote (2001), Hanushek *et al.* (2003), and Zimmerman (2003) with the use of the linear-in-means model; and Calvó-Armengol *et al.* (2009), Lin (2010), Boucher *et al.* (2010), and Liu *et al.* (2011) with the use of the network interactions model. For studies of peer effects on students' smoking behaviors, one can see evidence of peer effects on Gaviria and Raphael (2001), Powell *et al.* (2005), Lundborg (2006), Clark and Loheac (2007), and Fletcher (2010).

mation processes. They show that the endogenous effect obtained from the SAR model without controlling the endogeneity of the (spatial weight) network matrix can be upward biased. In the present study, we confirm Hsieh and Lee's (2001) finding that the endogenous effect would be smaller after controlling endogenous formation of friendship networks. Moreover, we will show that the benefit of interactions from academic learning is an important factor for students to form friendships.

4.1 The empirical specification of network structure effects in the link-associated utility

For the empirical application, we consider the following specification of the network structure effects in the link-associated utility of Eq. (6),[23]

$$
\begin{aligned}
&\varpi_{i,g}(w_{i\cdot,g}, W_{-i\cdot,g})\eta \\
&= \underbrace{\eta_1 \sum_{j=1}^{m_g} w_{ij,g}w_{ji,g}}_{\text{Reciprocality Effect}} + \underbrace{\eta_2 \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k\neq j}^{m_g} w_{ik,g}\right) + \eta_3 \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k\neq j}^{m_g} w_{ik,g}\right)^2}_{\text{Sender's Expansiveness Effect}} + \underbrace{\eta_4 \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k\neq i}^{m_g} w_{kj,g}\right)}_{\text{Receiver's Popularity Effect}} \\
&\quad + \underbrace{\eta_{51} \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k}^{m_g} w_{ik,g}w_{kj,g}\right) + \eta_{52} \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k}^{m_g} w_{ki,g}w_{kj,g}\right) + \eta_{53} \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k}^{m_g} w_{ik,g}w_{jk,g}\right)}_{\text{Transitive Triads Effect}} \\
&\quad + \underbrace{\eta_6 \sum_{j=1}^{m_g} w_{ij,g}\left(\sum_{k}^{m_g} w_{jk,g}w_{ki,g}\right)}_{\text{Three Cycles Effect}}. \qquad (31)
\end{aligned}
$$

In Eq. (31), the reciprocality effect reflects the utility from reciprocal friendships. Even though each link decision (naming friend) is made by one individual without mutual consent from another, the possibility of reciprocality may be a factor in an individual's link decision. The sender's expansiveness effect in Eq. (31) reflects

[23] The effects we consider here are mostly mentioned in Snijders *et al.* (2010) except for the squared term of the sender's outdegree to capture a nonlinear expansiveness effect. In practice, any other relevant network structure effects can be incorporated into our utility specification.

the utility from being an outgoing person who actively nominates friends. The statistics involved are the sender's outdegree and the outdegree square. We expect the coefficient, η_3, would be negative to reflect the reality that individuals might not make too many friends due to limited resources, e.g., limited time, energy, and money. The receiver's indegree is used to measure the receiver's popularity effect in Eq. (31), which reflects the utility from making friends with someone who is popular. The transitive triads effect and the three cycles effect reflect the utility from engaging in a transitive relationship, i.e., friends of my friends are my friends. However, they are distinguished by directions of links. From Kovářík and van der Leij (2012), transitive triads effects may be linked to an individual's sense of risk aversion. The three-cycles effect can be interpreted as an opposite hierarchy effect (Snjiders *et al.*, 2010). If the coefficient η_6 is negative, it implies a local hierarchy among linked individuals.

Given $\varpi_{i,g}(w_{i.,g}, W_{-i.,g})\eta$ in Eq. (31), the term $\sum_{i=1}^{m_g} \varpi_{i,g}(w_{i.,g}, W_{-i.,g})\eta$ in the transferable utility of Eq.(8) can be written as

$$\begin{aligned}
&\sum_{i=1}^{m_g} \varpi_{i,g}(w_{i.,g}, W_{-i.,g})\eta \\
&= \eta_1 \mathrm{tr}(W_g^2) + \eta_2(l_g' W_g' W_g l_g - l_g' W_g l_g) \\
&\quad + \eta_3(l_g' W_g' \mathrm{Diag}(W_g l_g) W_g l_g - 2 l_g' W_g' W_g l_g + l_g' W_g l_g) \\
&\quad + \eta_4(l_g' W_g W_g' l_g - l_g' W_g l_g) + (\eta_{51} + \eta_{52} + \eta_{53}) \mathrm{tr}(W_g^2 W_g') + \eta_6 \mathrm{tr}(W_g^3), \qquad (32)
\end{aligned}$$

where Diag(A) is a $n \times n$ diagonal matrix with its diagonal elements formed by the entries of a $n \times 1$ vector of A. One can see that parameters η_{51}, η_{52} and η_{53} are not separately identified from Eq. (32). Hence, without loss of generality, we will use η_5 for $\eta_{51} + \eta_{52} + \eta_{53}$ hereafter.

4.2 Data summary

To ease the computation burden, we only work with small networks in this study. The following steps are used to construct the sample. First, we group students by their school and grade level and consider friendships only inside the same group.[24] Second, we focus on senior high school students from 9^{th} to 12^{th} grades. Third, we restrict our network sample to those groups with sizes between 10 and 50 (10 and 60 for the smoking case). After removing missing observations on activity variables in each group, a total of 1,177 (1,476 for the smoking case) respondents from 47 networks (44 networks for the smoking case) are left for analysis.[25] These networks have the average size equal to 25.043 (33.546 for the smoking case), average density equal to 0.142 (0.108 for the smoking case), average outdegree equal to 2.564 (2.866 for the smoking case), and average clustering coefficient equal to 0.327 (0.332 for the smoking case).[26] In the network formation model, we capture individual-specific effects by a dummy variable of whether a student is older than the group average or not. Three other dummy variables – whether a pair of students has the same age, same sex, or same race – are used to capture dyad-specific effects.

For the network interaction model, the continuous variable, GPA, is calculated by the average of a respondent's reported grades from several subjects, including language, social science, mathematics, and science, of which each has a value between

[24] In the Add Health data, about 80% of friendship nominations happen within the same grade level. Hence, about 20% of links are missed due to the design of network boundary.

[25] The number of missing observations is equal to 113 (9.6%) for the GPA sample and 34 (2%) for the smoking sample.

[26] The outdegree for individual i is calculated by $\sum_j w_{ij,g}$. The average degree is $\sum_i \sum_{j \neq i} w_{ij,g}/m_g$. The network density is obtained by further dividing the average degree with $(m_g - 1)$. The clustering coefficient is calculated as the total fraction of transitive triples in the network, i.e.,

$$C(W_g) = \frac{\sum_{i;j\neq i;k\neq i,j} w_{ij,g}w_{jk,g}w_{ik,g}}{\sum_{i;j\neq i;k\neq i,j} w_{ij,g}w_{jk,g}}.$$

1 and 4. The Tobit-type variable, smoking, is obtained from students' response to the survey question, "*During the past twelve months, how often did you smoke cigarettes*?", which has a value between 0 and 7. We follow Lin (2010), Lee *et al.* (2010, 2013), and Hsieh and Lee (2011) to choose the independent variables which are used in the network interaction model. A complete list of variables is provided in Table 1. In Figures 1 and 2, we plot two typical networks from our sample – one is from the GPA sample and the other is from the smoking sample. From these two figures, one can observe that students who have higher GPAs tend to receive more friendship nominations than those who have lower GPAs. This observation does not seem to be evident for smoking behaviors, but one can find that smokers are friends with each other. Our estimation results shown in the following section provide evidence for the incentive stemming from interaction benefits on academic learning, but not from the pleasure of smoking together, on friendship decisions. Moreover, our results show that interaction effects on influencing GPA or smoking frequency are significant.

