A Structural Modeling Approach for Network Formation and Social Interactions – with Applications to Students' Friendship Choices and Selectivity on Activities

Chih-Sheng Hsieh and Lung-Fei Lee

Job Market Paper

Department of Economics
Ohio State University
Columbus, Ohio, USA

[**Preliminary Version: please do not quote without explicit permission from the authors**]

Abstract

In this paper, we propose an empirical modeling approach for social networks. To embed the concept of strategic network formation into the model, we motivate the model specification by an economic game in which individuals choose friendship links and economic activities with interactions. By allowing individuals to respond to economic incentives stemming from friend interactions on certain activities when making friendship decisions, our structural setting generates the following two advantages. First, one can evaluate the importance of economic incentives from certain interactions when individuals choose their friends. Second, the possible friendship selection bias in interaction outcomes will be corrected when the network formation is explicitly modeled. The proposed model is applied to American high school students' friendship networks in the Add Health data. From two activity outcomes, namely, students' GPAs and smoking frequencies, we find a significant economic incentive effect from GPA, but not from smoking, on friendship formation. These results suggest that helping one another on academic learning is important for forming friendships, while joy of smoking together may be not. However, outcomes of both GPA and smoking frequency are subject to significant positive network interactions.

JEL classification: C21, C25, I21, J13

Keywords: network formation, network interaction, selectivity, Bayesian estimation

1 Introduction

Economic research on social networks has grown rapidly over the past two decades. For many economic issues, the role of social networks as a major channel to disseminate information or facilitate activities is revealed.[1] Accompanying with wide applications of network concepts in economics, an immediate question faced by both theorists and practitioners is to understand how networks are formed. This question is not just interesting in its own right, but also important to analyze how changes of network structures affect economic outcomes.[2] From 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 enthusiastically apply this concept on building network formation models and discuss 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.[3]

In this paper, an empirical modeling approach for static networks is proposed in response to the literature.[4] A static network refers to a cross sectional case where only one observation of a network

[1]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, etc.)

[2]As a network might be formed in order to achieve favorable economic consequences, there is an empirical need to correct for possible endogeneity bias in network (or peer) effects on outcomes due to friendship selection. In the context of network interactions, no matter whether the research objects are labors, adolescents, or delinquents, in order to understand peer effects among these groups, one would like to know how individuals choose their friends. The choice of friendships might amplify observed peer interaction effects due to related unobserved factors behind both decisions of friendships and economic activities (Weinberg, 2008). Hence, to study network interaction effects on economic outcomes without considering endogenous friendship selections might lead to an upward biased estimate. With regard to this problem, Hsieh and Lee (2012) propose a latent variables approach in order to capture possibly important unobserved driving factors and use them to link network formation and social or peer interactions on economic activity outcomes.

[3]Some existing empirical 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), and Hsieh and Lee (2012) study friendship networks of American high school and college students.

[4]Except the economic literature, network formation models for static networks are also developed in the statistical literature. One example is the exponential random graph model (ERGM) proposed by Frank and Struss (1986), or

is available.[5] The specification of the model, motivated from the concept of strategic network formation, is based on an economic game where individuals determine both friendship links and economic activities with interactions. This game setting is modified from Hsieh and Lee (2012) by adding a device that individuals respond to economic incentives stemming from friend interactions on certain activities when making friendship decisions. This device is meaningful because in most of the survey data which contain friendship information, respondents are asked to nominate friends in general, but not for any specific purpose. Therefore, it remains interesting to see which activities with network interactions would provide significant economic incentives for forming friendships. The advantage of modeling both the network formation and network interactions on outcomes under a structural framework is in fact twofold. We can not only evaluate the importance of individuals' incentives as they stem from choosing their friends, the resulting model can also correct possible friendship selection biases in outcomes with interactions. We apply this modeling approach to study American high school students' friendship networks in the Add Health data. From two activity outcomes that are considered in the paper, namely, a student's GPA and how frequently a student smokes in an usual week, we find a significant economic incentive effect from GPA but not from smoking, which suggests that helping one another on academic learning is a factor for building friendships, while joy of smoking together is not. Furthermore, estimated endogenous effects from outcome equations under our structural setting are smaller than those from studying outcome equations alone with networks assumed to be exogenously given. The latter shows our structural approach is effective in correcting possible upward biases in interaction effects due to endogenous friendship selections.

The challenge of modeling static networks by the concept of strategic network formation comes from the fact that, as contrary to a dynamic setting, neither the order of linking nor how each link depends on previous existing links are observed. Therefore, one would model a network as a polychotomous choice with $2^{m(m-1)}$ alternatives jointly by individuals, where m is the size of the network. The estimation of such a model is computationally intensive because discrete choice

more generally, the p^* model by Wasserman and Pattison (1996), which formulates the propensity of the network formation from observed network features, such as stars, triangles, etc. Another example is the latent position model by Hoff *et al.* (2002) and Handcock *et al.* (2007), which introduces unobserved latent variables to create dependence between links and use them to visualize the network. See survey of statistical models in Goldenber *et al.* (2009). The main concern of using those statistical network models from an economic view is that they do not provide causal interpretations, which are, however, the central spirit of an economic study.

[5] 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).

variables increase exponentially with the size of the network. Furthermore, the implied model might contain multiple equilibria which requires investigators to specify an equilibrium selection rule or treat the model as an incomplete one. To prevent all of those complexities, few empirical studies choose to assume pairwise independence between network links. For example, Fafchamps and Gubert (2007a, 2007b) and Comola (2008) rely on the pairwise independence assumption which allows them to focus on individual and dyad-specific variables to explain network links. The estimation of those models can be done by a standard maximum likelihood approach since the likelihood of the whole network is just the product of likelihoods from each pairwise links. However, as noted by Bramoullé and Fortin (2009), the assumption of pairwise independence is too strong because it requires that the latent utility behind each pairwise link is separable.[6] One way suggested in Christakis *et al.* (2010) to bypass the problem of multiple equilibria from modeling networks without the pairwise independence assumption is to augment networks with a dynamic formation process. They assume a sequential process which only allows a single pair of individuals to establish (or terminate) a link at each period. By additionally assuming that individuals only concern the current state of the network but not future, they simplify the computation and eliminate the concern of multiple equilibria. With an artificial order of meetings simulated within a finite number of periods, they estimate model parameters using the Bayesian approach.

Mele (2010) also models static networks with a dynamic formation process. But instead of simulating artificially sequential meetings during estimation, he uses a random meeting technology and shows that observed networks can be viewed as realizations from a stable equilibrium distribution. His model is relevant to us as he also uses an economic game with individual utility maximization. However, there are three major differences between Mele's modeling approach and ours. First, the utility function proposed in Mele (2010) is restrictive. For the purpose of characterizing the network formation game as a potential game, which has an advantage of summarizing individual incentives by an aggregated potential function, he only include certain specific terms in the individual utility function. In this paper we show that, if one begins with specifying a negative potential (negpotential) function in an exponential distribution framework, an implied individual utility function from the negpotential function can be used to justify network formation as either a cooperative game or a non-cooperative game. The specification of the negpotential function can be very general, which is allowed to capture any relevant network characteristics or individual incentives. Also, any constraints on parameters in the implied individual utility can be understood.

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

Second, we focus on a static game with complete information rather than a dynamic game and therefore do not require any random meeting technology. Third, we primarily concern modeling both network and activity decision processes, which is more general than only a network decision process considered in Mele (2010).

The remainder of this paper is organized as follows. Section 2 presents an economic game of network formation which motivates the empirical model. 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 in the Add Health data. The paper is concluded in Section 5. We leave details of the MCMC sampling procedures, techniques and a simulation study in appendices.

2 An economic game for network formation

Assume individuals make their decisions on friendships and economic activities in a two-stage process.[7] In the first stage, individuals choose friends to maximize their link-associated utilities. In the second stage, individuals interact with their friends and choose economic activities to maximize their activity-associated utilities. One of our focuses on friendship formation is to consider the possibility that friendship be built upon pursuing economic incentives from friend interactions, e.g., students get along with other students with intention to help one another on academic learning, or delinquents hang out with other delinquents with intention to share knowledge or interests of their behaviors, etc. We allow the link-associated utility to contain economic incentives represented by activity-associated utilities which will be realized in the second stage. This two-stage process can be characterized as a two-stage static game. Individuals adopt strategies on choosing friends and economic activities in order to obtain utilities as payoffs of the game. There is perfect information between the two-stages. Within each stage, players move simultaneously with complete information. 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 as follows: First, determine equilibrium activity outcomes in the second stage with network interactions and

[7] One may interpret our modeling approach from simultaneous decisions on friendships and economic activities. In such a case, both types of decisions are made simultaneously and would affect each other. By interpreting it as a two-stage process, one can focus on friendship decisions in the first stage and activity decisions in the second stage. It gives us an advantage to emphasize the importance of economic incentives stemming from interactions on activity outcomes in the second stage.

calculate the corresponding equilibrium activity-associated utilities. Second, by incorporating the equilibrium activity-associated utilities into the link-associated utilities in the first stage, solve for the equilibrium network.

Assume that individuals are placed in a pre-specified group setting, such as students in a school-grade, workers in a company, delinquents in a neighborhood, etc.[8] Let W_g be a $m_g \times m_g$ adjacency matrix (sociomatrix) representing a network of m_g individuals (size) in the group g, where $g = 1, \cdots, G$ with G being the total number of groups in a sample. Each $w_{ij,g}$ is a dichotomous indicator which equals to one if individual i is making individual j as a friend and zero if not. The diagonal elements $w_{ii,g}$'s are set to structural zeros. 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 is a collection of such vectors in the group g. For economic activity outcomes, we consider two types of variables, continuous and Tobit-type.[9] Let $y_{i,cg}$ ($y_{i,tg}$) denotes individual i's continuous (Tobit-type) activity outcome in the group g, then $Y_{cg} = (y_{1,cg}, y_{2,cg}, \cdots, y_{m_g,cg})'$ and $Y_{tg} = (y_{1,tg}, y_{2,tg}, \cdots, y_{m_g,tg})'$ are m_g-dimensional column vectors for all members' continuous and Tobit-type outcomes in the group g.

2.1 Activity-associated utility with interaction effects – continuous and Tobit-type outcomes

We may first discuss the pursuit of economic activities in the second stage of the game.

Continuous Outcomes

Consider a continuous activity outcome variable $y_{i,cg}$. Depending on a network W_g, a profile of economic activities Y_{cg}, an unobserved group fixed-effect α_{cg}, and an idiosyncratic shock $\epsilon_{i,cg}$, which is known among individuals but not to the econometrician, an individual's activity-associated utility is assumed to take the following quadratic form:

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

[8] It would be of an interest to consider in certain situations that a specific group can be selected instead of given. Here we are focusing on the setting with reference to the Add Health data where grade as a group would be less subject to selectivity. Those issues of selection into groups will be left for future study.

[9] The case of binary variable will not be considered in the main article because of various model specifications of it, which deserve more detailed consideration. We have a discussion of it in the Appendix A.

for $i = 1, \cdots, m_g$. This quadratic utility function has been widely applied in the studies of peer effects, including Ballester *et al.* (2006), Calvó-Armengol *et al.* (2009), etc. The first and the second terms show that the utility is concave in the individual's own activity. The third term reflects a complementary effect (competitive effect) from peer's activities if $\lambda_c \geq 0$ ($\lambda_c \leq 0$). By the theorem of Ballester *et al.* (2006), as long as $|\lambda_c|$ is less than the largest eigenvalue of W_g,[10] an unique interior Nash equilibrium of activity outcomes from this simultaneous-move subgame is given by

$$\begin{aligned} y^*_{i,cg}(W_g) &= x_{i,g}\beta_{1c} + \sum_{j=1}^{m_g} w_{ij,g}x_{j,g}\beta_{2c} + \alpha_{cg} + \epsilon_{i,cg} \\ &+ \sum_{k=1}^{\infty} \lambda_c^k \sum_{j=1}^{m_g} \left(W_g^k\right)_{ij} \left(x_{j,g}\beta_{1c} + \sum_{l=1}^{m_g} w_{jl,g}x_{l,g}\beta_{2c} + \alpha_{cg} + \epsilon_{j,cg} \right), \end{aligned} \tag{2}$$

for $i = 1, \cdots, m_g$ and the corresponding activity-associated utility is $u_{i,cg}(y^*_{i,cg}(W_g)) = \frac{1}{2}y^{*2}_{i,cg}(W_g)$. By stacking individual equilibrium activity outcomes from Eq. (2), one can obtain a vector of equilibrium outcomes:

$$Y^*_{cg}(W_g) = S^{-1}_{cg}(W_g)(\mathbf{X}_g\beta_c + l_g\alpha_{cg} + \epsilon_{cg}), \tag{3}$$

where $S_{cg}(W_g) = I_{m_g} - \lambda_c W_g$, $\mathbf{X}_g = (X_g, W_g X_g)$, $\beta_c = (\beta'_{1c}, \beta'_{2c})'$, $\epsilon_{cg} = (\epsilon_{1,cg}, \cdots, \epsilon_{m_g,cg})'$, and l_g being a m_g-dimensional vector of ones. Eq. (3) can be recognized as the reduced form of the spatial autoregressive (SAR) model used in Lee *et al.* (2010), Lin (2010), Hsieh and Lee (2012), etc., on studies of network interactions.[11]

Tobit-type Outcomes

In certain cases an activity outcome might be continuous but nonnegative, i.e., a Tobit-type variable which is left-censored at the value zero. To model Tobit-type activity outcomes, we should impose a constraint, $y_{i,tg} \geq 0$, in the individual activity-associated utility of Eq (1) with $y_{i,cg}$ replaced by $y_{i,tg}$. Under this constraint, the Nash equilibrium of the outcome vector can be summarized by the equation:

$$Y^*_{tg}(W_g) = \max\left(0, \ddot{Y}^*_{tg}\right) \quad \text{with} \quad \ddot{Y}^*_{tg} = \lambda_t W_g Y^*_{tg} + \mathbf{X}_g\beta_t + l_g\alpha_{tg} + \epsilon_{tg}, \tag{4}$$

[10] One may also use a slightly stronger sufficient condition, $\| \lambda W_g \|_\infty < 1$, which is considered in Liu and Lee (2010).

[11] The SAR model has the specification $Y_{cg} = \lambda_c W_g Y_{cg} + \mathbf{X}_g\beta_c + l_g\alpha_{cg} + \epsilon_{cg}$, where the coefficient λ_c represents the endogenous effect, β_c represents the own and contextual effects of exogenous regressors. The group fixed-effect α_{cg} captures unobserved environmental factors shared by group members and hence will control for correlated effects and possible selections into a group.

where $\ddot{Y}^*_{tg}$ represents a vector of latent variables. We may call Y^*_{tg} the simultaneous-Tobit outcome.[12] The solution Y^*_{tg} must satisfy $Y^*_{tg} \geq \lambda_t W_g Y^*_{tg} + \mathbf{X}_g \beta_t + l_g \alpha_{tg} + \epsilon_{tg}$ such that $Y^*_{tg} \geq 0$, 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 $y^*_{i,tg} > 0$. Under the conditions as in Amemiya (1974) for a general simultaneous Tobit equation system,[13] the solution Y^*_{tg} is unique and can be obtained from a constrained quadratic programming problem

$$Y^*_{tg} = \min_{Y_{tg}} \left\{ \begin{array}{l} Y'_{tg}[(I_{m_g} - \lambda_t W_g)Y_{tg} - \mathbf{X}_g\beta_t - l_g\alpha_{tg} - \epsilon_{tg}] : \\ Y_{tg} \geq 0, (I_{m_g} - \lambda_t W_g)Y_{tg} - \mathbf{X}_g\beta_t - l_g\alpha_{tg} - \epsilon_{tg} \geq 0 \end{array} \right\}. \tag{5}$$

As an alternative, we show that with proper restricted parameter space on λ_t, the solution can be conveniently obtained via a contraction mapping algorithm provided in the Appendix B.