To obtain estimates from the Bayesian estimation in this empirical study, the values of hyperparameters in the prior distributions are set as follows: $\phi_0 = 0$; $\Phi_0 = 10I_{2\bar{s}+\bar{q}+8}$; $\beta_0 = 0$; $B_0 = 10I_{2k}$; $\sigma_0 = 0$; $\Sigma_0 = 10I_3$; $\alpha_0 = 0$; $A_0 = 400$. These specified values of hyperparameters are designed to allow relative flat prior densities over the ranges of the parameter spaces. The reported estimation results are based on the MCMC sampling draws which pass the convergence test provided by Geweke (1992).

4.3 Estimation results

4.3.1 The case of GPA

We first estimate the model with a single continuous variable, GPA, and report the results in Table 2. The values shown for each parameter are the mean and

Table 1: Summary Statistics

			GPA		Smoking	
variable	min	max	mean	s.d.	mean	s.d.
GPA	1	4	2.910	0.734	-	-
Smoking	0(57.86%)	7	-	-	1.257	2.511
Age	10	19	16.004	1.285	15.997	1.269
Male	0	1	0.493	0.500	0.482	0.499
Female	0	1	0.507	0.500	0.517	0.499
White	0	1	0.611	0.487	0.629	0.483
Black	0	1	0.246	0.430	0.230	0.421
Asian	0	1	0.016	0.125	0.016	0.123
Hispanic	0	1	0.068	0.251	0.067	0.250
Other race	0	1	0.059	0.236	0.058	0.233
Both parents	0	1	0.725	0.447	0.733	0.442
Less HS	0	1	0.114	0.318	0.109	0.312
HS	0	1	0.340	0.473	0.341	0.474
More HS	0	1	0.398	0.490	0.402	0.490
Edu missing	0	1	0.068	0.252	0.067	0.250
Professional	0	1	0.248	0.432	0.249	0.432
Staying home	0	1	0.220	0.414	0.228	0.419
Other Jobs	0	1	0.366	0.481	0.356	0.479
Job missing	0	1	0.076	0.265	0.077	0.266
Welfare	0	1	0.011	0.103	0.010	0.100
Num. of students at home	0	6	0.580	0.818	0.568	0.793
Network size			25.043	13.146	33.546	16.551
Network density			0.142	0.100	0.108	0.076
Outdegree			2.564	2.294	2.866	2.406
Indegree			2.564	2.418	2.866	2.596
Clustering Coef.			0.327	0.120	0.332	0.086
Sample size			1,177		1,476	
Num. of networks			47		44	

Both parents means living with both parents. Less HS means mother's education is less than high school.

Edu missing means mother's education level is missing.

Professional means mother's job is either scientist, teacher, executive, director and the like.

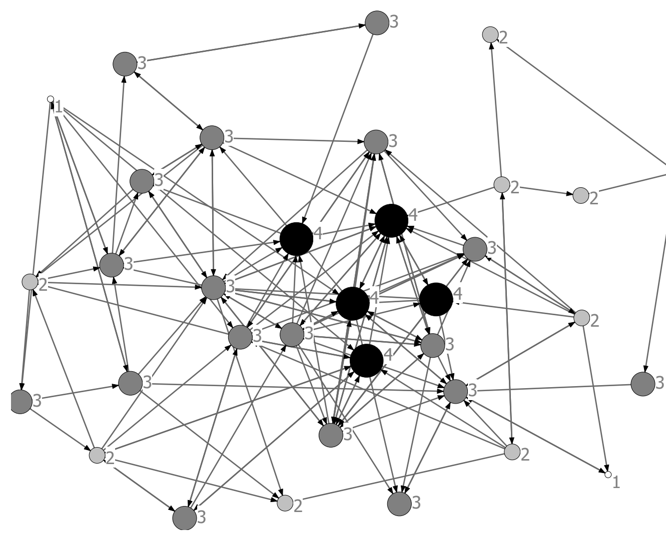
Other jobs means mother's occupation is not among "professional" or "staying home".

Welfare means mother participates in social welfare programs.

Number of students at home means how many other students of grade 7 to 12 living in the same household with you.

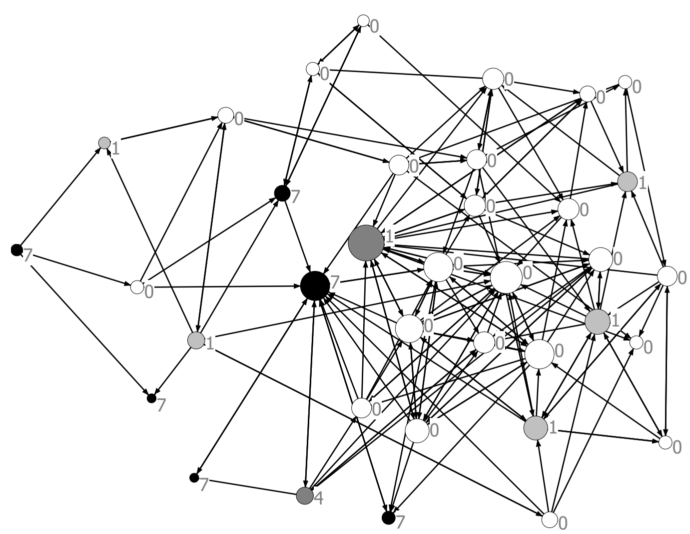
The variables in italics are omitted categories in the estimation.

Figure 1: A friendship network from the GPA sample



Note: The number (and color) for each node indicates the value of GPA. Nodes with a larger size means they have higher indegrees.

Figure 2: A friendship network from the Smoking sample



Note: The number (and color) for each node indicates the frequency of smoking. Nodes with a larger size means they have higher indegrees.

the standard deviation (in parentheses) from posterior draws. In the first column, we consider the full model with network formation and network interaction. In the second column, we consider the network interaction model alone with networks assumed exogenous. For examining the possible consequence of dropping 9.6% of the sample due to the missing observation on the dependent variable, GPA, the result in the third column is obtained from the full model with the Bayesian data augmentation approach to recover missing observations.[27]

From the part of network formation in the first column, we observe that being older than the group average or not does not have a significant effect on sending or receiving friendship nominations. However, the three dyad-specific effects are all positive and significant, where the effect of the same race is strongest, followed by the effects of the same sex and the same age. Among network structure effects, the positive and strong reciprocality effect is consistent with findings in the literature (Snijders *et al.*, 2010; Mele, 2010), which reflects that mutual friendship nominations are pervasive among students. In our sample, 49.8% of friendship links are reciprocal. The sender's expansiveness effect is concave, as the coefficient of the first order term is positive and the coefficient of the second order term is negative. This result confirms our conjecture that limited resources, e.g., limited time, energy, or money, might constrain students from making too many friends. The receiver's popularity effect is negative, which suggests that students between 9^{th} and 12^{th} grades in our sample are less willing to make friends with someone who is popular.

[27] At the time we drop observations with missing dependent variables, we also drop the potential links connected to these observations. Since we study network formation and network interaction, if many links were dropped, the resulting estimates of parameters might be biased. These missing observations can be treated as unobserved random variables and updated with other unknown parameters by the MCMC sampling. The advantage of doing this is that we could retrieve information provided by these missing observations and obtain consistent and efficient estimates. The problem of a missing independent variable can be much more involved and therefore we do not discuss this case in this paper.

The positive and strong transitive triads effect shows that students value transitive relationships. When the positive triads effect is accompanied by the negative three cycles effect, as discussed in Snijders *et al.* (2010), a certain degree of local hierarchy among students is revealed. The incentive effect from GPA is found to be large and significant. Therefore, for high school students in our sample, the potential benefit of learning from others in school work is a factor which determines their friendship decisions.

From the part of network interaction in the full model, the estimated endogenous effect is equal to 0.021 and significant. This value implies that, on average, one standard deviation increase in total friends' GPAs will increase a student's GPA by 0.154 units. The social multiplier effects across students and groups implied by this estimate have the maximum and average equal to 1.248 and 1.060, respectively.[28] From the estimated own and contextual effects, we observe that students who are older, male, or whose mothers have received less education than high school tend to have lower GPAs.[29] Also, students' GPAs could be negatively affected by having friends who are either older, male, Black, or Asian.

When estimating the network interaction model alone by treating the weight matrix as exogenously given, the result in the second column shows that the estimated endogenous effect and its standard deviation are nearly double of those obtained from the full model. Meanwhile, the estimated own and contextual effects are also different from those of the full model and have larger standard deviations. These differences between results in the first and second columns show the problem of friendship selection biases in the network interaction model.

In the third column, after augmenting missing observations on GPA during the estimation procedure, we do not observe significant changes on the estimates of pa-

[28]The vector of social multiplier effects can be calculated by $(I_{m_g} - \lambda W_g)^{-1} l_{m_g}$.

[29]We do not interpret estimates which are insignificant, i.e., the posterior standard deviation is close to or larger than the posterior mean.

rameters in the network formation model and the endogenous effect in the network interaction model, when comparing with the first column. Although there are few changes on the estimates of own and contextual effects, these are insignificant estimates. The results suggest that dropping 9.6% of missing observations on GPA does not cause any significant bias on the estimated interaction effects. However, the advantage of using the data augmentation approach to recover missing observations can be seen from smaller standard deviations of parameters in both the network formation and network interaction models.

4.3.2 The case of smoking

Next, we consider the model with a single incentive effect from smoking. Both the full model and the network interaction model alone with networks assumed exogenous are estimated. Their results are reported in the first and second columns in Table 3. As there are only 2% of missing observations on the variable of smoking, we do not consider to use the data augmentation approach to recover them. From the part of network formation in the first column, we still find that being older than the group average does not have a significant effect on sending or receiving friendship nominations. The estimates of dyad-specific effects show that being same sex or same race are important for friendship decisions, while being same age is not. Network structure effects are generally similar to those in the case of GPA. An important finding is that the incentive effect from smoking is small and insignificant. Hence, we can say that students in our sample do not consider the pleasure of smoking together as a factor for their friendship decisions. From the part of network interaction in the full model, the estimated endogenous interaction effect shown in the first column of Table 3 is equal to 0.080 and significant, which implies that, on average, one standard deviation increase in total friends' smoking frequencies will increase a student's smoking frequency by 0.425 units. The social multipliers across students and groups implied by this estimate have the maximum and average equal to 2.708 and 1.345, respectively. The estimated own and contextual effects show that students who are Black, or who live with both parents tend to smoke less than their counterparts. Also, a student may smoke less if he or she has friends who are Asian. In the second column, the estimated endogenous effect from the network interaction model alone is equal to 0.088, which is not significantly different from that obtained from the full model. Although few estimated own and contextual effects are different from those in the full model, they are insignificant effects. These results suggest that outcomes of network interactions for smoking are not subject to friendship selection bias.

Table 2: Estimation result based on GPA

	[Full]				[Network Interaction Only]				[Full with Missing]			
Network Formation												
Higher sender age (γ_1)	-0.036	(0.045)			-				-0.029	(0.045)		
Higher receiver age (γ_2)	-0.044	(0.051)			-				0.005	(0.047)		
Constant (γ_{31})	-4.305	(0.093)			-				-4.456	(0.089)		
Same age (γ_{32})	0.088	(0.041)			-				0.089	(0.041)		
Same sex (γ_{33})	0.365	(0.041)			-				0.369	(0.036)		
Same race (γ_{34})	0.428	(0.057)			-				0.451	(0.051)		
Reciprocality (η_1)	1.321	(0.043)			-				1.308	(0.038)		
Expansiveness (η_2)	0.240	(0.022)			-				0.253	(0.021)		
Expansiveness (η_3)	-0.029	(0.002)			-				-0.029	(0.002)		
Popularity (η_4)	-0.034	(0.006)			-				-0.028	(0.006)		
Trans. triads (η_5)	0.572	(0.022)			-				0.576	(0.019)		
Three cycles (η_6)	-0.247	(0.019)			-				-0.257	(0.016)		
Economic incentive (δ_c)	1.118	(0.225)			-				1.056	(0.218)		
Network Interaction												
Endogenous	0.021	(0.007)			0.040	(0.012)			0.023	(0.006)		
σ^2_ϵ	0.448	(0.019)			0.448	(0.019)			0.559	(0.023)		
	Own		Contextual		Own		Contextual		Own		Contextual	
Age	-0.254	(0.020)	-0.003	(0.002)	-0.209	(0.029)	-0.005	(0.003)	-0.245	(0.018)	-0.003	(0.001)
Male	-0.060	(0.029)	-0.019	(0.009)	-0.114	(0.042)	-0.001	(0.028)	-0.149	(0.023)	-0.018	(0.010)
Black	0.054	(0.027)	-0.056	(0.009)	0.055	(0.082)	-0.041	(0.024)	0.043	(0.044)	-0.048	(0.012)