2.2 Link-associated utility and the exponential probability distribution

Back to the first stage of the game where individuals make their friendship decisions, an individual link-associated utility $v_{i,g}$ from a network g may be specified to capture various observed network characteristics as well as economic incentives stemmed from the second stage. As in a static game setting with complete information, individual i's friendship choices may depend on network links of other individuals in the group. For example, individuals of popularity in a network may attract more links. If this static game is non-cooperative, with a specified individual utility function $v_{i,g}$, his/her choices will be

$$\max_{w_{i\cdot,g}} v_{i,g}\left(w_{i\cdot,g}, W_{-i\cdot,g}\right), \tag{6}$$

where $w_{i\cdot,g}$ represents the i^{th} row of W_g, and $W_{-i\cdot,g}$ is the W_g with its i^{th} row removed. Due to the nature of simultaneous moves, an equilibrium of this game determined by Eq. (6) for each individual in the group will be characterized as a Nash equilibrium. As the system via Eq. (6) is a simultaneous discrete choice one, there can be multiple Nash equilibria. However, with properly specified utility functions, the existence of an unique Nash equilibrium is possible. Instead of a

[12] Another possible way to handle the Tobit-type activity outcome is to assume individuals choose latent activities based on Eq (1) without the nonnegative constraint, but only the Tobit-type outcomes are observed, i.e., the observed outcome $y^*_{i,tg} = \max\{0, \ddot{y}^*_{i,tg}\}$ with a latent variable $\ddot{y}^*_{i,tg}$ from Eq (1). We may call this case the latent-Tobit outcome to distinguish it from the simultaneous-Tobit outcome.

[13] A sufficient condition for a unique solution of this quadratic programming problem is that the 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 $i = 1, \cdots, m_g$.

non-cooperative game, we may also consider this static game to be cooperative in which an unique equilibrium will be guaranteed with maximization of the aggregated utility. Both game concepts provide useful insights for us to model the observed equilibrium network and hence we pursue both strategies and show how they can be related.

We prefer that our empirical specification of the utility function will result in a statistical estimation framework which describes an observed network W_g as a realization from an exponential distribution on a network space. The exponential distribution framework is basic for random graph and p^* network models in the statistical literature (Frank and Struss, 1986; Wasserman and Pattison, 1996). A network space Ω_g for a group g consists of all possible network patterns W for that group. An exponential distribution for W_g has a probability specification in the form

$$P(W_g) = \exp(Q_g(W_g)) / \sum_{W \in \Omega_g} \exp(Q_g(W)) \tag{7}$$

for a function Q_g. A specification of the function Q_g gives a specific exponential distribution of networks. In the statistical mechanics literature (e.g., Ruelle 1969), $-Q_g$ plays a role of a potential energy function, and Q_g is referred to as a negpotential function. In the random graph statistical literature, $Q_g(W_g)$ may consist of various network statistics. The exponential probability function implies that conditional probabilities will also take the logistic form, which gives rise the notion of a Markov random field in spatial statistics (see, e.g., Cressie, 1993). The interest on the exponential distribution for networks is due to its computational tractability as witnessed by the statistical literature (See Strauss and Ikeda, 1990; Geyer and Thompson, 1992; Snijders, 2002; Liang, 2010). To incorporate economic rationality for this distribution, one may allow an individual's utility $v_{i,g}$ to incorporate observed or unobserved (from an econometrician's view) individual characteristics, and relevant network characteristics. The deterministic components of $v_{i,g}$ for all i in the group g will then be part of a negpotential function.

The specification of a model via a negpotential function in the exponential distribution will ensure the implied statistical model is model coherent, i.e., the econometric model has a well defined probability structure for an observed equilibrium network. For a multivariate system with discrete choices or limited dependent variables, a coherent model reflects an unique equilibrium generated by the system (see, e.g, Amemiya (1974) on a multivariate Tobit model). For our network formation process, in order to have the existence of an unique equilibrium, a possible modeling strategy is to embed individual link utilities into a negpotential function and use the specified negpotential function to study the network as a whole.

2.2.1 The specification of an individual link-associated utility

As motivated by several economic and statistical studies on network formation, we consider the following individual link-associated utility specification,

$$v_{i,g}(W_g) = \underbrace{\sum_{j=1}^{m_g} w_{ij,g}\psi_{ij,g}}_{\text{Exogenous Effects}} + \underbrace{\varpi_{i,g}(w_{i\cdot,g}, W_{-i\cdot,g})\eta}_{\text{Network Structure Effects}} + \underbrace{\sum_{d=1}^{\bar{d}} \frac{\delta_d}{2} y^{*2}_{i,dg}(W_g)}_{\text{Economic Incentive Effects}} \quad . \tag{8}$$

In Eq. (8), the exogenous effects capture influences from individual-specific and dyad-specific exogenous characteristics on the link utility. The function $\psi_{ij,g}$ has an explicit expression

$$\psi_{ij,g} = c_{i,g}\gamma_1 + c_{j,g}\gamma_2 + c_{ij,g}\gamma_3. \tag{9}$$

The variable $c_{i,g}$ ($c_{j,g}$) in Eq. (9) is a $\bar{s}$-dimensional row vector 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 the group g to capture the utility from homophily of observed characteristics in friendship formation. The idea of using $c_{i,g}$, $c_{j,g}$ and $c_{ij,g}$ in explaining the link decisions is from Fafchamps and Gubert (2007a, 2007b) in the study of risk-sharing network formation. For notational simplicity, we let $C_g = \{c_{i,g}, c_{j,g}, c_{ij,g}\}$ and $\gamma = (\gamma_1', \gamma_2', \gamma_3')'$. The network structure effects in Eq. (8) capture influences from link dependence within the network on individual i's 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 components of W_g and η is a corresponding vector of coefficients. The idea of considering network structure effects in Eq. (8) comes from the p^* models (Wasserman and Pattison, 1996; Snijders *et al.*, 2006), the actor-based dynamic network model (Snijders *et al.*, 2010), and the model of Mele (2010). In the p^* models, summary network statistics such as the number of k-stars and k-triangles, $k \in \mathbb{N}$, are used for the network structure effects to measure how likely those network patterns appear in observed networks. The coefficients of those network structure effects do not represent causal relationships. In the actor-based dynamic network model and the model of Mele (2010), the summary statistics used for the network structure effects are economically motivated and constructed from the number of individuals' reciprocal, outward, inward, and transitive links. The coefficients of those network structure effects provide causal interpretations. The empirical specification of the network structure effects used in this paper will be discussed later in Section 4.1.

The novel effects considered in this paper for an individual link-associated utility are economic incentive effects from network interactions, which are represented by the activity-associated utilities that will be realized in the second stage. There may be several ($\bar{d}$) economic activities which provide economic incentives for forming friendships. The coefficients of these economic incentive effects are denoted as $\delta = (\delta_1, \cdots, \delta_{\bar{d}})'$. As noted by Ballester *et al.* (2006), activity-associated utilities will always increase with the number of links in the network if network interactions provide complementary effects on activity outcomes. Since individuals' link decisions partially depend on the activity-associated utilities, they might choose to add as many links as possible if there were no cost on the link formation. To mitigate such a strong incentive to form links, we rely on the existence of nontrivial negative exogenous and network structure effects which represent possible costs of adding friendship links.[14]

Based on the link utility $v_{i,g}$ in Eq. (8), its specification implies that, in particular, each pairwise friendship link from individual i to individual j will depend on whether the difference of utilities, $v_{i,g}(w_{ij,g} = 1, W_{-ij,g}) - v_{i,g}(w_{ij,g} = 0, W_{-ij,g})$, is greater than zero or not, where $W_{-ij,g}$ represents the existing links except the entry (i, j) in W_g. Using $v_{i,g}(W_g)$ defined in Eq. (8), one has

$$
\begin{aligned}
& v_{i,g}(w_{ij,g} = 1, W_{-ij,g}) - v_{i,g}(w_{ij,g} = 0, W_{-ij,g}) \\
& = \psi_{ij,g} + (\varpi_{i,g}(w_{ij,g} = 1, W_{-ij,g}) - \varpi_{i,g}(w_{ij,g} = 0, W_{-ij,g}))\eta \\
& \quad + \sum_{d=1}^{\bar{d}} \frac{\delta_d}{2} \left(y^{*2}_{i,dg}(w_{ij,g} = 1, W_{-ij,g}) - y^{*2}_{i,dg}(w_{ij,g} = 0, W_{-ij,g}) \right). \qquad (10)
\end{aligned}
$$

One should note that if $W_{-ij,g}$ were not expected to have any effects on the link decision $w_{ij,g}$, including those in the economic incentive effects, then only the exogenous effects should be included in Eq. (8). In such a case, each pairwise link decisions would be independent.

[14]Based on our economic theory, the activity outcomes $\{y_{i,dg}\}_{d=1}^{\bar{d}}$ enter into the link-associated utility of Eq. (8) through economic incentive effects. It is also possible to have another economic theory which implies that $y_{i,dg}$, $y_{j,dg}$ (or $|y_{j,dg} - y_{i,dg}|$) appear in $\psi_{ij,g}$ of Eq. (8) for capturing individual-specific (or dyadic-specific) effects. Such a theory will emphasize that activity outcomes (or the absolute difference of activity outcomes) directly affect the utility of friendship links. For example, one may consider activities which are usually engaged by one person, e.g., watching TV, playing video games, etc. 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 one of those activities, should be specified in the function $\psi_{ij,g}$ to capture the influences on the link utility. However, those activities would still be subject to friendship interactions as friends may share information about new games or TV programs so that individuals spend even more time on those activities. In another example, we can consider that the frequency of delinquent behaviors is the activity outcome. Students care $|y_{j,dg} - y_{i,dg}|$ in forming friendships as differences in levels of their delinquent behaviors could create negative effects on their utilities. In terms of estimation, having $y_{i,dg}$, $y_{j,dg}$ or $|y_{j,dg} - y_{i,dg}|$ in the link utility will not cause large changes on the Bayesain approach and corresponding MCMC algorithms proposed in this paper.

2.2.2 The negpotential function in a cooperative game

Under a cooperative game setting, an unique equilibrium of the network formation game would be a realization, W_g, chosen by a social planner which maximizes the aggregated link utility $V_g(W_g)$. By summing individual utilities from Eq (8), one has

$$\begin{aligned} V_g(W_g) &= \sum_{i=1}^{m_g} v_{i,g}(W_g) \\ &= \sum_{i=1}^{m_g}\sum_{j=1}^{m_g} w_{ij,g}\psi_{ij,g} + \sum_{i=1}^{m_g} \varpi_{i,g}(w_{i\cdot,g}, W_{-i\cdot,g})\eta + \sum_{d=1}^{\bar{d}} \frac{\delta_d}{2} Y_{dg}^*(W_g)' Y_{dg}^*(W_g). \end{aligned} \tag{11}$$

To relate $V_g(W_g)$ to the exponential probability distribution of W_g, we introduce a disturbance ξ_W for each network pattern W in Ω_g additively to $V_g(W)$. The disturbance ξ_W is assumed to be observable to the planner but not the econometrician. Thus, W_g is the formed network if and only if $V_g(W_g) + \xi_{W_g} = \max_{W\in\Omega_g}\{V_g(W) + \xi_W\}$. By assuming that ξ_W's are i.i.d. type I extreme value distributed, we have a polychotomous choice logit model with the exponential probability in Eq. (7) with the negpotential function Q_g being V_g.

2.2.3 The negpotential function in a non-cooperative game

Instead of a cooperative game, one may model the formation process as a non-cooperative game with an individual utility function $\widetilde{v}_{i,g}$ which results in an unique Nash equilibrium. Let $\xi_{w_{i\cdot}}$ be a disturbance for each link pattern of individual i in Ω_{ig}, which is the set of all link patterns for individual i in the group g. The utility maximization for link decisions of individual i is $\widetilde{v}_{i,g}(w_{i\cdot,g}) + \xi_{w_{i\cdot,g}} = \max_{w_{i\cdot}\in\Omega_{ig}}\{\widetilde{v}_{i,g}(w_{i\cdot}, W_{-i\cdot}) + \xi_{w_{i\cdot}}\}$. The specification proposed for the function $v_{i,g}$ in Eq. (8) may not be able to directly apply to $\widetilde{v}_{i,g}$ as it might cause multiple equilibria. To find a specification of $\widetilde{v}_{i,g}$ which not only includes utility components in the function $v_{i,g}$ but also ensures an unique equilibrium, one possibility is to treat the game as a potential game and use a pre-specified function Q_g for the potential function. Monderer and Shapley (1996) shows that a potential game possesses an unique pure-strategy Nash equilibrium when the equilibrium outcome maximizes the value of the potential function.[15] The exponential distribution in Eq. (7) implies the conditional probability distribution for individual i's link decisions as

$$P(w_{i\cdot,g}|W_{-i\cdot,g}) = \frac{P(w_{i\cdot,g}, W_{-i\cdot,g})}{P(W_{-i\cdot,g})} = \frac{\exp(Q_g(w_{i\cdot,g}, W_{-i\cdot,g}))}{\sum_{w_{i\cdot}\in\Omega_{ig}} \exp(Q_g(w_{i\cdot}, W_{-i\cdot}))}, \tag{12}$$

[15] If one considers local maximum values of the potential function, then there may be multiple pure-strategy or mixed-strategy equilibria.

which is also a logistic probability. As in McFadden (1973), one may give an economic justification for this probability as a result from an individual's utility maximization over discrete choice alternatives as $Q_g(w_{i.,g}, W_{-i.,g}) + \xi_{w_{i.,g}} = \max_{w_{i.} \in \Omega_{ig}} (Q_g(w_{i.}, W_{-i.}) + \xi_{w_{i.}})$, where $\xi_{w_{i.}}$'s are i.i.d. type I extreme value distributed over $\Omega_{i,g}$. Thus, we may take the function $Q_g(w_{i.,g}, W_{-i.,g})$ as the individual utility function $\widetilde{v}_{i,g}(w_{i.,g})$ in a non-cooperative game. The exponential probability distribution further implies, for each binary link decision, a binary logit as

$$P(w_{ij,g}|W_{-ij,g}) = \frac{P(w_{ij,g}, W_{-ij,g})}{P(W_{-ij,g})} = \frac{\exp(Q_g(w_{ij,g}, W_{-ij,g}))}{\exp(Q_g(w_{ij,g}, W_{-ij,g})) + \exp(Q_g(1 - w_{ij,g}, W_{-ij,g}))}. \quad (13)$$

This conditional probability may also be justified with an utility maximization setting. Individual i will make a friend with individual j if

$$Q_g(w_{ij,g} = 1, W_{-ij,g}) + \xi_{w_{ij,g}=1, W_{-ij,g}} \geq Q_g(w_{ij,g} = 0, W_{-ij,g}) + \xi_{w_{ij,g}=0, W_{-ij,g}}. \quad (14)$$

If one considers to first write down an individual utility function for the network formation game, the negpotential function can always be constructed as the sum of these pre-specified individual utility functions from the whole network. However, under few cases one may have another negpotential function which is not equal to the sum of the pre-specified individual utility functions. One example can be found in Mele (2010). In his network formation model, some selected effects from direct, mutual and indirect friends are specified in an individual utility function with constrained coefficients. Those effects correspond to the exogenous effect, reciprocality effect, receiver's expansiveness effect, and sender's popularity effect by our paper's terminology. Based on this individual utility function, which we denote as $\breve{v}_{i,g}(W_g)$, he writes down a special negpotential function $\breve{Q}_g(W_g)$ where

$$\breve{Q}_g(w_{ij,g} = 1, W_{-ij,g}) - \breve{Q}_g(w_{ij,g} = 0, W_{-ij,g}) = \breve{v}_{i,g}(w_{ij,g} = 1, W_{-ij,g}) - \breve{v}_{i,g}(w_{ij,g} = 0, W_{-ij,g}),$$

however, $\breve{Q}_g(W_g) \neq \sum_{i=1}^{m_g} \breve{v}_{i,g}(W_g)$. Such a special negpotential function would only exist when the specified individual utility function has certain constraints on included terms or coefficients, as in Mele (2010). On the contrary, the negpotential function constructed by the sum of individual utility function always exists, but one might not prevent having a number of constraints on parameters to reflect externality or for parameter identification. Instead of a pre-specified individual utility functions, our modeling strategies also suggest that one can first specify a negpotential function and use its implied utility to justify an exponential distribution specification on network formation as either a cooperative game or a non-cooperative game. The specified negpotential function can be rich enough to incorporate a variety of interesting network statistics or individual incentives and parameter constraints in the implied utility function can be understood. For empirical applications, our network model is defined as the exponential probability function of Eq. (7) with the negpotential function Q_g replaced by the aggregated link utility function V_g of Eq. (11).