Continued on Next Page

Table – Continued

Asian	0.275	(0.045)	-0.186	(0.058)	0.236	(0.204)	-0.207	(0.206)	0.318	(0.035)	-0.057	(0.041)
Hispanic	0.361	(0.034)	0.013	(0.019)	0.266	(0.105)	0.025	(0.060)	0.229	(0.038)	0.034	(0.018)
Other race	-0.045	(0.038)	-0.012	(0.022)	-0.040	(0.097)	-0.115	(0.058)	-0.040	(0.030)	0.005	(0.019)
Both Parents	0.122	(0.036)	0.030	(0.011)	0.064	(0.048)	0.029	(0.033)	0.044	(0.028)	0.043	(0.012)
Less HS	-0.090	(0.042)	-0.020	(0.018)	-0.096	(0.068)	-0.083	(0.048)	-0.021	(0.035)	-0.027	(0.014)
More HS	0.096	(0.044)	0.006	(0.010)	0.140	(0.050)	-0.002	(0.032)	0.172	(0.019)	0.000	(0.010)
Edu missing	-0.055	(0.029)	-0.005	(0.021)	-0.103	(0.085)	0.031	(0.065)	-0.147	(0.037)	-0.040	(0.013)
Welfare	0.047	(0.033)	-0.058	(0.039)	0.161	(0.202)	-0.141	(0.164)	0.112	(0.037)	-0.085	(0.039)
Job missing	-0.142	(0.032)	-0.031	(0.020)	-0.073	(0.078)	-0.011	(0.055)	-0.066	(0.052)	-0.023	(0.013)
Professional	-0.040	(0.040)	-0.006	(0.011)	-0.031	(0.061)	0.007	(0.040)	0.041	(0.036)	0.006	(0.011)
Other Jobs	-0.015	(0.037)	-0.014	(0.011)	-0.004	(0.053)	0.002	(0.034)	0.054	(0.026)	-0.007	(0.008)
Num. of students at home	0.088	(0.029)	0.029	(0.009)	-0.000	(0.026)	-0.009	(0.017)	-0.033	(0.018)	-0.012	(0.006)

The MCMC runs for 100,000 iterations and the first 20,000 runs are dropped for the burn-in.

Table 3: Estimation result based on Smoking

	[Full]				[Network Interaction Only]			
Network Formation								
Higher sender age (γ_1)	-0.058	(0.040)			-			
Higher receiver age (γ_2)	-0.067	(0.040)			-			
Constant (γ_{31})	-4.344	(0.081)			-			
Same age (γ_{32})	0.057	(0.033)			-			
Same sex (γ_{33})	0.381	(0.032)			-			
Same race (γ_{34})	0.338	(0.039)			-			
Reciprocality (η_1)	1.357	(0.036)			-			
Expansiveness (η_2)	0.222	(0.016)			-			
Expansiveness (η_3)	-0.026	(0.001)			-			
Popularity (η_4)	-0.030	(0.004)			-			
Trans. triads (η_5)	0.624	(0.017)			-			
Three cycles (η_6)	-0.269	(0.015)			-			
Economic incentive (δ_t)	0.023	(0.017)			-			
Network Interaction								
Endogenous	0.080	(0.011)			0.088	(0.010)		
σ_ϵ^2	18.761	(1.253)			18.580	(1.456)		
	Own		Contextual		Own		Contextual	
Age	0.122	(0.091)	0.002	(0.015)	0.164	(0.071)	-0.003	(0.015)
Male	0.422	(0.265)	-0.279	(0.187)	0.343	(0.254)	-0.217	(0.167)
Black	-2.580	(0.559)	-0.008	(0.144)	-2.080	(0.639)	-0.037	(0.141)

Continued on Next Page

Table – Continued

Asian	-0.602	(0.698)	-1.101	(0.428)	-0.134	(0.624)	-1.394	(0.670)
Hispanic	-0.179	(0.464)	0.361	(0.293)	-0.000	(0.384)	0.129	(0.314)
Other race	0.368	(0.429)	-0.100	(0.343)	0.175	(0.410)	-0.099	(0.317)
Both Parents	-0.823	(0.263)	-0.331	(0.190)	-0.523	(0.261)	-0.353	(0.217)
Less HS	0.223	(0.349)	0.092	(0.250)	0.134	(0.318)	0.259	(0.207)
More HS	-0.334	(0.284)	0.137	(0.184)	-0.248	(0.259)	0.201	(0.173)
Edu missing	0.106	(0.480)	0.530	(0.276)	0.124	(0.425)	0.545	(0.354)
Welfare	0.823	(0.713)	-0.351	(0.641)	1.366	(0.521)	0.155	(0.689)
Job missing	0.387	(0.369)	-0.181	(0.288)	0.193	(0.391)	-0.003	(0.295)
Professional	0.115	(0.314)	-0.263	(0.221)	-0.087	(0.360)	-0.257	(0.209)
Other Jobs	-0.109	(0.230)	-0.094	(0.185)	-0.324	(0.313)	-0.037	(0.180)
Num. of students at home	-0.156	(0.171)	-0.059	(0.096)	-0.116	(0.145)	-0.071	(0.096)

The MCMC runs for 100,000 iterations and the first 20,000 runs are dropped for the burn-in.

4.3.3 The case of combining both GPA and smoking

Lastly, we estimate the full model with incentive effects from both GPA and smoking.[30] From the results reported in Table 4, the estimates of incentive effects on GPA and smoking are very close to those in Tables 2 and 3. Therefore, the joint modeling of both incentive effects does not affect the estimate of each single effect, which confirms the separability of incentive effects. By comparing the results between the full model and the network interaction model alone with networks assumed exogenous, we observe a significant friendship selection bias on the estimated endogenous effect for GPA, which changes from 0.049 when estimating the network interaction model alone to 0.025 when estimating the full model. For smoking, due to a small and insignificant incentive effect, we do not find evidence of friendship selection bias in the estimate of its interaction effects. Furthermore, the covariance of disturbances in the network interaction models between GPA and smoking is found to be -0.653 and significant.

[30]The sample used to estimate this model is based on the original GPA sample where we remove missing observations on smoking.

Table 4: Estimation result based on both GPA and Smoking

	[Full]				[Network Interaction Only]			
Network Formation								
Higher sender age (γ_1)	-0.016 (0.018)				-			
Higher receiver age (γ_2)	0.061 (0.024)				-			
Constant (γ_{31})	-4.616 (0.066)				-			
Same age (γ_{32})	0.120 (0.026)				-			
Same sex (γ_{33})	0.342 (0.026)				-			
Same race (γ_{34})	0.620 (0.020)				-			
Reciprocality (η_1)	1.376 (0.034)				-			
Expansiveness (η_2)	0.246 (0.014)				-			
Expansiveness (η_3)	-0.028 (0.002)				-			
Popularity (η_4)	-0.038 (0.006)				-			
Trans. triads (η_5)	0.560 (0.011)				-			
Three cycles (η_6)	-0.235 (0.010)				-			
Incentive from GPA (δ_c)	1.013 (0.022)				-			
Incentive from Smoking (δ_t)	0.026 (0.014)				-			
Network Interaction	[GPA]		[Smoking]		[GPA]		[Smoking]	
Endogenous	0.025 (0.005)		0.079 (0.013)		0.049 (0.015)		0.085 (0.013)	
σ_ϵ^2	0.499 (0.022)		20.115 (1.446)		0.479 (0.022)		19.368 (1.778)	
$\sigma_{\epsilon_{tc}}$	-0.653 (0.117)							
	Own	Contextual	Own	Contextual	Own	Contextual	Own	Contextual
Age	-0.203 (0.018)	-0.002 (0.001)	0.171 (0.070)	-0.043 (0.017)	-0.200 (0.022)	-0.005 (0.003)	0.216 (0.090)	-0.030 (0.020)