2.2.4 Identification

The identification of our network model is similar to standard discrete choice models, comes from differences in utility. To identify (or estimate) parameters in the aggregated utility function V_g, it requires that the parameters in $V_g(w_{ij,g} = 1, W_{-ij,g}) - V_g(w_{ij,g} = 0, W_{-ij,g})$ can be identified (or estimated). The identification of parameters would be guaranteed as long as the summary network statistics being considered in V_g are not linearly dependent. We also need to normalize the variance of the disturbance ξ_w to one to eliminate the concern of arbitrary scaling problem in discrete choice models. After all, we require the parameters of economic incentive effects from activity outcomes to be nonnegative. This constraint is not needed for identification but can helps us to prevent the negative case which violates the spirit of our economic model.

There are two advantages of modeling the endogenous network formation and activity outcomes jointly under our structural framework. First, it allows us to study how individuals respond to economic incentives from network interactions when choosing their friends, which are revealed by the coefficients δ_d's. Second, it handles the bias problem on the interaction effects caused by friendship selections. The disturbance term ϵ_g appears in both the friendship and activity decision processes. Hence, it captures unobserved factors which contribute to these two decisions. In the next section we will discuss estimation issues of our model.

3 Model estimation

3.1 The likelihood functions of models

We first provide the likelihood function of our model with economic incentives from either a single continuous or a single Tobit-type activity outcome. Then we introduce correlations between disturbances for the bivariate case. The joint likelihood function based on this bivariate case will be used for the posterior analysis in section 3.2.

Continuous Outcomes

Given the model structure built by continuous activity outcomes of Eq. (3), the aggregated link utility of Eq. (11), and the assumption of the exponential distribution for networks, we consider the parametric approach to estimate those equations. Individual idiosyncratic shocks $\epsilon_{i,cg}$'s in the outcome equation are assumed i.i.d. normally distributed with a zero mean and a variance equal to $\sigma^2_{\epsilon_c}$. With the economic incentive effect from one continuous outcome Y^*_{cg}, the joint probability function of the outcome Y^*_{cg} and the network W^*_g can be written as

$$\begin{aligned} P(W^*_g, Y^*_{cg}|\theta_c, \alpha_{cg}) &= P(Y^*_{cg}|W^*_g, \theta_c, \alpha_{cg}) \cdot P(W^*_g|\theta_c, \alpha_{cg}) \\ &= |S_{cg}(W^*_g)| \cdot f(\epsilon_{cg}|W^*_g, \theta_c, \alpha_{cg}) \cdot P(W^*_g|\theta_c, \alpha_{cg}) \\ &= |S_{cg}(W^*_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot P(W^*_g|\epsilon_{cg}, \theta_c, \alpha_{cg}) \\ &= |S_{cg}(W^*_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot \frac{\exp(V_g(W^*_g, \epsilon_{cg}, \theta_c, \alpha_{cg}))}{\sum_W \exp\left(V_g(W, \epsilon_{cg}, \theta_c, \alpha_{cg})\right)}, \end{aligned} \tag{15}$$

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 vector of parameters.

Tobit-type Outcomes

For the simultaneous Tobit-type activity outcome, we can divide the m_g agents in the network g into two blocks such that the first m_{g1} agents have outcome variables equal to zero and the remaining agents from $m_{g1}+1$ to m_g have positive outcome variables. According to Eq. (4), the observed activity outcome vector Y^*_{tg} and network W^*_g can be conformably decomposed into

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

where $Y^*_{tg1} = 0$, $Y^*_{tg2} > 0$, and the latent variables $\ddot{Y}^*_{tg1} \leq 0$. Individual idiosyncratic shocks $\epsilon_{i,tg}$'s are assumed i.i.d. normally distributed with a zero mean and a variance equal to $\sigma^2_{\epsilon_t}$. If economic incentive effect is only from one single Tobit-type outcome, the likelihood function of Y^*_{tg} and W^*_g

can be written as

$$
\begin{aligned}
&P(Y^*_{tg}, W^*_g|\theta_t, \alpha_{tg}) \\
&= P(Y^*_{tg1} = 0, Y^*_{tg2}, W^*_g|\theta_t, \alpha_{tg}) \\
&= \int I(Y^*_{tg1} = 0, \ddot{Y}^*_{tg1}) \cdot P(\ddot{Y}^*_{tg1}, Y^*_{tg2}, W^*_g|\theta_t, \alpha_{tg}) \cdot d\ddot{Y}^*_{tg1} \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{tg2} + X_{1g}\beta_{1t} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{2t}\right)} P(\epsilon_{tg1}, Y^*_{tg2}, W^*_g|\theta_t, \alpha_{tg}) \cdot d\epsilon_{tg1} \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{tg2} + X_{1g}\beta_{1t} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{2t}\right)} \left|I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right| \cdot f(\epsilon_{tg1}, \epsilon_{tg2}|W^*_g, \theta_t, \alpha_{tg}) \cdot \\
&\qquad P(W^*_g|\theta_t, \alpha_{tg}) \cdot d\epsilon_{tg1} \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{tg2} + X_{1g}\beta_{1t} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{2t}\right)} \left|I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right| \cdot f(\epsilon_{tg1}, \epsilon_{tg2}|\theta_t, \alpha_{tg}) \cdot \\
&\qquad P(W^*_g|\epsilon_{tg1}, \epsilon_{tg2}, \theta_t, \alpha_{tg}) \cdot d\epsilon_{tg1} \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{tg2} + X_{1g}\beta_{1t} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{2t}\right)} \left|I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right| \cdot f\left(\epsilon_{tg1}, \epsilon_{tg2}|\theta_t, \alpha_{tg}\right) \cdot \\
&\qquad \frac{\exp\left(V_g(W^*_g, \epsilon_{tg1}, \epsilon_{tg2}; \theta_t, \alpha_{tg})\right)}{\sum_W \exp\left(V_g(W, \epsilon_{tg1}, \epsilon_{tg2}; \theta_t, \alpha_{tg})\right)} \cdot d\epsilon_{tg1}, \qquad (16)
\end{aligned}
$$

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. Also, $\epsilon_{tg2} = \left(I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right) Y^*_{tg2} - X_{2g}\beta_{1t} - (W^*_{21,g} X_{1g} + W^*_{22,g} X_{2g})\beta_{2t} - l_{2g}\alpha_{tg}$ and $\theta_t = (\gamma', \eta', \delta_t, \lambda_t, \beta'_t, \sigma^2_{\epsilon_t})$.

Economic incentive effects can be from $\bar{d}$ outcomes mixed with the continuous ones and the Tobit-type ones. For simplicity, considering a model of $\bar{d} = 2$ which consists of one continuous and one Tobit-type outcomes where the disturbances $\epsilon_{i,tg}$ and $\epsilon_{i,cg}$ follow a joint normal distribution,

$$
(\epsilon_{i,tg}, \epsilon_{i,cg}) \sim i.i.d.\ \mathscr{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\epsilon_t} & \sigma_{\epsilon_{tc}} \\ \sigma_{\epsilon_{ct}} & \sigma^2_{\epsilon_c} \end{pmatrix}\right), \quad i = 1, \cdots, m_g. \qquad (17)
$$

From Eq. (17), one has

$$
\epsilon_{tg} = \sigma_{\epsilon_{tc}} \sigma^{-2}_{\epsilon_c} \epsilon_{cg} + u_g, \quad u_g \sim \mathscr{N}_{m_g}(0, \sigma^2_u I_{m_g}), \qquad (18)
$$

where $\sigma^2_u = (\sigma^2_{\epsilon_t} - \sigma_{\epsilon_{tc}} \sigma^{-2}_{\epsilon_c} \sigma_{\epsilon_{ct}})$. Let $\theta_{ct} = (\gamma', \eta', \delta_c, \delta_t, \lambda_c, \lambda_t, \beta'_c, \beta'_t, \sigma^2_{\epsilon_c}, \sigma^2_{\epsilon_t}, \sigma_{\epsilon_{tc}})$, the joint likelihood function of Y^*_{tg}, Y^*_{cg} and W^*_g can be written as

$$
\begin{aligned}
&P(Y^*_{tg}, Y^*_{cg}, W^*_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg}) \\
&= P(Y^*_{tg}|Y^*_{cg}, W^*_g, \theta_{ct}, \alpha_{cg}, \alpha_{tg}) \cdot P(Y^*_{cg}|W^*_g, \theta_{ct}, \alpha_{cg}) \cdot P(W^*_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg}) \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{t2g} + X_{1g}\beta_{t1} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{t2}\right)} \left|I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right| \cdot f\left(\epsilon_{tg}|\epsilon_{cg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}\right) \cdot \\
&\quad |S_{cg}(W^*_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot P(W^*_g|\epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{cg}, \alpha_{tg}) \cdot d\epsilon_{t1g} \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{t2g} + X_{1g}\beta_{t1} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{t2}\right)} \left|I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right| \cdot f\left(u_g|\epsilon_{cg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}\right) \cdot \\
&\quad |S_{cg}(W^*_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot \frac{\exp(V_g(W^*_g, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}))}{\sum_W \exp\left(V_g(W, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg})\right)} \cdot d\epsilon_{tg1}. \qquad (19)
\end{aligned}
$$

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

$$
\begin{aligned}
&P(Y^*_{cg}, Y^*_{tg}, W^*_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg}) \\
&= \int_{-\infty}^{-\left(\lambda_t W^*_{12,g} Y^*_{t2g} + X_{1g}\beta_{t1} + (W^*_{11,g} X_{1g} + W^*_{12,g} X_{2g})\beta_{t2}\right)} \left|I_{m_g - m_{g1}} - \lambda_t W^*_{22,g}\right| \cdot f\left(\epsilon_{tg}|\theta_t, \alpha_{tg}\right) \cdot \\
&\quad |S_{cg}(W^*_g)| \cdot f(\epsilon_{cg}|\theta_c, \alpha_{cg}) \cdot \frac{\exp(V_g(W^*_g, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg}))}{\sum_W \exp\left(V_g(W, \epsilon_{cg}, \epsilon_{tg}, \theta_{ct}, \alpha_{tg}, \alpha_{cg})\right)} \cdot d\epsilon_{tg1}. \qquad (20)
\end{aligned}
$$

One 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.[16] Hence, the standard maximum likelihood estimation approach would be infeasible. This problem applies to all the p^* models 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 developed to handle an intractable normalizing term in the the posterior density function.[17]

[16]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.

[17]There are also several classical approaches that have been proposed. The first is the maximum pseudo-likelihood approach (MPL). This approach was first mentioned in Besag (1974) and later be 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 approaches. Another approach is the Monte Carlo maximum likelihood (MCML) estimation approach which simulates auxiliary networks for approximating the denominator of the exponential distribution density 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

Suppose we have a likelihood function of y given the parameter θ which takes the form $P(y|\theta) = f(y;\theta)/D(\theta)$, where $D(\theta)$ is an intractable normalizing term. In the usual Metropolis-Hastings (M-H) algorithm, we generate a new parameter $\tilde{\theta}$ from a proposal distribution $q(\cdot|\theta)$. Then updating the previous draw θ to the new draw $\tilde{\theta}$ with an acceptance probability α. Denoting $\pi(\theta)$ as the prior probability of θ, the acceptance probability α is computed as

$$\alpha(\tilde{\theta}|\theta) = \min\left\{1, \frac{P(\tilde{\theta}|y)q(\theta|\tilde{\theta})}{P(\theta|y)q(\tilde{\theta}|\theta)}\right\} = \min\left\{1, \frac{\pi(\tilde{\theta})f(y;\tilde{\theta})q(\theta|\tilde{\theta})}{\pi(\theta)f(y;\theta)q(\tilde{\theta}|\theta)} \cdot \frac{D(\theta)}{D(\tilde{\theta})}\right\}.$$

We can see that the normalizing terms are left in both the numerator and denominator and will not cancel out, so the evaluation of the acceptance-rejection criterion with α would be intractable. Murray *et al.* (2006) first show that by introducing auxiliary variables into the model, the acceptance probability can be replaced with

$$\alpha(\tilde{\theta}|\theta, x) = \min\left\{1, \frac{\pi(\tilde{\theta})P(y|\tilde{\theta})q(\theta|\tilde{\theta})}{\pi(\theta)P(y|\theta)q(\tilde{\theta}|\theta)} \cdot \frac{P(x|\theta)}{P(x|\tilde{\theta})}\right\} = \min\left\{1, \frac{\pi(\tilde{\theta})f(y;\tilde{\theta})q(\theta|\tilde{\theta})}{\pi(\theta)f(y;\theta)q(\tilde{\theta}|\theta)} \cdot \frac{f(x;\theta)}{f(x;\tilde{\theta})}\right\}, \quad (21)$$

where the auxiliary variable x is simulated from the likelihood function $f(x;\tilde{\theta})/D(\tilde{\theta})$ with the exact sampling (Propp and Wilson, 1996). In this acceptance probability, 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 x is simulated by m iterations of the usual M-H algorithm starting from y, the acceptance probability of the exchange algorithm 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 to mention is that, in this paper we have provided a technical modification on the standard double M-H algorithm to make it simplify the simulation and better fit into our application. When using the standard double M-H algorithm to update θ from $P(\theta|\{Y_g^*\}, \{W_g^*\})$, it requires to simulate auxiliary networks $\{\widetilde{W}_g\}$ and outcomes $\{\widetilde{Y}_g\}$. However, $\{\widetilde{Y}_g\}$ would be redundant as they can be fully replaced by a function of $\{\widetilde{W}_g\}$. Therefore, we modify the standard double M-H acceptance probability in Eq. (21) to

$$\alpha(\tilde{\theta}|\theta, x) = \min\left\{1, \frac{\pi(\tilde{\theta})P(y|\tilde{\theta})q(\theta|\tilde{\theta})}{\pi(\theta)P(y|\theta)q(\tilde{\theta}|\theta)} \cdot \frac{P^*(x|\theta)}{P^*(x|\tilde{\theta})}\right\} = \min\left\{1, \frac{\pi(\tilde{\theta})f(y;\tilde{\theta})q(\theta|\tilde{\theta})}{\pi(\theta)f(y;\theta)q(\tilde{\theta}|\theta)} \cdot \frac{f^*(x;\theta)}{f^*(x;\tilde{\theta})}\right\}, \quad (22)$$

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.

where the density function $P^*(x|\theta) = f^*(x;\theta)/D(\theta)$ can be different from $P(x|\theta)$ but shares the same normalizing term $D(\theta)$.[18] This modification allows us to construct the double M-H algorithm acceptance probability by only simulating auxiliary networks. The validity of the Markov chain based on the acceptance probability in Eq. (22), i.e., the reversibility condition of the transition density $p(\tilde{\theta}|\theta) = \alpha(\tilde{\theta}|\theta)q(\tilde{\theta}|\theta)$ is provided in the Appendix C-1. Moreover, to improve the performance of the double M-H algorithm when the vector of parameters is in high-dimension, we use the adaptive algorithm by Robert and Rosenthal (2009) for proposing candidate draws. The detail of the adaptive algorithm is provided in the Appendix C-2.