Continued on Next Page

Table – Continued

Male	-0.097	(0.019)	-0.021	(0.010)	-0.060	(0.290)	0.081	(0.218)	-0.079	(0.034)	-0.002	(0.026)	0.058	(0.282)	-0.135	(0.227)
Black	0.082	(0.040)	-0.045	(0.009)	-2.585	(0.610)	0.185	(0.182)	0.013	(0.044)	-0.037	(0.021)	-2.475	(0.864)	0.120	(0.177)
Asian	0.215	(0.051)	-0.179	(0.057)	0.221	(0.684)	-1.372	(0.612)	0.242	(0.058)	-0.143	(0.109)	-0.890	(1.230)	-0.730	(0.866)
Hispanic	0.336	(0.035)	0.001	(0.013)	-0.981	(0.491)	0.178	(0.273)	0.175	(0.074)	-0.012	(0.041)	-0.557	(0.667)	0.047	(0.438)
Other race	-0.044	(0.020)	0.035	(0.017)	0.179	(0.572)	0.147	(0.324)	-0.065	(0.056)	-0.103	(0.054)	0.587	(0.378)	-0.016	(0.412)
Both Parents	0.131	(0.030)	0.041	(0.013)	-0.227	(0.313)	-0.190	(0.226)	0.077	(0.032)	0.020	(0.029)	-0.158	(0.350)	-0.251	(0.278)
Less HS	-0.145	(0.025)	-0.044	(0.010)	0.290	(0.440)	0.237	(0.305)	-0.069	(0.048)	-0.065	(0.036)	-0.052	(0.493)	0.296	(0.334)
More HS	0.091	(0.021)	-0.010	(0.008)	-0.701	(0.305)	-0.077	(0.235)	0.155	(0.040)	0.018	(0.027)	-0.553	(0.372)	0.008	(0.258)
Edu missing	-0.055	(0.027)	-0.037	(0.018)	0.078	(0.474)	0.495	(0.385)	-0.185	(0.037)	0.039	(0.037)	0.122	(0.549)	0.245	(0.415)
Welfare	0.183	(0.059)	-0.099	(0.020)	-0.059	(0.540)	-0.071	(0.649)	0.112	(0.052)	-0.214	(0.047)	0.435	(0.846)	0.094	(0.880)
Job missing	0.039	(0.036)	-0.053	(0.017)	-0.169	(0.424)	0.327	(0.389)	-0.027	(0.040)	-0.007	(0.037)	-0.011	(0.458)	0.100	(0.350)
Professional	0.003	(0.022)	-0.013	(0.014)	0.161	(0.287)	0.054	(0.335)	0.037	(0.045)	-0.033	(0.037)	-0.261	(0.361)	-0.002	(0.317)
Other Jobs	0.024	(0.033)	-0.027	(0.010)	-0.154	(0.322)	0.570	(0.277)	-0.023	(0.035)	-0.029	(0.024)	-0.191	(0.344)	0.435	(0.281)
Num. of students at home	-0.037	(0.016)	0.000	(0.007)	-0.183	(0.203)	-0.148	(0.140)	-0.002	(0.023)	-0.010	(0.018)	-0.221	(0.212)	-0.143	(0.151)

The MCMC runs for 100,000 iterations and the first 20,000 runs are dropped for the burn-in.

5 Conclusion

An important reason why researchers study a network structure is to analyze its impact on outcomes. As mentioned in Jackson (2011, section 5), if networks only serve as conduits for diffusion, e.g., diseases or ideas, given the network structure, their impact on outcomes is sort of mechanical and one need not worry about any feedback effects from outcomes. However, for studying the impact of a friendship network on outcomes, both the network structure and the strategic interactions between the network and outcomes should be considered. This extra consideration should be reflected in a dynamic or static equilibrium model. In this paper, we propose a static equilibrium model which takes into account these features. The modeling approach in this paper assumes that students respond to incentives stemming from interaction benefits with friends in making their friendship decisions. The empirical results show that American high school students regard the interaction benefit from academic learning as a significant incentive for forming friendships, while the incentive effect of smoking together is not found in the friendship decision. Another valuable contribution of our approach to the social interaction literature is to correct possible friendship selection biases in interaction effects.

Some issues that are not emphasized in this paper remain important for future extensions. The first is the problem of possible multiple equilibria in the simultaneous non-cooperative network formation game. We circumvent this problem in the present paper by assuming a benevolent social planner who manages the overall network links to maximize the aggregated utility, or individuals who coordinate their friendship formation processes. Those assumptions in friendship formation may be appropriate for a school setting, but are questionable for other circumstances. When discarding these assumptions, one could either provide an equilibrium selection rule or characterize the estimation problem with moment inequalities. The second issue to consider is missing links which are prevalent in empirical network data. Missing links could happen due to the specification of the network boundary, survey non-responses or the fixed choice design, e.g., nominate best ten friends by the survey design. Those three causes are all relevant to our use of the Add Health data. Kossinets (2006) uses simulation methods to examine the impact of missing links due to these causes and finds that biases of missing links in estimated network statistics due to the network boundary specification and the fixed choice design are dramatic.[31] A simple solution to overcome missing links due to network boundary is to examine results under various network boundary specifications as robustness checks. This has not been carried out in the present paper, as we face difficulty on handling computation with large networks. For dealing with missing links from the second and third causes, the likelihood-based approach (Robins *et al.*, 2004, Gile and Handcock, 2006)[32] and imputation (Huisman, 2009) provide possible solutions. For potential biases brought by missing network links in outcomes with network interactions, Chandrasekhar and Lewis (2012) and Liu (2012) provide useful discussions. The third issue to consider is the dynamic evolution of networks and outcomes. The work of Snijders *et al.* (2010) is surely leaning in that direction. Last, we are interested in applying our modeling strategies to study the formation of other types of networks, e.g., criminal network, physician referral network, or academic co-author network, and the economic activities in these networks.

[31] However, for the purpose of data collection, researchers tend to believe that if an individual is allowed to fill in as many friends as possible, that might be a difficult task for the individual, and the filled-in responses might not reflect what one would hope for from a survey. There are various opinions on this issue by survey scholars.

[32] The likelihood-based approach expresses the distribution of the observed data by

$$\Pr(W_{obs} = w_{obs}|X) = \sum_{s} \Pr(W_{obs}, W_{miss} = (w_{obs}, s)|X),$$

where $s \in \Omega(W_{miss})$ and $\Omega(W_{miss})$ contains all possible realizations of W_{miss} in the network. In practice, there are far too many realizations of s to be considered. Therefore, one could approximate the observed data likelihood via simulation.