3.2 Posterior distributions of parameters and the MCMC

Here we show the posterior distributions of parameters based on the model which considers both continuous and Tobit-type activity outcomes. For dealing with Tobit-type outcome variables under the Bayesian approach, it is natural to include the sampling of latent variables $\{\ddot{Y}^*_{tg1}\}$ during the MCMC procedure along with other unobservables as an augmentation (Albert and Chib, 1993). By Bayes' theorem, the joint posterior distribution of the parameters and unobservables in the model is

$$
\begin{aligned}
&P\left(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\}, \{\ddot{Y}^*_{tg1}\} | \{Y^*_{cg}\}, \{Y^*_{tg}\}, \{W^*_g\}\right) \\
&\propto \pi(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\}) \prod_{g=1}^{G} \left\{ \left(\prod_{i=1}^{m_{g1}} I(y^*_{i,tg} = 0) \cdot I(\ddot{y}^*_{i,tg} \le 0) \right) \cdot P\left(Y^*_{tg}, Y^*_{cg}, W^*_g, \ddot{Y}^*_{tg1} | \theta_{ct}, \alpha_{cg}, \alpha_{tg}\right) \right\},
\end{aligned}
\tag{23}
$$

where $\pi(\cdot)$ represents the density function of the prior distribution and exogenous variables $\{X_g\}$ and $\{C_g\}$ are suppressed from the above expression for simplicity. We assume independence between prior distributions, i.e., $\pi(\theta_{ct}, \{\alpha_{cg}\}, \{\alpha_{tg}\}) = \pi_1(\theta_{ct})\pi_2(\{\alpha_{cg}\})\pi_3(\{\alpha_{tg}\})$. It is not easy to directly simulate draws from the joint posterior density in Eq. (23). But one can use the Gibbs sampling algorithm and work on the marginal posterior densities of parameters. By properly blocking parameters in θ_{ct} into subgroups, we define prior distributions for parameters in the model as

[18] More explicitly, $P(y|\theta)$ is the joint density function of $\{Y^*_g\}$ and $\{W^*_g\}$ and $P^*(x;\theta)$ is simply the density function of $\{\widetilde{W}_g\}$.

follows:

$$\phi = (\gamma', \eta', \delta_c, \delta_t) \sim \mathscr{T}\mathscr{N}_{2\bar{s}+\bar{q}+\bar{h}}(\phi_0, \Phi_0), \tag{24}$$

$$\lambda_c, \lambda_t \sim U[-1/\tau_G, 1/\tau_G], \tag{25}$$

$$\beta_c, \beta_t \sim \mathscr{N}_{2k}(\beta_0, B_0), \tag{26}$$

$$\sigma = (\sigma^2_{\epsilon_c}, \sigma^2_{\epsilon_t}, \sigma_{\epsilon_{ct}}) \sim \mathscr{T}\mathscr{N}_3(\sigma_0, \Sigma_0), \tag{27}$$

$$\alpha_{cg}, \alpha_{tg} \sim \mathscr{N}(\alpha_0, A_0), \quad g = 1, \cdots, G, \tag{28}$$

where $\mathscr{T}\mathscr{N}_s$ represents a truncated multivariate normal distribution of dimension s and $\mathscr{I}\mathscr{G}(a, b)$ represents a inverse-gamma distribution with a shape parameter a and a scale parameter b. 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 $V_g(W_g)$. The prior distribution of ϕ is the truncated normal which is defined on a convex area O where δ_c and δ_t are nonnegative. For λ_c and λ_t, they are independent and we employ a uniform prior for each suggested in Smith and LeSage (2002). We restrict the valid value of λ_c between $-1/\tau_G$ to $1/\tau_G$, where $\tau_G = \max\{\tau_1^*, \cdots, \tau_G^*\}$ and $\tau_g^* = \min\{\max_{1\le i\le m_g}\sum_{j=1}^{m_g}|w_{ij,g}|, \max_{1\le j\le m_g}\sum_{i=1}^{m_g}|w_{ij,g}|\}$[19]. $\sigma^2_{\epsilon_c}$, $\sigma^2_{\epsilon_t}$ and $\sigma_{\epsilon_{ct}}$ are put into a group called σ. We specify a truncated distribution for σ to an area T where $\sigma^2_{\epsilon_c}$, $\sigma^2_{\epsilon_t}$ and $\sigma_{\epsilon_{ct}}$ together form a proper correlation matrix. For group effects, α_{cg} and α_{tg}, as they are treated as fixed effects, hyperparameters in its prior distribution will not be updated. Applying the Gibbs sampling, 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) $P\left(\ddot{Y}^*_{tg1}\middle|\,\theta_{ct}, \alpha_{cg}, \alpha_{tg}, Y^*_{cg}, Y^*_{tg}, W^*_g\right)$, $g = 1, \cdots, G$.

By the Bayes' theorem,

$$P\left(\ddot{Y}^*_{tg1}\middle|\,\theta_{ct}, \alpha_{cg}, \alpha_{tg}, Y^*_{cg}, Y^*_{tg}, W^*_g\right) \propto \left(\prod_{i=1}^{m_{g1}} I(y^*_{i,g}=0) I(\ddot{y}^*_{i,g}\le 0)\right) f(\ddot{Y}^*_{tg1}, Y^*_{cg}, Y^*_{tg}, W^*_g|\theta_{ct}, \alpha_{cg}, \alpha_{tg}), \quad g = 1, \cdots, G. \tag{29}$$

(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} without ϕ. By

[19] 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.

Bayes' theorem, we have

$$\begin{aligned}&P(\phi|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}-\phi,\{\alpha_{cg}\},\{\alpha_{tg}\})\\&\qquad\propto \mathscr{T}\mathscr{N}_{2\bar{s}+\bar{q}+8}(\phi;\phi_0,\Phi_0)\cdot\prod_{g=1}^{G}P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\quad \phi\in O.\end{aligned}\tag{30}$$

(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

$$\begin{aligned}&P(\lambda_c|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}\backslash\lambda_c,\{\alpha_{cg}\},\{\alpha_{tg}\})\\&\qquad\propto\prod_{g=1}^{G}P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\end{aligned}\tag{31}$$

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

$$\begin{aligned}&P(\lambda_t|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}\backslash\lambda_t,\{\alpha_{cg}\},\{\alpha_{tg}\})\\&\qquad\propto\prod_{g=1}^{G}P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\end{aligned}\tag{32}$$

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

$$\begin{aligned}&P(\beta_c|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}\backslash\beta_c,\{\alpha_{cg}\},\{\alpha_{tg}\})\\&\qquad\propto\mathscr{N}(\beta_c;\beta_0,B_0)\cdot\prod_{g=1}^{G}P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}).\end{aligned}\tag{33}$$

(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

$$\begin{aligned}&P(\beta_t|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}\backslash\beta_t,\{\alpha_{cg}\},\{\alpha_{tg}\})\\&\qquad\propto\mathscr{N}(\beta_t;\beta_0;B_0)\cdot\prod_{g=1}^{G}P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}).\end{aligned}\tag{34}$$

(vii) $P(\sigma|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}\backslash\sigma,\{\alpha_{cg}\},\{\alpha_{tg}\})$.

After removing irrelevant arguments from the conditional posterior distribution of σ, by applying Bayes' theorem, we have

$$P(\sigma|\{\ddot{Y}^*_{tg1}\},\{Y^*_{cg}\},\{Y^*_{tg}\},\{W^*_g\},\theta_{ct}\backslash\sigma,\{\alpha_{cg}\},\{\alpha_{tg}\}) \propto \mathscr{T}\mathscr{N}_3(\sigma;\sigma_0,\Sigma_0)\cdot\prod_{g=1}^{G}P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}). \tag{35}$$

(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

$$P(\alpha_{cg}|\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g,\theta_{ct},\alpha_{tg}) \propto \mathscr{N}(\alpha_{cg};\alpha_0,A_0)\cdot P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\quad g=1,\cdots,G. \tag{36}$$

(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

$$P(\alpha_{tg}|\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g,\theta_{ct},\alpha_{cg}) \propto \mathscr{N}(\alpha_{tg};\alpha_0,A_0)\cdot P(\ddot{Y}^*_{tg1},Y^*_{cg},Y^*_{tg},W^*_g|\theta_{ct},\alpha_{tg},\alpha_{cg}),\quad g=1,\cdots,G. \tag{37}$$

All of the conditional posterior distributions are not available in a closed form and hence, we use the double M-H algorithm to draw from those conditional distributions. It has been shown in Tierney (1994), Chib and Greenberg (1996) that the combination of Markov chains (Metropolis-within-Gibbs) is still a Markov chain with the invariant distribution equal to the correct objective distribution. The procedure of the MCMC sampling can start with arbitrary initial values for $\{\alpha^{(0)}_{cg}\}$, $\{\alpha^{(0)}_{tg}\}$, and $\theta^{(0)}_{ct}$, and then sampling sequentially from the above set of conditional posterior distributions. The detailed implementation steps of the MCMC sampling based on the model with both the continuous and Tobit-type activity outcomes are provided in the Appendix C-3. To examine the computational aspects of our MCMC algorithms, especially with the use of double M-H algorithm, we have conducted a simple simulation study and its results show that the double M-H algorithm can handle the problem of normalizing terms without problem. The details of this simulation study are left in the Appendix D.

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 7^{th} to 12^{th} grades from 132 schools.[20] Four waves of surveys were conducted between 1994 to 2008. In the wave I in-school survey, the total of 90,182 students were interviewed. Each respondent answered questions about their demographic backgrounds, academic performances, health related behaviors, etc., and most uniquely, they were asked to nominate up to five male and five female friends which provide information of their friendship networks. In the following waves of in-home surveys, more information about students' families and living neighborhoods are available for a subset of the total sample. To accommodate most of students' nominated friends into our studying framework, the sample used in this study is constructed from the wave I in-school survey. We consider two activity outcomes 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] Another is how frequently a student smokes in an usual week, which is represented by a Tobit-type variable.

In the context of social interactions, both students' academic performances and smoking behaviors 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 difficulties of identification from endogenous selections into groups, correlated effects from group-level unobservables (Moffitt, 2001), and separating the endogenous interaction effect from contextual effects in a linear model (the reflection problem by Manski, 1993). Sizes of peer effects reported in the literature usually differ from each other due to uses of different data sets or different strategies to deal with the identification problem. However, researchers generally provide evidence of the existence for peer effects. Hsieh and Lee (2012) further considers the identification problem on peer effects

[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.

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

[22]For studying peer effects on students' academic performance, see Hoxby (2000), Hanushek *et al.* (2003), Sacerdote (2001), Zimmerman (2003), etc., on using the linear-in-means model and Calvó-Armengol *et al.* (2009), Lin (2010), Boucher *et al.* (2010), Liu *et al.* (2011) on using the SAR model. For students' smoking behaviors, see evidences of peer effects on Powell *et al.* (2005), Gaviria and Raphael (2001), Clark and Loheac (2007), Lundborg (2006) and Fletcher (2010).

caused by endogenous friendship selections within a group when using the SAR model for studying network interactions. They show that the endogenous effect obtained from the SAR model without controlling the endogeneity of the adjacency matrix will be upward biased and this endogeneity problem can be resolved by a modeling approach with unobservables in both the outcome and network formation processes. In the present study, we confirm their findings that the endogenous effect would be smaller after controlling endogenous formation of friendship networks. Moreover, we will show that the benefit from helping one another in 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 empirical applications, we consider the following specification of the network structure effects in the link-associated utility of Eq. (8),[23]

$$
\begin{aligned}
&\varpi_{i,g}(w_{i\cdot,g}, W_{-i\cdot,g})\eta \\
&= \underbrace{\eta_1 \sum_{j=1}^{m_g} w_{ij,g} w_{ji,g}}_{\text{Reciprocality Effect}} + \underbrace{\eta_2 \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k \neq j}^{m_g} w_{ik,g} \right) + \eta_3 \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k \neq j}^{m_g} w_{ik,g} \right)^2}_{\text{Sender's Expansiveness Effect}} + \underbrace{\eta_4 \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k \neq i}^{m_g} w_{kj,g} \right)}_{\text{Receiver's Popularity Effect}} \\
&\quad \underbrace{+ \eta_{51} \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k}^{m_g} w_{ik,g} w_{kj,g} \right) + \eta_{52} \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k}^{m_g} w_{ki,g} w_{kj,g} \right) + \eta_{53} \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k}^{m_g} w_{ik,g} w_{jk,g} \right)}_{\text{Transitive Triads Effect}} \\
&\quad \underbrace{+ \eta_6 \sum_{j=1}^{m_g} w_{ij,g} \left(\sum_{k}^{m_g} w_{jk,g} w_{ki,g} \right)}_{\text{Three Cycles Effect}}. \qquad (38)
\end{aligned}
$$

In Eq. (38), the reciprocality effect reflects the utility from reciprocal friendships. Since each link decision is made by one individual without mutual consents from the other, the possibility of reciprocality may be a factor in an individual's link decision. The sender's expansiveness effect in Eq. (38) reflects the utility from being an outgoing person who actively nominate friends. The statistics involved are the sender's outdegrees and their squares. We expect the coefficient η_3 would be negative to reflect the reality that individuals might not make friends with everybody due to limited resources, e.g., time, energy, etc. The receiver's indegree is used to measure the receiver's

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

popularity effect in Eq. (38), which reflects the utility from making a friend 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. We allow different coefficients to capture their possibly distinctive effects on the utility. From Kovářík and van der Leij (2012), transitive triads effects may be linked to 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. (38), the term $\sum_{i=1}^{m_g} \varpi_{i,g}(w_{i.,g}, W_{-i.,g})\eta$ in the aggregated link utility of Eq.(11) can be written as

$$\begin{aligned}
&\sum_{i=1}^{m_g} \varpi_{i,g}(w_{i.,g}, W_{-i.,g})\eta \\
&= \eta_1 \mathrm{tr}(W_g^2) + \eta_2(l_g' W_g' W_g l_g - l_g' W_g l_g) \\
&\quad + \eta_3(l_g' W_g' \mathrm{Diag}(W_g l_g) W_g l_g - 2 l_g' W_g' W_g l_g + l_g' W_g l_g) \\
&\quad + \eta_4(l_g' W_g W_g' l_g - l_g' W_g l_g) + (\eta_{51} + \eta_{52} + \eta_{53}) \mathrm{tr}(W_g^2 W_g') + \eta_6 \mathrm{tr}(W_g^3),
\end{aligned} \tag{39}$$

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. (39). 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-grades and friendships are considered inside groups.[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 of the size between 10 to 50 (10 to 60 for the smoking case). After removing missing observations on outcome variables in each group, there are total 1,177 (1,476 for the smoking case) respondents from 47 networks (44 networks for the smoking case) left for analysis.[25] Those 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),

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

[25] For the smoking sample, we remove groups in which there are fewer than three students who have ever smoked. The number of missing observation is equal to 113 (9.6%) for the GPA sample and 34 (2%) for the smoking sample.

average outdegree equal to 2.564 (2.866 for the smoking case), and average clustering coefficient[26] equal to 0.327 (0.332 for the smoking case). In the network 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 a same age, same sex, or same race are used to capture dyadic-specific effects. For the activity outcome equation, the continuous variable, GPA, is calculated by the average of respondent's reported grades from several subjects, including language, social science, mathematics, and science, and each of which has a value between 1 to 4. The Tobit-type variable, smoking, is obtained from the survey question, "*During the past twelve months, how often did you smoke cigarettes*?" and we transform the responses to a weekly base. The choice of independent variables used in the outcome equation follows from Lin (2010), Lee *et al.* (2007, 2010) and Hsieh and Lee (2012). A complete list of variables is in Table 1. In Figure 1 and 2, we plot two networks – one is from the GPA sample and another 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 of each other. Our estimation results shown in the next subsection will provide evidences for the economic incentive stemming from benefits of helping one another in academic learning, but not from joy of smoking together, on friendship decisions. Moreover, our results show that interaction effects on improving GPA or increasing 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: $\phi_0 = 0$; $\Phi_0 = 10I_{2\bar{s}+\bar{q}+\bar{h}}$; $\beta_0 = 0$; $B_0 = 10I_{2k}$; $\kappa_0 = 0.1$; $\varrho_0 = 2$; $\alpha_0 = 0$; $A_0 = 400$. These specified values of hyperparameters are designed to allow relative flat prior densities over the range of the parameter spaces.

4.3 Estimation results

We first estimate the model with a single economic incentive effect from GPA and report results in Table 2. Results in the first two columns are obtained separately from the full model and the outcome equation alone with networks assumed exogenous. For examining possible consequences of dropping the 10% missing observations on GPA in the sample, results in the third column are obtained from the full model with a Bayesian data augmentation approach to recover the missing

[26]The clustering coefficient is calculated as the total fraction of transitive triples in the network, i.e.,

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

observations.[27] The values shown for each parameter are the mean and the standard deviation (in parenthesis) from posterior draws. From the network model in the first column, we observe that whether being older than the group average or not does not have a significant effect on sending or receiving friendship nominations. However, three dyadic-specific effects we consider are all positive and significant, where the effect of the same race is strongest, followed by the effect of the same sex. 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 percent 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., time, energy, money, etc., might constraint students from making too many friends. The receiver's popularity effect is negative, which suggests that students between 9^{th} to 12^{th} grades in our sample do not prefer to make a friend with someone who is popular, i.e., having received many friendship nominations. The positive and strong transitive triads effect shows that transitive relationships are valued by students. When accompanying with the negative three cycles effect, as discussed in Snijders *et al.* (2010), it reveals a certain degree of local hierarchy among students. The last parameter of the network model, the economic incentive effect, is found to be positive and significant. Therefore, for high school students in our sample, potential benefits from helping one another in academic learning is a factor which determines their friendship decisions.

From the outcome equation in the first column, the estimated endogenous effect is equal to 0.021 and significant. It 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 estimated own and contextual effects, we observe that for students who are older, male, or whose moms have education less 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

[27] At the time we drop observations with missing outcome variables, we also drop the potential links connected to these observations. Since we study network formation and network interactions, 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.

[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.

Asian.

When estimating the outcome equation alone by treating the adjacency matrix as exogenously given, results in the second column show that, the estimate of the endogenous effect as well as its standard deviation are nearly double than those obtained from the full model. Meanwhile, estimated own and contextual effects are also different from those of the full model and have larger standard deviations. Those differences between results in the first and second columns show the problem of friendship selection bias in outcomes and this problem can be remedied by our proposed system of models. In the third column, with missing observations on GPA augmented during the estimation procedure, we do not observe significant changes on estimates of parameters in the network model and the endogenous effect in the outcome equation, when comparing to the first column. Although there are few changes on estimates of own and contextual effects, these are insignificant estimates. The advantage of using this data augmentation approach can be seem from smaller standard deviations of posterior draws for parameters in both the network model and the outcome equation.

Next we turn to the model with a single economic incentive effect from smoking. Estimation are done separately for the full model and the outcome equation alone with networks assumed exogenous and results are reported in the first and the second columns of Table 3. As there are only 2% of missing observations on the variable of smoking, we do not think it is necessary to recover them. From the network model in the first column, we again find that being older than the group average or not 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 found similar to those in the case of GPA. An exceptional finding is that the economic incentive effect from smoking is small and insignificant. Hence we can know that, for students in our sample, they do not consider joy of smoking together as a factor for their friendship decisions. From the outcome equation of the full model, the estimated endogenous interaction effect 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. 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 having friends who are male or Asian. As shown in the second column, the estimated endogenous effect from the outcome equation alone is equal to 0.088, which is not significantly different from that obtained from the full model. Although several estimated own and contextual effects are different from what are obtained from the full model, those are insignificant effects. These results suggest that outcomes of network interactions for smoking are not subject to friendship selection biases.

Lastly, for robustness checks, we estimate the model with economic incentive effects from both GPA and smoking and report results in Table 4.[30] In the network model, a strong and significant economic incentive effect is still found from GPA, while it is not found from smoking. By comparing results between the full model and the outcome equations 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 outcome equation alone to 0.025 when estimating the full model. For smoking, due to a small and insignificant incentive effect, we do not find clear evidences of friendship selection biases in effects of network interactions. After all, the covariance of disturbances in outcome equations between GPA and smoking is found to be -0.653 and significant.

5 Conclusion

An important reason why researchers study network structures is to analyze network impacts 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, the impacts on outcomes are sort of mechanical and one do not need to worry any feedback effects from outcomes. However, for studying impacts of friendship networks on outcomes, both the network structure and the strategic interactions between networks and outcomes should be considered. This extra consideration should be reflected on a dynamic or static equilibrium model. In this paper, we propose a static equilibrium model which takes into account those features. The modeling approach used in this paper assumes that students respond to economic incentives stemming from interactions with friends on certain behaviors when making their friendship decisions. The empirical results show that, American high school students regard achieving better academic outcomes through friendship interactions as a significant incentive for forming friendships, while the same incentive effect is not found from smoking behaviors. A valuable byproduct of our approach, which contributes to the literature of social interactions, is to correct the possible friendship selection biases in interaction effects.

Some issues which are not emphasized in this paper remain important for future extensions.

[30]The sample used for estimating this model is based on the original GPA sample where we remove missing observations on smoking within groups and remove groups in which there are fewer than three students who have ever smoked. The resulting sample has 1,062 observations and 37 groups.

The first is the problem of possible multiple equilibria in the simultaneous 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 each individuals internalize the generated externality during the friendship formation process. Those assumptions are questionable for general friendship networks. 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 impacts 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 the network boundary is to examine results under various network boundaries as robustness checks. This has not been done in the present paper due to the capacity to handle 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) and imputation (Huisman, 2009) provide few possible solutions. For potential biases brought by missing network links in outcomes with network interactions, Chandrasekhar and Lewis (2012) and Liu (2012) have useful discussions. The third issue to consider is the dynamic evolution of networks and outcomes. The work of Snijders *et al.* (2010) is surely leading the direction of this research. Soon or later, panel network data would become more available for network researchers and different modeling and estimation approaches should be highly desirable.

[31] However, for data collection purpose, researchers tend to believe that if an individual is allowed to fill in 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.

APPENDIX A: The case of binary activity outcome variable

For binary outcomes variables, we assume the vector of latent outcomes follow the SAR model, i.e.,

$$\ddot{Y}_{bg}^* = \lambda_b W_g \ddot{Y}_{bg}^* + \mathbf{X}_g \beta_b + l_g \alpha_{bg} + \epsilon_{bg} \quad \epsilon_{bg} \sim \mathcal{N}_{m_g}(0, I_{m_g}), \tag{40}$$

and the observed outcomes are determined by

$$\begin{aligned} y_{i,bg}^* &= I\{\ddot{y}_{i,bg}^* \geq 0\} \\ &= I\left\{(I_{m_g} - \lambda_b W_g)_i^{-1}(\mathbf{X}_g \beta_b + l_g \alpha_{bg} + \epsilon_{bg}) \geq 0\right\} \\ &= I\left\{S_{bg,i}^{-1}(\mathbf{X}_g \beta_b + l_g \alpha_{bg} + \epsilon_{bg}) \geq 0\right\}, \end{aligned} \tag{41}$$

where $S_{bg,i}^{-1}$ is the ith row of S_{bg}^{-1}, and $I\{\cdot\}$ represents an indicator function. The normalization of the variance of ϵ_{bg} to one is a standard practice to deal with the arbitrary scaling problem in binary choice models. This specification is meaningful in the sense that the latent variables in $\ddot{Y}_{bg}^*$ can be treated as unobserved utilities, motivations or intensions from individuals. Interactions between individuals may be generated via inter-dependences of these latent variables and then their interactions are reflecting on the observed binary outcomes of Y_{bg}^*. To estimate the model with economic incentive effects from binary outcomes, we can follow a similar approach used for the Tobit-type outcomes by including the sampling of latent variables $\{\ddot{Y}_{bg}^*\}$ along with other unobservables in the MCMC procedure. Hence, the details of the estimation procedure will not be repeated again.

APPENDIX B: A contraction mapping algorithm for solving the unique solution from the simultaneous Tobit outcome equation

To find out the solution of Eq. (5), 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

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

where $Z_g = X_g\beta + l_g\alpha_g + \epsilon_g$, $w_{i.,g}$ is the ith element 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. 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.

APPENDIX C-1: The reversibility condition of the modified double M-H algorithm

To show 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. (22), we need to show the Markov chain based on the transition density $p(\tilde{\theta}|\theta) = \alpha(\tilde{\theta}|\theta)q(\tilde{\theta}|\theta)$ is reversible, and therefore, $P(\theta|y)$ is an invariant distribution. One should note that the acceptance probability of Eq. (22) is conditioning on drawing x from the density $P^*(x|\tilde{\theta})$. To check the reversibility condition of this modified double M-H algorithm, we need the unconditional acceptance probability, that is

$$\begin{aligned}
\alpha(\tilde{\theta}|\theta) &= \int \alpha(\tilde{\theta}|\theta, x) P^*(x|\tilde{\theta}) dx \\
&= \int \min\left\{ \frac{\pi(\tilde{\theta})P(y|\tilde{\theta})q(\theta|\tilde{\theta})}{\pi(\theta)P(y|\theta)q(\tilde{\theta}|\theta)} \cdot \frac{P^*(x|\theta)}{P^*(x|\tilde{\theta})}, 1 \right\} P^*(x|\tilde{\theta}) dx \\
&= \int \min\left\{ \frac{P(\tilde{\theta}|y)q(\theta|\tilde{\theta})}{P(\theta|y)q(\tilde{\theta}|\theta)} P^*(x|\theta), P^*(x|\tilde{\theta}) \right\} dx
\end{aligned}$$

because x has the density $P^*(x|\tilde{\theta})$. We want to check if the following equality holds,

$$P(\theta|y)\alpha(\tilde{\theta}|\theta)q(\tilde{\theta}|\theta) = P(\tilde{\theta}|y)\alpha(\theta|\tilde{\theta})q(\theta|\tilde{\theta}).$$

From the left hand side,

$$P(\theta|y)\alpha(\tilde{\theta}|\theta)q(\tilde{\theta}|\theta) = \int \min\left\{ q(\theta|\tilde{\theta})P(\tilde{\theta}|y)P^*(x|\theta), q(\tilde{\theta}|\theta)P(\theta|y)P^*(x|\tilde{\theta}) \right\} dx.$$

From the right hand side,

$$P(\tilde{\theta}|y)\alpha(\theta|\tilde{\theta})q(\theta|\tilde{\theta}) = \int \min\left\{q(\tilde{\theta}|\theta)P(\theta|y)P^*(x|\tilde{\theta}), q(\theta|\tilde{\theta})P(\tilde{\theta}|y)P^*(x|\theta)\right\} dx.$$

Those two are equal. Hence, the reversibility condition is satisfied.

APPENDIX C-2: The Adaptive Metropolis (AM) algorithm

One difficulty faced by the M-H algorithm is that, when the dimension of the draw is large, the convergence of the Markov chain might be slow when proposals were poor. To improve the efficiency of the M-H algorithm, we consider the Adaptive Metropolis (AM) algorithm introduced by Haario *et al.*(2001). The idea of the AM algorithm is to provide effective proposals for the M-H algorithm. The standard M-H algorithm uses the random walk proposal which has mean equal to the previous draw and its covariance equal to an identity matrix. It is often the case that the target distribution will not be well characterized with an identity matrix. The AM algorithm suggests to use the covariance of historical MCMC draws to form the covariance matrix of the proposal distribution. Suppose we use the M-H algorithm to update the parameter vector θ, say dimension K. At the iteration t with historical MCMC draws $(\theta^{(0)}, \theta^{(1)}, \cdots, \theta^{(t-1)})$, the AM proposal suggested by Robert and Rosenthal (2009)[32] is

$$\begin{aligned} & q_t(\theta|\theta^{(0)}, \cdots, \theta^{(t-1)}) \\ &= \begin{cases} \mathscr{N}_K\left(\theta^{(t-1)}, I_K \frac{0.1^2}{K}\right) & t \leq 2K \\ (1-\rho)\mathscr{N}_K\left(\theta^{(t-1)}, \mathrm{Cov}(\theta^{(0)}, \cdots, \theta^{(t-1)})\frac{2.38^2}{K}\right) + \rho\mathscr{N}_K\left(\theta^{(t-1)}, I_K \frac{0.1^2}{K}\right) & t > 2K, \end{cases} \end{aligned} \tag{42}$$

where the scaling factor $2.38^2/K$ is suggested in Gelman *et al.* (1996), which optimizes the mixing properties of the Metropolis search in the case of Gaussian proposals. This AM proposal is a mixture of two normal distributions with a ratio parameter ρ. This works as an extra safe scheme to prevent us to generate the proposal from a problematic value of $\mathrm{Cov}(\theta^{(0)}, \theta^{(1)}, \cdots, \theta^{(t-1)})$, e.g., singular. In this paper we set ρ as 0.05 following Robert and Rosenthal (2009).

APPENDIX C-3: The algorithm of the MCMC sampling for the model with both continuous and simultaneous Tobit outcome variables

[32]They have proved the ergodicity of the resulting MCMC using this AM proposal in their paper.

At the r^{th} run of the iteration, perform the following steps:

Step I. For $g = 1, \cdots, G$, simulate $\ddot{Y}_{tg1}^{*(r)}$ from $P(\ddot{Y}_{tg1}^{*}|\theta_{ct}^{(r-1)}, \alpha_{tg}^{(r-1)}, \alpha_{cg}^{(r-1)}, Y_{tg}^{*}, Y_{cg}^{*}, W_g^{*})$ by the double M-H algorithm.

(a) propose $\widetilde{\ddot{Y}_{tg1}^{*}}$ from a AM proposal $q\left(\ddot{Y}_{tg1}^{*} \middle| \ddot{Y}_{tg1}^{*(0)}, \cdots, \ddot{Y}_{tg1}^{*(r-1)}\right)$.

(b) Calculate
$\widetilde{\epsilon}_{tg1} = \widetilde{\ddot{Y}_{tg1}^{*}} - \left(\lambda_t^{(r-1)} W_{12,g}^{*} Y_{tg2}^{*} + X_{1g}\beta_{1t}^{(r-1)} + (W_{11,g}^{*}X_{1g} + W_{12,g}^{*}X_{2g})\beta_{2t}^{(r-1)} + l_{g1}\alpha_{tg}^{(r-1)}\right)$
and $\epsilon_{tg2}^{(r-1)} = \left(I_{m_g - m_{g1}} - \lambda_t^{(r-1)} W_{22,g}^{*}\right) Y_{tg2}^{*} - X_{2g}\beta_{1t}^{(r-1)} - (W_{21,g}^{*}X_{1g} + W_{22,g}^{*}X_{2g})\beta_{2t}^{(r-1)} - l_{g2}\alpha_{tg}^{(r-1)}$. Denote $\widetilde{\epsilon}_{tg} = \left(\widetilde{\epsilon}'_{tg1}, \epsilon_{tg2}^{(r-1)'}\right)'$. To make a distinction, denote $\epsilon_{tg}^{(r-1)} = \left(\epsilon_{tg1}^{(r-1)'}, \epsilon_{tg2}^{(r-1)'}\right)'$ with
$\epsilon_{tg1}^{(r-1)} = \ddot{Y}_{tg1}^{*(r-1)} - \left(\lambda_t^{(r-1)} W_{12,g}^{*} Y_{tg2}^{*} + X_{1g}\beta_{1t}^{(r-1)} + (W_{11,g}^{*}X_{1g} + W_{12,g}^{*}X_{2g})\beta_{2t}^{(r-1)} + l_{g1}\alpha_{tg}^{(r-1)}\right)$.
Also calculate $\epsilon_{cg}^{(r-1)} = \left(I_{m_g} - \lambda_c^{(r-1)} W_g^{*}\right) Y_{cg}^{*} - X_g\beta_{1c}^{(r-1)} - W_g^{*}X_g\beta_{2c}^{(r-1)} - l_g\alpha_{cg}^{(r-1)}$.

(c) Given $\widetilde{\epsilon}_{tg}$ and $\epsilon_{cg}^{(r-1)}$ from (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on[33]

$$P(W_g|\widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \theta_{ct}^{(r-1)}, \alpha_{tg}^{(r-1)}, \alpha_{cg}^{(r-1)}) = \frac{\exp\left(V_g(W_g, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \theta_{ct}^{(r-1)}, \alpha_{tg}^{(r-1)}, \alpha_{cg}^{(r-1)})\right)}{\sum_W \exp\left(V_g(W, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \theta_{ct}^{(r-1)}, \alpha_{tg}^{(r-1)}, \alpha_{cg}^{(r-1)})\right)}$$

starting from W_g^{*}, i.e., first set the initial auxiliary network W_g equal to W_g^{*}. For each entry of W_g, $w_{ij,g}$, $i \neq j$, in turn, we propose $\widetilde{w}_{ij,g} = 1 - w_{ij,g}$. With the acceptance probability

$$\alpha(\widetilde{w}_{ij,g}|w_{ij,g}) = \min\left\{\frac{\exp\left(V_g(\tilde{w}_{ij,g}, W_{-ij,g}, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \theta_{ct}^{(r-1)}, \alpha_{tg}^{(r-1)}, \alpha_{cg}^{(r-1)})\right)}{\exp\left(V_g(w_{ij,g}, W_{-ij,g}, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \theta_{ct}^{(r-1)}, \alpha_{tg}^{(r-1)}, \alpha_{cg}^{(r-1)})\right)}, 1\right\},$$

updating $w_{ij,g}$ to $\widetilde{w}_{ij,g}$.

[33] In practice, we set $m = 2m_g(m_g - 1)$ where m_g is the size of the network g.