APPENDIX A: A contraction mapping for the simultaneous Tobit-type activity

To find out the solution of Eq. (4), we may consider a contraction mapping algorithm. Denote $a \vee 0 = \max\{a, 0\}$ for a scalar a. Consider a mapping $h : R_+^{m_g} \to R_+^{m_g}$ where $R_+^{m_g} = \{Y : Y \in R^{m_g}, Y \geq 0\}$ defined by

$$h(Y) = (\lambda W_g Y + Z_g) \vee 0 = \begin{pmatrix} (\lambda w_{1.,g} Y + Z_{1,g}) \vee 0 \\ \vdots \\ \left(\lambda w_{m_g.,g} Y + Z_{m_g,g}\right) \vee 0 \end{pmatrix},$$

where $Z_g = X_g\beta + l_g\alpha_g + \epsilon_g$, $w_{i.,g}$ is the ith row of W_g, and $Z_{i,g}$ is the ith row of Z_g. For any Y_1 and Y_2 in $R_+^{m_g}$,

$$\begin{aligned} \| h(Y_1) - h(Y_2) \|_\infty &= \| ((\lambda W_g Y_1 + Z_g) \vee 0) - ((\lambda W_g Y_2 + Z_g) \vee 0) \|_\infty \\ &= \max_{i=1,\cdots,m_g} |((\lambda w_{i.,g} Y_1 + Z_{i,g}) \vee 0) - ((\lambda w_{i.,g} Y_2 + Z_{i,g}) \vee 0)| \\ &\leq \max_{i=1,\cdots,m_g} |\lambda w_{i.,g}(Y_1 - Y_2)| = \| \lambda W_g (Y_1 - Y_2) \|_\infty \\ &\leq \| \lambda W_g \|_\infty \cdot \| Y_1 - Y_2 \|_\infty . \end{aligned}$$

Thus, if $\| \lambda W_g \|_\infty < 1$, $h(Y)$ is a contraction mapping.[33] As $h(Y)$ is a contraction mapping, there exists a unique fixed point Y_g such that $h(Y_g) = Y_g$. This Y_g is the unique solution for this simultaneous Tobit equation because $Y_g = h(Y_g) = (\lambda W_g Y_g + Z_g) \vee 0$, which gives $Y_g \geq 0$, $Y_g \geq \lambda W_g Y_g + Z_g$ and $y_{i,g} = \lambda w_{i.,g} Y_g + Z_{i,g}$ whenever $y_{i,g} > 0$ for any i in the group g. This contraction mapping feature suggests a simple iterative algorithm to solve for Y_g given values of λ, W_g and Z_g.

---

[33] The assumption $||\lambda W_g||_\infty < 1$ is quite often used for the analysis of a SAR model in the spatial literature. A SAR process becomes stable in the cross section dimension under such an assumption because it implies the series expansion of the spatial transformation $(I - \lambda W)^{-1} = \sum_{l=0}^{\infty} \lambda^l W^l$, where W is a spatial weights matrix.

APPENDIX B: A Simulation experiment

In this simulation experiment, we consider that the network formation model contains the incentive effect from either a continuous activity variable or a Tobit-type variable. The artificial activity variables are generated from the network interaction models for continuous variables in Eq. (1), and Tobit-type variables in Eq. (3). The artificial networks are generated from the exponential distribution of Eq. (5) with the transferable utility in Eq. (8). The network size is fixed at 30 and we generate a total of 100 networks. In the network interaction model (for both continuous and Tobit-type variables), the exogenous variable X_i is generated from $N(0, 36)$. The group effect α is generated from $N(0, 0.5)$. The disturbance term ϵ_i is generated from $N(0, 0.5)$. For simplicity, the contextual variable WX is not included. We set the endogenous effect λ to 0.05, and the exogenous effect β to 0.50.

In the network formation model, exogenous effects for each link ij are captured by a constant term and a dyad-specific exogenous variable C_{ij} which is generated as follows: first drawing two vectors of uniform random variables from $U(0, 1)$, which are denoted as U_1 and U_2. If the i^{th} element of U_1 and the j^{th} element of U_2 are both larger than 0.7 or less than 0.3, then we set C_{ij} equal to one. Otherwise, we set it to zero. The parameters for exogenous and network structure effects are set as: $\gamma_{31} = -3.2$; $\gamma_{32} = 0.4$; $\eta_1 = 0.4$; $\eta_2 = 0.2$; $\eta_3 = -0.03$; $\eta_4 = 0.03$; $\eta_5 = 0.3$; $\eta_6 = -0.2$. For both cases of continuous and Tobit-type activity variables, the incentive effect is set to 0.3, i.e., $\delta_c = \delta_t = 0.3$. Each artificial network W is simulated by the M-H algorithm from an empty network based on $P(W|\theta)$ in Eq. (5) (See details in footnote 19). Activity variables are simulated along with the network. The M-H algorithm runs through the whole network for a total of 10,000 iterations and realizations of the network and the activity variables from the last iteration are used for the data. These generated networks have the average density equal to 0.078 (0.095 for the Tobit-type case), the average outdegree equal to 1.875 (2.270 for the Tobit-type case), and the average clustering coefficient equal to 0.117 (0.150 for the Tobit-type case).

The purpose of this simulation experiment is to demonstrate the identification of two key parameters, δ and λ, in Bayesian estimation theory through comparing the posterior distributions of parameters under different levels of data informativeness, i.e., when 20, 60, or 100 sample networks are used for estimating the model. The estimation is done by the Bayesian approach with the double M-H algorithm discussed in Section 3.3. The hyperparameters used in prior distributions are specified as follows: $\phi_0 = 0$; $\Phi_0 = 10I_9$; $\beta_0 = 0$; $B_0 = 10$; $\sigma_0 = 0.0$; $\Sigma_0 = 1.0$; $\alpha_0 = 0$; $A_0 = 100$. These parameters are designed to allow relative flat prior densities over the range of the parameter spaces.

The total of 100,000 draws from the conditional posterior distributions of parameters are simulated. we present plots of the priors and the posteriors of δ and λ based on three different network numbers in Figure (3). From both Panel (a) (the continuous case) and Panel (b) (the Tobit-type case) of Figure (3), one can observe that, our choice of hyperparameters lead to uninformative priors. However, the posterior distributions of δ and λ are getting more concentrated at the true values when more networks, i.e., more data, are used for estimation. This evidence supports the identification of these two important parameters in our model.

6 Supplementary Materials

MCMC algorithm: MCMC algorithm provides the details of each MCMC sampling step for estimating our model.

References

[1] Albert, J. H. and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of American Statistical Association*, 88, 669-679.

[2] Amemiya, T. (1974), "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables are Truncated Normal," *Econometrica*, 42, 999-1012.

[3] Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006), "Who's Who in Netowrks. Wanted: the Key Player," *Econometrica*, 74, 1403-1417.

[4] Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2010), "Delinquent Networks," *Journal of the European Economic Associations*, 8, 34-61.

[5] Bartz, K., Blitzstein, J. K., and Liu, J. S. (2008), "Monte Carlo Maximum Likelihood for Exponential Random Graph Models: From Snowballs to Umbrella Densities," Unpublished paper.

[6] Bayer, P., Ross, S. L., and Topa, G. (2008). "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 116, 1150-1196.

[7] Bayer, P., Hjalmarsson, R., and Pozen, D. (2009), "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections," *Quarterly Journal of Economics*, 124, 105-147.

[8] Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society B*, 36, 192-236.

[9] Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2010), "Do Peers Affect Student achievement? Evidence from Canada Using Group Size Variation," IZA Discussion Papers 4723, Institute for the Study of Labor (IZA).

[10] Bramoullé, Y., Djebbari, H., and Fortin, B. (2009), "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150, 41-55.

[11] Bramouillé, Y., and Fortin, B. (2009), "The Econometrics of Social Networks," Cahiers de recherche 0913, CIRPEE.

[12] Caimo, A., and Friel, N. (2010), "Bayesian Inference for Exponential Random Graph Models," *Social Networks*, 33, 41-55.

[13] Calvó-Armengol, A., and Jackson, M. O. (2004), "The Effects of Social Networks on Employment and Inequality," *American Economic Review*, 94, 426-454.

[14] Calvó-Armengol, A., and Jackson, M. O. (2007), "Networks in Labor Markets: Wage and Employment Dynamics and Inequality," *Journal of Economic Theory*, 132, 27-46.

[15] Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). "Peer Effects and Social Networks in Education," *Review of Economic Studies*, 76, 1239-1267.

[16] Chandrasekhar A. G., and Lewis, R. (2012), "Econometrics of Sampled Networks," Working paper.

[17] Chib, S., and Greenberg, E. (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 409-31.

[18] Christakis, N.A., and Fowler, J.H. (2007), "The Spread of Obesity in a Large Social Network Over 32 Years," *New England Journal of Medicine*, 357, 370-379.

[19] Christakis, N. A., Fowler, J. H., Imbens, G. W., and Jalyanaraman, K. (2010), "An Empirical Model for Strategic Network Formation," Working paper.

[20] Clark, A., and Loheac, Y. (2007), "It wasn't Me. It was Them! Social Influence in Risky Behaviors among Young Adults," *Journal of Health Economics*, 26, 763-784.

[21] Conley, T. G., and Udry, C. R. (2001), "Social Learning through Networks: The Adoption of New Agricultural Technologies in Ghana," *American Journal of Agricultural Economics*, 83, 668-673.

[22] Conley, T. G., and Udry, C. R. (2010), "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100, 35-69.

[23] Comola, M. (2008), "The Network Structure of Informal Arrangements: Evidence from Rural Tanzania," Working paper.

[24] Currarini, S., Jackson, M. O., and Pin, P. (2010), "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proceedings of the National Academy of Sciences of the United States of America*, 107, 4857-61.

[25] Fafchamps, M., and Gubert, F. (2007a), "The Formation of Risk Sharing Networks," *Journal of Development Economics*, 83, 326-350.

[26] Fafchamps, M., and Gubert, F. (2007b), "Risk Sharing and Network Formation," *The American Economic Review*, 97, 75-79.

[27] Fletcher, J. M. (2010), "Social Interactions and Smoking: Evidence Using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects," *Journal of Health Economics*, 19, 466-484.

[28] Fosco, C., and Mengel, F. (2011), "Cooperation through Imitation and Exlucsion in Networks," *Journal of Economic Dynamics and Control*, 35, 641-658.

[29] Fowler, J. H., and Christakis, N.A. (2008), "Estimating Peer Effects on Health in Social Networks," *Journal of Health Economics*, 27, 1386-1391.

[30] Frank, O., and Strauss, D. (1986). "Markov Graphs," *Journal of the American Statistical Association*, 81, 832-42.

[31] Fu, F., Hauert, C., Nowak, M. A., and Wang, L. (2008), "Reputation-Based Partner Choice Promotes Cooperation in Social Networks," *Physical Review*, 78, 026117.

[32] Gaviria, A., and Raphael, S. (2001). "School-based Peer Effects and Juvenile Behavior," *Review of Economics and Statistics*, 83, 257-268.

[33] Geyer, C. J., and Thompson, E. A. (1992). "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *Journal of the Royal Statistical Society. Series B*, 54, 657-99.

[34] Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to calculating posterior moments," *Bayesian Statistics* 4, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith). Clarendon Press, Oxford, UK.

[35] Gile, K., and Handcock, M. S. (2006), "Model-based Assessment of the Impacts of Missing Data on Inference for Networks," working paper No. 66, Center for Statistics and Social Sciences, University of Washington.

[36] Goldenberg, A, Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2009), "A Survey of Statistical Network Models," *Foundations and Trends in Machine Learning* 2, 1-117.

[37] Goldsmith-Pinkham, P., and Imbens, G. W. (2011). "Social networks and the identification of peer effects," Working paper.

[38] Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003), "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics*, 18, 527-544.

[39] Hoxby, C. (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working paper 7867.

[40] Hruschka, D. J., and Henrich, J. (2006), "Friendship, Cliquishness, and the Emergence of Cooperation," *Journal of Theoretical Biology*, 239, 1-15.

[41] Hsiao, C. (1983), "Identification," In Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Chapter 4, Vol. 1. pp.224-283. Amsterdam, North-Holland.

[42] Hsieh, C. S., and Lee, L. F. (2011), "A Social Interaction Model with Endogenous Friendship Formation and Selectivity," Working paper.

[43] Huisman, M (2009), "Imputation of missing network data: some simple procedures," *Journal of Social Structure*, 10.

[44] Jackson, M. O., and Wolinsky, J. (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44-74.

[45] Jackson, M. O. (2008), "Social Networks in Economics," in *Handbook of Social Economics*, J. Benhabib, A. Bisin, and M. O. Jackson, Eds., New York: Elsevier.

[46] Jackson, M. O. (2009), "Social Networks in Economics," *The Annual Review of Economics*, 1, 489-513.

[47] Jackson, M. O. (2011), "An Overview of Social Networks and Economic Applications," in *Handbook of Social Economics*, Eds J. Benhabib, A. Bisin, and M. O. Jackson., North Holland.

[48] Kadane, J. B. (1974), "The Role of Identification in Bayesian Theory," in *Studies in Bayesian Econometrics and Statistics*, eds. S. E. Fienberg and A. Zellner, Amsterdam, North-Holland Publishing Co.

[49] Kelejian, H. H., and Prucha, I. R. (2010), "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, 157, 53-67.

[50] Kovářík, J., and van der Leij, M. J. (2012), "Risk Aversion and Social Networks," Working paper.

[51] Kossinets, G. (2006), "Effects of Missing Data in Social Networks," *Social Networks*, 28, 247-268.

[52] Lee, L. F., Li, J., and Lin, X. (2013), "Binary Choice Models with Social Network under Heterogeneous Rational Expectations," Forthcoming in *Review of Economics and Statistics* 2013.

[53] Lee, L. F., Liu, X., and Lin, X. (2010), "Specification and Estimation of Social Interaction Models with Network Structures," *The Econometrics Journal*, 13, 145-76.

[54] Liang, F. (2010), "A Double Metropolis-Hastings Sampler for Spatial Models with Intractable Normalizing Constants," *Journal of Statistical Computating and Simulation*, 80, 1007-1022.

[55] Lin, Xu (2010), "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables," *Journal of Labor Economics*, 28, 825-60.

[56] Liu, X., and Lee, L. F. (2010), "GMM Estimation of Social Interaction Models with Centrality," *Journal of Econometrics*, 159, 99-115.

[57] Liu, X., Patacchini, E., and Zenou, Y. (2011), "Peer Effects in Education, Sport, and Screen Activities: Local Aggregate or Local Average," Working paper.

[58] Lundberg, P. (2006), "Having the Wrong Friends? Peer Effects in Adolescent Substance Use," *Journal of Health Economics*, 25, 214-233.

[59] Majolo, B., Ames, K., Brumption, R., Garratt, R., Hall, K., and N. Wilson (2006), "Human Friednship Favours Cooperation in the Iterated Prisoner's Dilemma," *Behaviour*, 143, 1383-1395.

[60] Manski, C. F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-42.