(d) With the acceptance probability equal to

$$
\begin{aligned}
&\alpha\left(\widetilde{\ddot{Y}^*_{tg1}}\,\middle|\ddot{Y}^{*(r-1)}_{tg1}\right)\\
&=\min\left\{\frac{P\left(\widetilde{\ddot{Y}^*_{tg1}},Y^*_{tg},Y^*_{cg},W^*_g\,\middle|\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)}{P\left(\ddot{Y}^{*(r-1)}_{tg1},Y^*_{tg},Y^*_{cg},W^*_g\,\middle|\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)}\cdot\right.\\
&\quad\left.\frac{P\left(\widetilde{W}_g\,\middle|\epsilon^{(r-1)}_{tg},\epsilon^{(r-1)}_{cg},\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)}{P\left(\widetilde{W}_g\,\middle|\widetilde{\epsilon}_{tg},\epsilon^{(r-1)}_{cg},\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)}\cdot\frac{I\left(\widetilde{\ddot{Y}^*_{tg1}}<0\right)}{I\left(\ddot{Y}^{*(r-1)}_{tg1}<0\right)},1\right\}\\
&=\min\left\{\frac{f\left(\widetilde{\epsilon}_{tg}-\sigma^{(r-1)}_{\epsilon_{tc}}(\sigma^{2(r-1)}_{\epsilon_c})^{-1}\epsilon^{(r-1)}_{cg}\right)\exp\left(V_g\left(W^*_g,\widetilde{\epsilon}_{tg},\epsilon^{(r-1)}_{cg},\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)\right)}{f\left(\epsilon^{(r-1)}_{tg}-\sigma^{(r-1)}_{\epsilon_{tc}}(\sigma^{2(r-1)}_{\epsilon_c})^{-1}\epsilon^{(r-1)}_{cg}\right)\exp\left(V_g\left(W^*_g,\epsilon^{(r-1)}_{tg},\epsilon^{(r-1)}_{cg},\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)\right)}\cdot\right.\\
&\quad\left.\frac{\exp\left(V_g\left(\widetilde{W}_g,\epsilon^{(r-1)}_{tg},\epsilon^{(r-1)}_{cg},\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)\right)}{\exp\left(V_g\left(\widetilde{W}_g,\widetilde{\epsilon}_{tg},\epsilon^{(r-1)}_{cg},\theta^{(r-1)}_{ct},\alpha^{(r-1)}_{tg},\alpha^{(r-1)}_{cg}\right)\right)}\cdot\frac{I\left(\widetilde{\ddot{Y}^*_{tg1}}<0\right)}{I\left(\ddot{Y}^{*(t-1)}_{tg1}<0\right)},1\right\}
\end{aligned}
$$

set $\ddot{Y}^{*(r)}_{tg1}$ with $\widetilde{\ddot{Y}^*_{tg1}}$. Otherwise, set $\ddot{Y}^{*(r)}_{tg1}=\ddot{Y}^{*(r-1)}_{tg1}$.

Step II. Simulate $\phi^{(r)}$ from $P\left(\phi\,\middle|\{\ddot{Y}^{*(r)}_{tg1}\},\{Y^*_{tg}\},\{Y^*_{cg}\},\{W^*_g\},\Upsilon^{(r-1)}\right)$ by the double M-H algorithm, where $\Upsilon^{(r-1)}$ denotes the rest of paramters evaluted at the $(r-1)^{th}$ iteration.

(a) propose $\widetilde{\phi}$ from a AM proposal $q(\phi|\phi^{(0)},\cdots,\phi^{(r-1)})$

(b) For $g=1,\cdots,G$, calculate $\epsilon^{(r-1)}_{tg1}=\ddot{Y}^{*(r)}_{tg1}-\left(\lambda^{(r-1)}_t W^*_{12,g}Y^*_{tg2}+X_{1g}\beta^{(r-1)}_{1t}+(W^*_{11,g}X_{1g}+W^*_{12,g}X_{2g})\beta^{(r-1)}_{2t}+l_{g1}\alpha^{(r-1)}_{tg}\right)$ and $\epsilon^{(r-1)}_{tg2}=\left(I_{m_g-m_{g1}}-\lambda^{(r-1)}_t W^*_{22,g}\right)Y^*_{tg2}-X_{2g}\beta^{(r-1)}_{1t}-(W^*_{21,g}X_{1g}+W^*_{22,g}X_{2g})\beta^{(r-1)}_{2t}-l_{g2}\alpha^{(r-1)}_{tg}$. Denote $\epsilon^{(r-1)}_{tg}=\left(\epsilon^{(r-1)'}_{tg1},\epsilon^{(r-1)'}_{t2g}\right)'$. Also calculate $\epsilon^{(r-1)}_{cg}=\left(I_{m_g}-\lambda^{(r-1)}_c W^*_g\right)Y^*_{cg}-X_g\beta^{(r-1)}_{1c}-W^*_gX_g\beta^{(r-1)}_{2c}-l_g\alpha^{(r-1)}_{cg}$.

(c) For $g=1,\cdots,G$, given $\epsilon^{(r-1)}_{tg}$ and $\epsilon^{(r-1)}_{cg}$ from (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$P(W_g|\epsilon^{(r-1)}_{tg},\epsilon^{(r-1)}_{cg},\widetilde{\phi},\Upsilon^{(r-1)})=\frac{\exp\left(V_g\left(W_g,\epsilon^{(r-1)}_{tg},\epsilon^{(r-1)}_{cg},\widetilde{\phi},\Upsilon^{(r-1)}\right)\right)}{\sum_W\exp\left(V_g\left(W,\epsilon^{(r-1)}_{tg},\epsilon^{(r-1)}_{cg},\widetilde{\phi},\Upsilon^{(r-1)}\right)\right)}$$

starting from W^*_g. See details in Step I. part (c).

(d) With the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\phi}|\phi^{(r-1)}) \\
&= \min\left\{\prod_{g=1}^{G}\left(\frac{P\left(W_g^*|\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\widetilde{\phi},\Upsilon^{(r-1)}\right)}{P\left(W_g^*|\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\phi^{(r-1)},\Upsilon^{(r-1)}\right)}\cdot\right.\right. \\
&\qquad \left.\left.\frac{P\left(\widetilde{W}_g|\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\phi^{(r-1)},\Upsilon^{(r-1)}\right)}{P\left(\widetilde{W}_g|\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\widetilde{\phi},\Upsilon^{(r-1)}\right)}\right)\frac{\mathcal{N}_{2\bar{s}+\bar{q}+\bar{h}}(\widetilde{\phi}|\phi_0,\Phi_0)}{\mathcal{N}_{2\bar{s}+\bar{q}+\bar{h}}(\phi^{(r-1)}|\phi_0,\Phi_0)}\cdot\frac{I(\widetilde{\phi}\in O)}{I(\phi^{(r-1)}\in O)},1\right\} \\
&= \min\left\{\prod_{g=1}^{G}\left(\frac{\exp\left(V_g\left(W_g^*,\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\widetilde{\phi},\Upsilon^{(r-1)}\right)\right)}{\exp\left(V_g\left(W_g^*,\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\phi^{(r-1)},\Upsilon^{(r-1)}\right)\right)}\cdot\right.\right. \\
&\qquad \left.\left.\frac{\exp\left(V_g\left(\widetilde{W}_g,\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\phi^{(r-1)},\Upsilon^{(r-1)}\right)\right)}{\exp\left(V_g\left(\widetilde{W}_g,\epsilon_{tg}^{(r-1)},\epsilon_{cg}^{(r-1)},\widetilde{\phi},\Upsilon^{(r-1)}\right)\right)}\right)\frac{\mathcal{N}_{2\bar{s}+\bar{q}+\bar{h}}(\widetilde{\phi}|\phi_0,\Phi_0)}{\mathcal{N}_{2\bar{s}+\bar{q}+\bar{h}}(\phi^{(r-1)}|\phi_0,\Phi_0)}\cdot\frac{I(\widetilde{\phi}\in O)}{I(\phi^{(r-1)}\in O)},1\right\},
\end{aligned}
$$

set $\phi^{(r)}$ with $\widetilde{\phi}$. Otherwise, set $\phi^{(r)}=\phi^{(r-1)}$.

Step III. Simulate $\lambda_c^{(r)}$ from $P(\lambda_c|\{\ddot{Y}_{tg1}^{*(r)}\},\{Y_{tg}^*\},\{Y_{cg}^*\},\{W_g^*\},\phi^{(r)},\Upsilon^{(r-1)})$ by the double M-H algorithm.

(a) propose $\widetilde{\lambda}_c$ from a random walk proposal density $q(\lambda_c|\lambda_c^{(r-1)})$

(b) For $g=1,\cdots,G$, calculate $\widetilde{\epsilon}_{cg}=\left(I_{m_g}-\widetilde{\lambda}_c W_g^*\right)Y_{cg}^*-X_g\beta_{1c}^{(r-1)}-W_g^*X_g\beta_{2c}^{(r-1)}-l_g\alpha_{cg}^{(r-1)}$. and $\epsilon_{cg}^{(r-1)}=\left(I_{m_g}-\lambda_c^{(r-1)}W_g^*\right)Y_{cg}^*-X_g\beta_{1c}^{(r-1)}-W_g^*X_g\beta_{2c}^{(r-1)}-l_g\alpha_{cg}^{(r-1)}$. Also calculate
$\epsilon_{tg1}^{(r-1)}=\ddot{Y}_{tg1}^{*(r)}-\left(\lambda_t^{(r)}W_{12,g}^*Y_{tg2}^*+X_{1g}\beta_{1t}^{(r-1)}+(W_{11,g}^*X_{1g}+W_{12,g}^*X_{2g})\beta_{2t}^{(r-1)}+l_{g1}\alpha_{tg}^{(r-1)}\right)$
and $\epsilon_{tg2}^{(r-1)}=(I_{m_g-m_{g1}}-\lambda_t^{(r)}W_{22,g}^*)Y_{tg2}^*-X_{2g}\beta_{1t}^{(r-1)}-(W_{21,g}^*X_{1g}+W_{22,g}^*X_{2g})\beta_{2t}^{(r-1)}-l_{g2}\alpha_{tg}^{(r-1)}$ Denote $\epsilon_{tg}^{(r-1)}=\left(\epsilon_{tg1}^{(r-1)'},\epsilon_{tg2}^{(r-1)'}\right)'$.

(c) For $g=1,\cdots,G$, given $\widetilde{\epsilon}_{cg}$ and $\epsilon_{tg}^{(r-1)}$ in (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$
P(W_g|\widetilde{\epsilon}_{cg},\epsilon_{tg}^{(r-1)},\phi^{(r)},\widetilde{\lambda}_c,\Upsilon^{(r-1)})=\frac{\exp\left(V_g\left(W_g,\widetilde{\epsilon}_{cg},\epsilon_{tg}^{(r-1)},\phi^{(r)},\widetilde{\lambda}_c,\Upsilon^{(r-1)}\right)\right)}{\sum_W\exp\left(V_g\left(W,\widetilde{\epsilon}_{cg},\epsilon_{tg}^{(r-1)},\phi^{(r)},\widetilde{\lambda}_c,\Upsilon^{(r-1)}\right)\right)}
$$

starting from W_g^*. See details in Step I. part (c).

(d) Let $A = [-1/\tau_G, 1/\tau_G]$, with the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\lambda}_c|\lambda_c^{(r-1)}) \\
&= \min\left\{\prod_{g=1}^{G}\left(\frac{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^*|\phi^{(r)}, \widetilde{\lambda}_c, \Upsilon^{(r-1)}\right)}{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^*|\phi^{(r)}, \lambda_c^{(r-1)}, \Upsilon^{(r-1)}\right)}\cdot\right.\right. \\
&\qquad \left.\left.\frac{P\left(\widetilde{W}_g|\epsilon_{cg}^{(r-1)}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r-1)}, \Upsilon^{(r-1)}\right)}{P\left(\widetilde{W}_g|\widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \widetilde{\lambda}_c, \Upsilon^{(r-1)}\right)}\right)\cdot\frac{I(\widetilde{\lambda}_c\in A)}{I(\lambda_c^{(r-1)}\in A)}, 1\right\} \\
&= \min\left\{\prod_{g=1}^{G}\left(\frac{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r-1)}(\sigma_{\epsilon_c}^{2(r-1)})^{-1}\widetilde{\epsilon}_{cg}\right)}{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r-1)}(\sigma_{\epsilon_c}^{2(r-1)})^{-1}\epsilon_{cg}^{(r-1)}\right)}\cdot\frac{\left|I_{m_g} - \widetilde{\lambda}_c W_g^*\right|}{\left|I_{m_g} - \lambda_c^{(r-1)} W_g^*\right|}\cdot\frac{f(\widetilde{\epsilon}_{cg})}{f(\epsilon_{cg}^{(r-1)})}\cdot\right.\right. \\
&\qquad \left.\left.\frac{\exp\left(\frac{\delta_c^{(r)}}{2}Y_{cg}^{*\prime}(\widetilde{W}_g, \epsilon_{cg}^{(r-1)})Y_{cg}^*(\widetilde{W}_g, \epsilon_{cg}^{(r-1)})\right)}{\exp\left(\frac{\delta_c^{(r)}}{2}Y_{cg}^{*\prime}(\widetilde{W}_g, \widetilde{\epsilon}_{cg})Y_{cg}^*(\widetilde{W}_g, \widetilde{\epsilon}_{cg})\right)}\right)\cdot\frac{I(\widetilde{\lambda}_c\in A)}{I(\lambda_c^{(r-1)}\in A)}, 1\right\},
\end{aligned}
$$

set $\lambda_t^{(r)}$ with $\widetilde{\lambda}_t$. Otherwise, set $\lambda_t^{(r)} = \lambda_t^{(r-1)}$.

Step IV. Simulate $\lambda_t^{(r)}$ from $P(\lambda_t|\{\ddot{Y}_{tg1}^{*(r)}\}, \{Y_{tg}^*\}, \{Y_{cg}^*\}, \{W_g^*\}, \phi^{(r)}, \Upsilon^{(r-1)})$ by the double M-H algorithm.

(a) propose $\widetilde{\lambda}_t$ from a random walk proposal density $q(\lambda_t|\lambda_t^{(r-1)})$

(b) For $g = 1, \cdots, G$, calculate
$\widetilde{\epsilon}_{tg1} = \ddot{Y}_{tg1}^{*(r)} - \left(\widetilde{\lambda}_t W_{12,g}^* Y_{tg2}^* + X_{1g}\beta_{1t}^{(r-1)} + (W_{11,g}^* X_{1g} + W_{12,g}^* X_{2g})\beta_{2t}^{(r-1)} + l_{g1}\alpha_{tg}^{(r-1)}\right)$ and $\widetilde{\epsilon}_{tg2} = (I_{m_g - m_{g1}} - \widetilde{\lambda}_t W_{22,g}^*)Y_{tg2}^* - X_{2g}\beta_{1t}^{(r-1)} - (W_{21,g}^* X_{1g} + W_{22,g}^* X_{2g})\beta_{2t}^{(r-1)} - l_{g2}\alpha_{tg}^{(r-1)}$. Denote $\widetilde{\epsilon}_{tg} = \left(\widetilde{\epsilon}_{tg1}', \widetilde{\epsilon}_{tg2}'\right)'$. To make a distinction, denote $\epsilon_{tg}^{(r-1)} = \left(\epsilon_{tg1}^{(r-1)\prime}, \epsilon_{tg2}^{(r-1)\prime}\right)'$ with $\epsilon_{tg1}^{(r-1)\prime}$ and $\epsilon_{tg2}^{(r-1)\prime}$ calculated based on $\lambda_t^{(r-1)}$. Also calculate
$\epsilon_{cg}^{(r-1)} = \left(I_{m_g} - \lambda_c^{(r)} W_g^*\right) Y_{cg}^* - X_g\beta_{1c}^{(r-1)} - W_g^* X_g \beta_{2c}^{(r-1)} - l_g\alpha_{cg}^{(r-1)}$.

(c) For $g = 1, \cdots, G$, given $\widetilde{\epsilon}_{tg}$ and $\epsilon_{cg}^{(r-1)}$ in (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$
P(W_g|\widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \widetilde{\lambda}_t, \Upsilon^{(r-1)}) = \frac{\exp\left(V_g\left(W_g, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \widetilde{\lambda}_t, \Upsilon^{(r-1)}\right)\right)}{\sum_W \exp\left(V_g\left(W, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \widetilde{\lambda}_t, \Upsilon^{(r-1)}\right)\right)}
$$

starting from W_g^*. See details in Step I. part (c).

(d) Let $A = [-1/\tau_G, 1/\tau_G]$, with the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\lambda}_t|\lambda_t^{(r-1)}) \\
&= \min\left\{ \prod_{g=1}^{G} \left(\frac{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^*|\phi^{(r)}, \lambda_c^{(r)}, \widetilde{\lambda}_t, \Upsilon^{(r-1)}\right)}{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, , W_g^*|\phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r-1)}, \Upsilon^{(r-1)}\right)} \cdot \right.\right. \\
&\qquad \left.\left. \frac{P\left(\widetilde{W}_g|\epsilon_{tg}^{(r-1)}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r-1)}, \Upsilon^{(r-1)}\right)}{P\left(\widetilde{W}_g|\widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \widetilde{\lambda}_t, \Upsilon^{(r-1)}\right)} \right) \cdot \frac{I(\widetilde{\lambda}_t \in A)}{I(\lambda_t^{(r-1)} \in A)}, 1 \right\} \\
&= \min\left\{ \prod_{g=1}^{G} \left(\frac{\left|I_{m_g - m_{g1}} - \widetilde{\lambda}_t W_{22,g}^*\right| f\left(\widetilde{\epsilon}_{tg} - \sigma_{\epsilon_{tc}}^{(r-1)} (\sigma_{\epsilon_c}^{2(r-1)})^{-1} \epsilon_{cg}^{(r-1)}\right)}{\left|I_{m_g - m_{g1}} - \lambda_t^{(r-1)} W_{22,g}^*\right| f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r-1)} (\sigma_{\epsilon_c}^{2(r-1)})^{-1} \epsilon_{cg}^{(r-1)}\right)} \cdot \right.\right. \\
&\qquad \frac{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(W_g^*, \widetilde{\epsilon}_{tg}) Y_{tg}^*(W_g^*, \widetilde{\epsilon}_{tg})\right)}{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(W_g^*, \epsilon_{tg}^{(r-1)}) Y_{tg}^*(W_g^*, \epsilon_{tg}^{(r-1)})\right)} \cdot \\
&\qquad \left.\left. \frac{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(\widetilde{W}_g, \epsilon_{tg}^{(r-1)}) Y_{tg}^*(\widetilde{W}_g, \epsilon_{tg}^{(r-1)})\right)}{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(\widetilde{W}_g, \widetilde{\epsilon}_{tg}) Y_{tg}^*(\widetilde{W}_g, \widetilde{\epsilon}_{tg})\right)} \right) \frac{I(\widetilde{\lambda}_t \in A)}{I(\lambda_t^{(r-1)} \in A)}, 1 \right\},
\end{aligned}
$$

set $\lambda_t^{(r)}$ with $\widetilde{\lambda}_t$. Otherwise, set $\lambda_t^{(r)} = \lambda_t^{(r-1)}$.

Step V. Simulate $\beta_c^{(r)}$ from $P(\beta_c|\{\ddot{Y}_{1g}^{*(r)}\}, \{Y_{tg}^*\}, \{Y_{cg}^*\}, \{W_g^*\}, \phi^{(r)}, \lambda_t^{(r)}, \lambda_c^{(r)}, \Upsilon^{(r-1)})$ by the double M-H algorithm.

(a) propose $\widetilde{\beta}_c$ from a AM proposal $q(\beta_c|\beta_c^{(0)}, \cdots, \beta_c^{(r-1)})$.

(b) For $g = 1, \cdots, G$, calculate $\widetilde{\epsilon}_{cg} = \left(I_{m_g} - \lambda_c^{(r)} W_g^*\right) Y_{cg}^* - X_g \widetilde{\beta}_{1c} - W_g^* X_g \widetilde{\beta}_{2c} - \alpha_{cg}^{(r-1)}$ and $\epsilon_{cg}^{(r-1)} = \left(I_{m_g} - \lambda_c^{(r)} W_g^*\right) Y_{cg}^* - X_g \beta_{1c}^{(r-1)} - W_g^* X_g \beta_{2c}^{(r-1)} - l_g \alpha_{cg}^{(r-1)}$. Also calculate $\epsilon_{tg2}^{(r-1)} = \left(I_{m_g - m_{g1}} - \lambda_t^{(r)} W_{22,g}^*\right) Y_{tg2}^* - X_{2g} \beta_{1t}^{(r-1)} - (W_{21,g}^* X_{1g} + W_{22,g}^* X_{2g}) \beta_{2t}^{(r-1)} - l_{g2} \alpha_{tg}^{(r-1)}$ and
$\epsilon_{tg1}^{(r-1)} = \ddot{Y}_{tg1}^{*(r)} - \left(\lambda^{(r)} W_{12,g}^* Y_{tg2}^* + X_{1g} \beta_{1t}^{(r-1)} + (W_{11,g}^* X_{1g} + W_{12,g}^* X_{2g}) beta_{2t}^{(r-1)} + l_{g1} \alpha_{tg}^{(r-1)}\right)$.
Denote $\epsilon_{tg}^{(r-1)} = \left(\epsilon_{tg1}^{(r-1)\prime}, \epsilon_{tg2}^{(r-1)\prime}\right)^{\prime}$.

(c) For $g = 1, \cdots, G$, given $\widetilde{\epsilon}_{cg}$ and $\epsilon_{tg}^{(r-1)}$ from (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$
\begin{aligned}
&P(W_g|\widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \widetilde{\beta}_c, \Upsilon^{(r-1)}) \\
&\qquad = \frac{\exp\left(V_g(W_g, \widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \widetilde{\beta}_c, \Upsilon^{(r-1)})\right)}{\sum_W \exp\left(V_g(W, \widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \widetilde{\beta}_c, \Upsilon^{(r-1)})\right)}
\end{aligned}
$$

starting from W_g^*. See details in Step I. part (c).

(d) With the acceptance probability equal to

$$\begin{aligned}
&\alpha(\widetilde{\beta}_c|\beta_c^{(r-1)}) \\
&= \min\Bigg\{\prod_{g=1}^{G}\Bigg(\frac{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^* \left|\phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \widetilde{\beta}_c, \Upsilon^{(r-1)}\right.\right)}{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^* \left|\phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r-1)}, \Upsilon^{(r-1)}\right.\right)}\cdot \\
&\quad \frac{P\left(\widetilde{W}_g \left|\epsilon_{cg}^{(r-1)}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r-1)}, \Upsilon^{(r-1)}\right.\right)}{P\left(\widetilde{W}_g \left|\widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \widetilde{\beta}_c, \Upsilon^{(r-1)}\right.\right)}\Bigg)\cdot\frac{\mathcal{N}_{2k}(\widetilde{\beta}_c|\beta_0, B_0)}{\mathcal{N}_{2k}(\beta_c^{(r-1)}|\beta_0, B_0)}, 1\Bigg\} \\
&= \min\Bigg\{\prod_{g=1}^{G}\Bigg(\frac{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r-1)}(\sigma_{\epsilon_c}^{2(r-1)})^{-1}\widetilde{\epsilon}_{cg}\right)}{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r-1)}(\sigma_{\epsilon_c}^{2(r-1)})^{-1}\epsilon_{cg}^{(r-1)}\right)}\cdot\frac{f\left(\widetilde{\epsilon}_{cg}\right)}{f\left(\epsilon_{cg}^{(r-1)}\right)}\cdot \\
&\quad \frac{\exp\left(\frac{\delta_c^{(r)}}{2}Y_{cg}^{*\prime}(\widetilde{W}_g, \epsilon_{cg}^{(r-1)})Y_{cg}^*(\widetilde{W}_g, \epsilon_{cg}^{(r-1)})\right)}{\exp\left(\frac{\delta_c^{(r)}}{2}Y_{cg}^{*\prime}(\widetilde{W}_g, \widetilde{\epsilon_{cg}})Y_{cg}^*(\widetilde{W}_g, \widetilde{\epsilon_{cg}})\right)}\Bigg)\cdot\frac{\mathcal{N}_{2k}(\widetilde{\beta}_c|\beta_0, B_0)}{\mathcal{N}_{2k}(\beta_c^{(r-1)}|\beta_0, B_0)}, 1\Bigg\}
\end{aligned}$$

set $\beta_c^{(r)}$ with $\widetilde{\beta}_c$. Otherwise, set $\beta_c^{(r)} = \beta_c^{(r-1)}$.

Step VI. Simulate $\beta_t^{(r)}$ from $P(\beta_t|\{\ddot{Y}_{1g}^{*(r)}\}, \{Y_{tg}^*\}, \{Y_{cg}^*\}, \{W_g^*\}, \phi^{(r)}, \lambda_t^{(r)}, \lambda_c^{(r)}, \beta_c^{(r)}, \Upsilon^{(r-1)})$ by the double M-H algorithm.

(a) propose $\widetilde{\beta}_t$ from a AM proposal $q(\beta_t|\beta_t^{(0)}, \cdots, \beta_t^{(r-1)})$.

(b) For $g = 1, \cdots, G$, calculate
$\widetilde{\epsilon}_{tg1} = \ddot{Y}_{tg1}^{*(r)} - \left(\lambda_t^{(r)}W_{12,g}^*Y_{tg2}^* + X_{1g}\widetilde{\beta}_{t1} + (W_{11,g}^*X_{1g} + W_{12,g}^*X_{2g})\widetilde{\beta}_{t2} + l_{g1}\alpha_{tg}^{(r-1)}\right)$ and
$\widetilde{\epsilon}_{tg2} = \left(I_{m_g - m_{g1}} - \lambda_t^{(r)}W_{22,g}^*\right)Y_{tg2}^* - X_{2g}\widetilde{\beta}_{t1} - (W_{21,g}^*X_{1g} + W_{22,g}^*X_{2g})\widetilde{\beta}_{t2} - l_{g2}\alpha_{tg}^{(r-1)}$.
Denote $\widetilde{\epsilon}_{tg} = \left(\widetilde{\epsilon}_{tg1}', \widetilde{\epsilon}_{tg2}'\right)'$. To make a distinction, $\epsilon_{tg}^{(r-1)} = \left(\epsilon_{tg1}^{(r-1)\prime}, \epsilon_{tg2}^{(r-1)\prime}\right)'$ is calculated based on $\beta_t^{(r-1)}$. Also calculate $\epsilon_{cg}^{(r-1)} = \left(I_{m_g} - \lambda_c^{(r)}W_g^*\right)Y_{cg}^* - X_g\beta_{c1}^{(r)} - W_g^*X_g\beta_{c2}^{(r)} - l_g\alpha_{cg}^{(r-1)}$.

(c) For $g = 1, \cdots, G$, given $\widetilde{\epsilon}_{tg}$ and $\epsilon_{cg}^{(r-1)}$ from (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$\begin{aligned}
&P(W_g|\widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \widetilde{\beta}_t, \Upsilon^{(r-1)}) \\
&\qquad = \frac{\exp\left(V_g(W_g, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \widetilde{\beta}_t, \Upsilon^{(r-1)})\right)}{\sum_W \exp\left(V_g(W, \widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \widetilde{\beta}_t, \Upsilon^{(r-1)})\right)}
\end{aligned}$$

See details in Step I. part (c).

(d) With the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\beta}_t|\beta_t^{(r-1)}) \\
&= \min\left\{ \prod_{g=1}^{G} \left(\frac{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^* \,\middle|\, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \widetilde{\beta}_t, \Upsilon^{(r-1)}\right)}{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{tg}^*, Y_{cg}^*, W_g^* \,\middle|\, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \beta_t^{(r-1)}, \Upsilon^{(r-1)}\right)} \cdot \right.\right. \\
&\qquad \left.\left. \frac{P\left(\widetilde{W}_g \,\middle|\, \epsilon_{cg}^{(r-1)}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \beta_t^{(r-1)}, \Upsilon^{(r-1)}\right)}{P\left(\widetilde{W}_g \,\middle|\, \widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \widetilde{\beta}_t, \Upsilon^{(r-1)}\right)} \right) \cdot \frac{\mathcal{N}_{2k}(\widetilde{\beta}_t|\beta_0, B_0)}{\mathcal{N}_{2k}(\beta_t^{(r-1)}|\beta_0, B_0)}, 1 \right\} \\
&= \min\left\{ \prod_{g=1}^{G} \left(\frac{f\left(\widetilde{\epsilon}_{tg} - \sigma_{\epsilon_{tc}}^{(r-1)}(\sigma_{\epsilon_c}^{2(r-1)})^{-1}\epsilon_{cg}^{(r-1)}\right)}{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r-1)}(\sigma_{\epsilon_c}^{2(r-1)})^{-1}\epsilon_{cg}^{(r-1)}\right)} \cdot \right.\right. \\
&\qquad \frac{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(W_g^*, \widetilde{\epsilon}_{tg}) Y_{tg}^*(W_g^*, \widetilde{\epsilon}_{tg})\right)}{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(W_g^*, \epsilon_{tg}^{(r-1)}) Y_{tg}^*(W_g^*, \epsilon_{tg}^{(r-1)})\right)} \cdot \\
&\qquad \left.\left. \frac{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(\widetilde{W}_g, \epsilon_{tg}^{(r-1)}) Y_{cg}^*(\widetilde{W}_g, \epsilon_{tg}^{(r-1)})\right)}{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(\widetilde{W}_g, \widetilde{\epsilon_{cg}}) Y_{tg}^*(\widetilde{W}_g, \widetilde{\epsilon_{tg}})\right)} \right) \cdot \frac{\mathcal{N}_{2k}(\widetilde{\beta}_t|\beta_0, B_0)}{\mathcal{N}_{2k}(\beta_t^{(r-1)}|\beta_0, B_0)}, 1 \right\}
\end{aligned}
$$

set $\beta_t^{(r)}$ with $\widetilde{\beta}_t$. Otherwise, set $\beta_t^{(r)} = \beta_t^{(r-1)}$.

Step VII. Simulate $\sigma^{(r)}$ from $P(\sigma|\{\ddot{Y}_{1g}^{*(r)}\}, \{Y_{tg}^*\}, \{Y_{cg}^*\}, \{W_g^*\}, \phi^{(r)}, \lambda_t^{(r)}, \lambda_c^{(r)}, \beta_c^{(r)}, \beta_t^{(r)}, \Upsilon^{(r-1)})$ by the standard M-H algorithm.

(a) propose $\widetilde{\sigma}$ from a AM proposal $q(\sigma|\sigma^{(0)}, \cdots, \sigma^{(r-1)})$.

(b) For $g = 1, \cdots, G$, calculate $\epsilon_{tg1}^{(r-1)} = \ddot{Y}_{tg1}^{*(r)} - \left(\lambda_t^{(r)} W_{12,g}^* Y_{tg2}^* + X_{1g}\beta_{t1}^{(r)} + (W_{11,g}^* X_{1g} + W_{12,g}^* X_{2g})\beta_{t2}^{(r)} + l_{g1}\alpha_{tg}^{(r-1)}\right)$ and $\epsilon_{tg2}^{(r-1)} = \left(I_{m_g - m_{g1}} - \lambda_t^{(r)} W_{22,g}^*\right) Y_{tg2}^* - X_{2g}\beta_{t1}^{(r)} - (W_{21,g}^* X_{1g} + W_{22,g}^* X_{2g})\beta_{t2}^{(r)} - l_{g2}\alpha_{tg}^{(r-1)}$. Denote $\epsilon_{tg}^{(r-1)} = \left(\epsilon_{tg1}^{(r-1)\prime}, \epsilon_{tg2}^{(r-1)\prime}\right)^{\prime}$. Also calculate $\epsilon_{cg}^{(r-1)} = \left(I_{m_g} - \lambda_c^{(r)} W_g^*\right) Y_{cg}^* - X_g\beta_{c1}^{(r)} - W_g^* X_g \beta_{c2}^{(r)} - l_g \alpha_{cg}^{(r-1)}$.