[61] Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995), Microeconomic Theory. Oxford University Press, New York.

[62] Mayer, A. and Puller, S. L. (2008), "The Old Boy (and Girl) Network: Social Network Formation on University Campuses," *Journal of Public Economics*, 92, 329-247.

[63] Mele, Angelo (2010), "A Structure Model of Segregation in Social Networks," Working paper.

[64] Moffitt, R. A. (2001), "Policy Interventions, Low-Level Qquilibria, and Social Interactions," In *Social Dynamics* Eds. Durlauf, S., and H. P. Young, MIT Press, Combridge, MA, 45-82.

[65] Murray, I., Ghahramani, Z., and Mackay, D. J. C. (2006), "MCMC for Doubly-Intractable Distributions," In Proceedings of the 22nd Annual Conference on Uncertainity in Artificial Intelligence (UAI).

[66] Patacchini, E., and Zenou, Y. (2008). "The Strength of Weak Ties in Crime," *European Economic Review*, 52, 209-236.

[67] Patacchini, E., and Zenou, Y. (2012), "Juvenile Delinquency and Conformism," *Journal of Law, Economics, and Organization*, 28, 1-31.

[68] Peck, J. R. (1993), "Friendship and the Evoultion of Co-operation," *Journal of Theoreticl Biology*, 162, 195-228.

[69] Poirier, D. J., and Tobias, J. L. (2003), "On the Predictive Distributions of Outcome Gains in the Presence of an Unidentified Parameter," *Journal of Business and Economic Statistics*, 21. 258-268.

[70] Powell, L., Taurus, J., and Ross, H. (2005), "The Importance of Peer Effects, Cigarette Prices, and Tobacco Control Policies for Youth Smoking Behavior," *Journal of Health Economics*, 24, 950-968.

[71] Prop, J. G., and Wilson, D. B. (1996), "Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics," *Random Structures and Algorithms*, 9, 223-52.

[72] Robins, G., Pattison, P., and Woolcock, J. (2004), "Missing data in networks: Exponential random graph (p^*) models for networks with non-respondents," *Social Networks*, 26, 257-283.

[73] Sacerdote, B. (2001), "Peer Effects with Random Assignments: Results from Dartmouth Roommates," *Quarterly Journal of Economics*, 116, 681-704.

[74] Smith, T. E., and LeSage, J. P. (2004), "A Bayesian Probit Model with Spatial Dependencies," In *Spatial and Spatiotemporal Econometrics. Advances in Econometrics*, Eds J. P. LeSage and R. K. Pace, Vol. 18. Elsevier Science, Oxford, UK, 127V160.

[75] Snijders, T. A. B. (2002), "Markov chain Monte Carlo estimation of exponential random graph models," *Journal of Social Structure*, 3.2. Available from http://www2.heinz.cmu.edu/project/INSNA/joss/index1.html.

[76] Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006), "New specifications for exponential random graph models," *Sociological Methodology*, 36, 99-153.

[77] Snijders, T. A. B., van de Bunt, G. G., and Stehlich, C. E. G. (2010), "Introduction to Stochastic Actor-based Models for Network Dynamics," *Social Networks*, 32, 44-60.

[78] Strauss, D., and Ikeda, M. (1990), "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, 85, 204-12.

[79] Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions (with discussion)," *Annuals of Statistics*, 22, 1701-62.

[80] Udry, J. R. (2003), The National Longitudinal Study of Adolescent Health (add health), waves 1.

[81] Ule, A. (2008). *Partner Choice and Cooperation in Networks – Theory and Experiment Evidence*, Springer-Verlag Berlin Heidelberg.

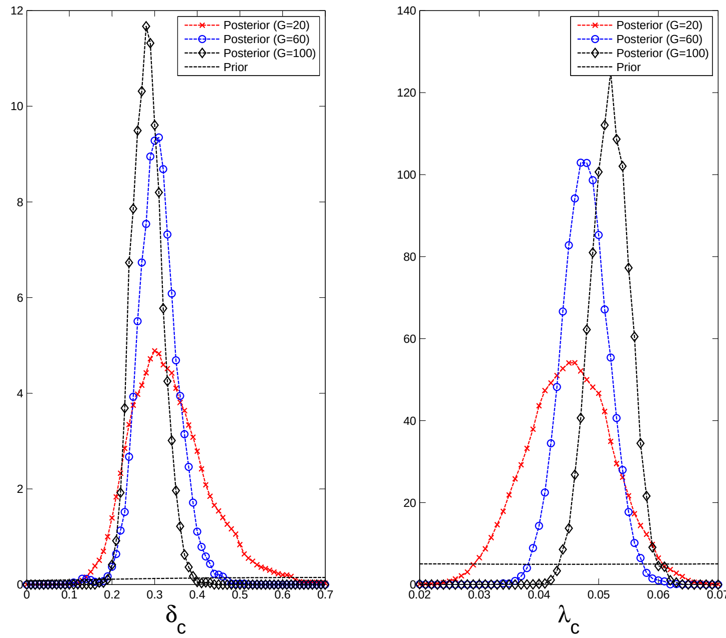
[82] Wasserman, S. S., and Pattison, P. E. (1996), "Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^*," *Psychometrika*, 61, 401-25.

[83] Weinberg, B. A. (2008), "Social Interactions with Endogenous Associations," NBER Working Paper Series, Vol. w13038.

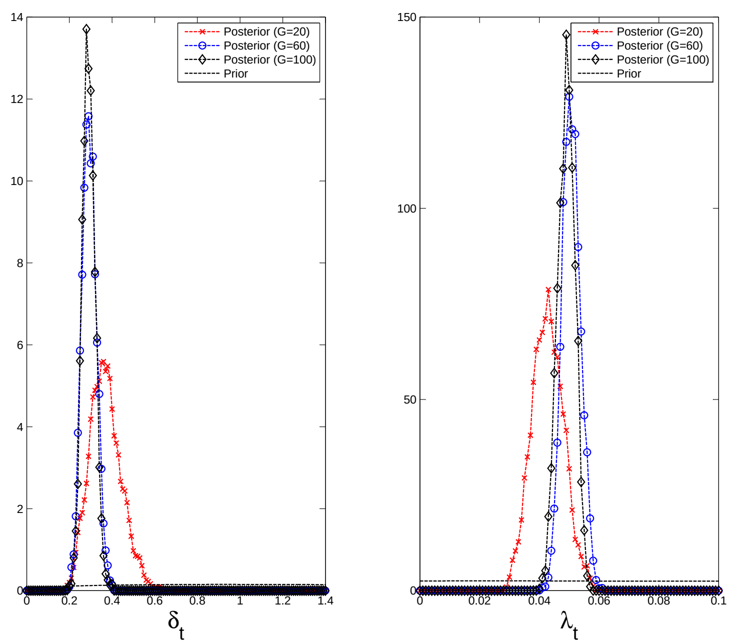
[84] Zimmerman, D. J. (2003), "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 85, 9-23.

7 Acknowledgement

This work was supported in part by an allocation of computing time from the Ohio Supercomputer Center.



(a) The prior and posterior distributions of δ_c and λ_c



(b) The prior and posterior distributions of δ_t and λ_t

Figure 3: The plot of the prior and the posterior distributions