(c) With the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\sigma}|\sigma^{(r-1)}) \\
&= \min\left\{\prod_{g=1}^{G}\left(\frac{P\left(\ddot{Y}^{*(r)}_{tg1}, Y^*_{tg}, Y^*_{cg}, W^*_g|\phi^{(r)}, \lambda^{(r)}_c, \lambda^{(r)}_t, \beta^{(r)}_c, \beta^{(r)}_t, \widetilde{\sigma}, \Upsilon^{(r-1)}\right)}{P\left(\ddot{Y}^{*(r)}_{tg1}, Y^*_{tg}, Y^*_{cg}, W^*_g|\phi^{(r)}, \lambda^{(r)}_c, \lambda^{(r)}_t, \beta^{(r)}_c, \beta^{(r)}_t, \sigma^{(r-1)}, \Upsilon^{(r-1)}\right)}\right)\cdot\right. \\
&\qquad \left.\frac{\mathscr{N}_3(\widetilde{\sigma}|\sigma_0, \Sigma_0)}{\mathscr{N}_3(\sigma^{(r-1)}|\sigma_0, \Sigma_0)}\cdot\frac{I(\tilde{\sigma}\in T)}{I(\sigma^{(r-1)}\in T)}, 1\right\} \\
&= \min\left\{\prod_{g=1}^{G}\left(\frac{f\left(\epsilon^{(r-1)}_{tg} - \widetilde{\sigma}_{\epsilon_{tc}}(\widetilde{\sigma}^2_{\epsilon_c})^{-1}\epsilon^{(r-1)}_{cg}; \widetilde{\sigma}\right)}{f\left(\epsilon^{(r-1)}_{tg} - \sigma^{(r-1)}_{\epsilon_{tc}}(\sigma^{2(r-1)}_{\epsilon_c})^{-1}\epsilon^{(r-1)}_{cg}; \sigma^{(r-1)}\right)}\cdot\frac{f\left(\epsilon^{(r-1)}_{cg}; \widetilde{\sigma}\right)}{f\left(\epsilon^{(r-1)}_{cg}; \sigma^{(r-1)}\right)}\right)\cdot\right. \\
&\qquad \left.\frac{\mathscr{N}_3(\widetilde{\sigma}|\sigma_0, \Sigma_0)}{\mathscr{N}_3(\sigma^{(r-1)}|\sigma_0, \Sigma_0)}\cdot\frac{I(\tilde{\sigma}\in T)}{I(\sigma^{(r-1)}\in T)}, 1\right\},
\end{aligned}
$$

set $\sigma^{(r)}$ with $\widetilde{\sigma}$. Otherwise, set $\sigma^{(r)} = \sigma^{(r-1)}$.

Step VIII. For $g = 1, \cdots, G$, simulate $\alpha^{(r)}_{cg}$ from $P(\alpha_{cg}|\ddot{Y}^{*(r)}_{1g}, Y^*_{tg}, Y^*_{cg}, W^*_g, \phi^{(r)}, \lambda^{(r)}_t, \lambda^{(r)}_c, \beta^{(r)}_c, \beta^{(r)}_t, \sigma^{2(r)}_{\epsilon_c}, \sigma^{2(r)}_{\epsilon_t}, \sigma^{(r)}_{\epsilon_{tc}}, \alpha^{(r-1)}_{tg})$ by the double M-H algorithm.

(a) propose $\widetilde{\alpha}_{cg}$ from a random walk proposal density $q(\alpha_{cg}|\alpha^{(r-1)}_{cg})$

(b) Calculate $\widetilde{\epsilon}_{cg} = \left(I_{m_g} - \lambda^{(r)}_c W^*_g\right) Y^*_{cg} - X_g\beta^{(r)}_{c1} - W^*_g X_g \beta^{(r)}_{c2} - l_g\widetilde{\alpha}_{cg}$ and $\epsilon^{(r-1)}_{cg} = \left(I_{m_g} - \lambda^{(r)}_c W^*_g\right) Y^*_{cg} - X_g\beta^{(r)}_{c1} - W^*_g X_g\beta^{(r)}_{c2} - l_g\alpha^{(r-1)}_{cg}$. Also calculate $\epsilon^{(r-1)}_{tg2} = \left(I_{m_g - m_{g1}} - \lambda^{(r)}_t W^*_{22,g}\right) Y^*_{tg2} - X_{2g}\beta^{(r)}_{1t} - (W^*_{21,g}X_{1g} + W^*_{22,g}X_{2g})\beta^{(r)}_{2t} - l_{g2}\alpha^{(r-1)}_{tg}$ and $\epsilon_{tg1} = \ddot{Y}^{*(r)}_{tg1} - \left(\lambda^{(r)}_t W^*_{12,g}Y^*_{tg2} + X_{1g}\beta^{(r)}_{1t} + (W^*_{11,g}X_{1g} + W^*_{12,g}X_{2g})\beta^{(r)}_{2t} + l_{g1}\alpha^{(r-1)}_{tg}\right)$. Denote $\epsilon^{(r-1)}_{tg} = \left(\epsilon^{(r-1)'}_{tg1}, \epsilon^{(r-1)'}_{tg2}\right)'$.

(c) Given $\widetilde{\epsilon}_{cg}$ and $\epsilon^{(r-1)}_{tg}$ from (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$
\begin{aligned}
&P(W_g|\widetilde{\epsilon}_{cg}, \epsilon^{(r-1)}_{tg}, \phi^{(r)}, \lambda^{(r)}_c, \lambda^{(r)}_t, \beta^{(r)}_c, \sigma^{2(r)}_{\epsilon_c}, \sigma^{2(r)}_{\epsilon_t}, \sigma^{(r)}_{\epsilon_{tc}}, \widetilde{\alpha}_{cg}, \alpha^{(r-1)}_{tg}) \\
&\quad = \frac{\exp\left(V_g(W_g, \widetilde{\epsilon}_{cg}, \epsilon^{(r-1)}_{tg}, \phi^{(r)}, \lambda^{(r)}_c, \lambda^{(r)}_t, \beta^{(r)}_c, \sigma^{2(r)}_{\epsilon_c}, \sigma^{2(r)}_{\epsilon_t}, \sigma^{(r)}_{\epsilon_{tc}}, \widetilde{\alpha}_{cg}, \alpha^{(r-1)}_{tg})\right)}{\sum_W \exp\left(V_g(W, \widetilde{\epsilon}_{cg}, \epsilon^{(r-1)}_{tg}, \phi^{(r)}, \lambda^{(r)}_c, \lambda^{(r)}_t, \beta^{(r)}_c, \sigma^{2(r)}_{\epsilon_c}, \sigma^{2(r)}_{\epsilon_t}, \sigma^{(r)}_{\epsilon_{tc}}, \widetilde{\alpha}_{cg}, \alpha^{(r-1)}_{tg})\right)}
\end{aligned}
$$

See details in Step I. part (c).

(d) With the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\alpha}_{cg}|\alpha_{cg}^{(r-1)}) \\
&= \min\left\{ \frac{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{cg}^{*}, Y_{tg}^{*}, W_g^{*} \,\middle|\, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \widetilde{\alpha}_{cg}, \alpha_{tg}^{(r-1)}\right)}{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{cg}^{*}, Y_{tg}^{*}, W_g^{*} \,\middle|\, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r-1)}, \alpha_{tg}^{(r-1)}\right)} \cdot \right. \\
&\qquad \frac{P\left(\widetilde{W}_g \,\middle|\, \epsilon_{cg}^{(r-1)}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r-1)}, \alpha_{tg}^{(r-1)}\right)}{P\left(\widetilde{W}_g \,\middle|\, \widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \widetilde{\alpha}_{cg}, \alpha_{tg}^{(r-1)}\right)} \cdot \\
&\qquad \left. \frac{\mathscr{N}(\widetilde{\alpha}_{cg}|\alpha_0, A_0)}{\mathscr{N}(\alpha_{cg}^{(r-1)}|\alpha_0, A_0)}, 1 \right\} \\
&= \min\left\{ \left(\frac{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r)}(\sigma_{\epsilon_c}^{2(r)})^{-1}\widetilde{\epsilon}_{cg}\right)}{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r)}(\sigma_{\epsilon_c}^{2(r)})^{-1}\epsilon_{cg}^{(r-1)}\right)} \cdot \frac{g\left(\widetilde{\epsilon}_{cg}\right)}{g\left(\epsilon_{cg}^{(r-1)}\right)} \cdot \right.\right. \\
&\qquad \left.\left. \frac{\exp\left(\frac{\delta_c^{(r)}}{2} Y_{cg}^{*\prime}(\widetilde{W}_g, \epsilon_{cg}^{(r-1)}) Y_{cg}^{*}(\widetilde{W}_g, \epsilon_{cg}^{(r-1)})\right)}{\exp\left(\frac{\delta_c^{(r)}}{2} Y_{cg}^{*\prime}(\widetilde{W}_g, \widetilde{\epsilon_{cg}}) Y_{cg}^{*}(\widetilde{W}_g, \widetilde{\epsilon}_{cg})\right)} \right) \cdot \frac{\mathscr{N}(\widetilde{\alpha}_{cg}|\alpha_0, A_0)}{\mathscr{N}(\alpha_{cg}^{(r-1)}|\alpha_0, A_0)}, 1 \right\}
\end{aligned}
$$

set $\alpha_{cg}^{(r)}$ with $\widetilde{\alpha}_{cg}$. Otherwise, set $\alpha_{cg}^{(r)} = \alpha_{cg}^{(r-1)}$.

Step IX. For $g = 1, \cdots, G$, simulate $\alpha_{tg}^{(r)}$ from $P(\alpha_{tg}|\ddot{Y}_{1g}^{*(r)}, Y_{tg}^{*}, Y_{cg}^{*}, W_g^{*}, \phi^{(r)}, \lambda_t^{(r)}, \lambda_c^{(r)}, \beta_c^{(r)}, \beta_t^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)})$ by the double M-H algorithm.

(a) propose $\widetilde{\alpha}_{tg}$ from a random walk proposal density $q(\alpha_{tg}|\alpha_{tg}^{(r-1)})$

(b) For $g = 1, \cdots, G$, calculate $\widetilde{\epsilon}_{tg1} = \ddot{Y}_{tg1}^{*(r)} - \left(\lambda_t^{(r)} W_{12,g}^{*} Y_{tg2}^{*} + X_{1g}\beta_{t1}^{(r)} + (W_{11,g}^{*}X_{1g} + W_{12,g}^{*}X_{2g})\beta_{t2}^{(r)} + l_{g1}\widetilde{\alpha}_{tg}\right)$ and $\widetilde{\epsilon}_{tg2} = \left(I_{m_g - m_{g1}} - \lambda_t^{(r)} W_{22,g}^{*}\right) Y_{tg2}^{*} - X_{2g}\beta_{t1}^{(r)} - (W_{21,g}^{*}X_{1g} + W_{22,g}^{*}X_{2g})\beta_{t2}^{(r)} - l_{g2}\widetilde{\alpha}_{tg}$. Denote $\widetilde{\epsilon}_{tg} = \left(\widetilde{\epsilon}'_{tg1}, \widetilde{\epsilon}'_{tg2}\right)'$. To make a distinction, $\epsilon_{tg}^{(r-1)} = \left(\epsilon_{tg1}^{(r-1)\prime}, \epsilon_{tg2}^{(r-1)\prime}\right)'$ is calculated based on $\alpha_{tg}^{(r-1)}$. Also calculate $\epsilon_{cg}^{(r)} = \left(I_{m_g} - \lambda_c^{(r)} W_g^{*}\right) Y_{cg}^{*} - X_g\beta_{c1}^{(r)} - W_g^{*}X_g\beta_{c2}^{(r)} - l_g\alpha_{cg}^{(r)}$.

(c) Given $\widetilde{\epsilon}_{tg}$ and $\epsilon_{cg}^{(r)}$ from (b), simulate an auxiliary network $\widetilde{W}_g$ by m runs of the M-H algorithm based on

$$
\begin{aligned}
&P(W_g|\widetilde{\epsilon}_{tg}, \epsilon_{cg}^{(r)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \widetilde{\alpha}_{tg}) \\
&\quad = \frac{\exp\left(V_g(W_g, \widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \widetilde{\alpha}_{tg})\right)}{\sum_W \exp\left(V_g(W, \widetilde{\epsilon}_{cg}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \widetilde{\alpha}_{tg})\right)}
\end{aligned}
$$

See details in Step I. part (c).

(d) With the acceptance probability equal to

$$
\begin{aligned}
&\alpha(\widetilde{\alpha}_{tg}|\alpha_{tg}^{(r-1)})\\
&= \min\left\{ \frac{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{cg}^{*}, Y_{tg}^{*}, W_g^{*} \,\middle|\, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \widetilde{\alpha}_{tg}\right)}{P\left(\ddot{Y}_{tg1}^{*(r)}, Y_{cg}^{*}, Y_{tg}^{*}, W_g^{*} \,\middle|\, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \alpha_{tg}^{(r-1)}\right)} \cdot \right.\\
&\qquad \frac{P\left(\widetilde{W}_g \,\middle|\, \epsilon_{cg}^{(r)}, \epsilon_{tg}^{(r-1)}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \alpha_{tg}^{(r-1)}\right)}{P\left(\widetilde{W}_g \,\middle|\, \epsilon_{cg}^{(r)}, \widetilde{\epsilon}_{tg}, \phi^{(r)}, \lambda_c^{(r)}, \lambda_t^{(r)}, \beta_c^{(r)}, \sigma_{\epsilon_c}^{2(r)}, \sigma_{\epsilon_t}^{2(r)}, \sigma_{\epsilon_{tc}}^{(r)}, \alpha_{cg}^{(r)}, \widetilde{\alpha}_{tg}\right)} \cdot \\
&\qquad \left. \frac{\mathcal{N}(\widetilde{\alpha}_{tg}|\alpha_0, A_0)}{\mathcal{N}(\alpha_{tg}^{(r-1)}|\alpha_0, A_0)}, 1 \right\}\\
&= \min\left\{ \left(\frac{f\left(\widetilde{\epsilon}_{tg} - \sigma_{\epsilon_{tc}}^{(r)}(\sigma_{\epsilon_c}^{2(r)})^{-1}\epsilon_{cg}^{(r)}\right)}{f\left(\epsilon_{tg}^{(r-1)} - \sigma_{\epsilon_{tc}}^{(r)}(\sigma_{\epsilon_c}^{2(r)})^{-1}\epsilon_{cg}^{(r)}\right)} \cdot \frac{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(W_g^{*}, \widetilde{\epsilon}_{tg}) Y_{tg}^{*}(W_g^{*}, \widetilde{\epsilon}_{tg})\right)}{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(W_g^{*}, \epsilon_{tg}^{(r-1)}) Y_{tg}^{*}(W_g^{*}, \epsilon_{tg}^{(r-1)})\right)} \cdot \right.\right.\\
&\qquad \left.\left. \frac{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(\widetilde{W}_g, \epsilon_{tg}^{(r-1)}) Y_{tg}^{*}(\widetilde{W}_g, \epsilon_{tg}^{(r-1)})\right)}{\exp\left(\frac{\delta_t^{(r)}}{2} Y_{tg}^{*\prime}(\widetilde{W}_g, \widetilde{\epsilon_{tg}}) Y_{tg}^{*}(\widetilde{W}_g, \widetilde{\epsilon_{tg}})\right)} \right) \cdot \frac{\mathcal{N}(\widetilde{\alpha}_{tg}|\alpha_0, A_0)}{\mathcal{N}(\alpha_{tg}^{(r-1)}|\alpha_0, A_0)}, 1 \right\}
\end{aligned}
$$

set $\alpha_{tg}^{(r)}$ with $\widetilde{\alpha}_{tg}$. Otherwise, set $\alpha_{tg}^{(r)} = \alpha_{tg}^{(r-1)}$.

APPENDIX D: Computational performances of our MCMC algorithms

We conduct a simulation study to examine computational aspects of the proposed MCMC algorithms. The artificial data is generated by a data generating process (DGP) based on continuous outcomes of Eq. (3)[34] and the aggregated link utility function of Eq. (11). For simplicity, we only specify one exogenous variable X and do not include WX as another regressor in the outcome equation. As for the link utility function, we capture the exogenous effect with a dyad-specific variable C and a matrix of ones for intercept terms. The networks are generated in three mixed sizes: 20, 30, and 40, and the corresponding number of networks for each size is 10. The exogenous variable X for outcomes are generated from a normal distribution with a zero mean and a variance equal to 25. The group effect α's are generated from a normal distribution with a mean equal to the group average of X times 0.3 and a variance equal to 1[35]. The disturbance term ϵ's are generated from a normal distribution with a zero mean and a variance equal to σ_ϵ^2 as will be specified below. The elements of the dyad-specific variable C for the exogenous effect are generated by first drawing two vectors of random variables from $U(0, 1)$, U_1 and U_2. If the ith element of U_1 and the jth element of U_2 are both larger than 0.7 or less than 0.3, we set C_{ij} equal to one. Otherwise, we set it to zero.

By assigning the following parameters

- Network: $\gamma_{31} = -3.2$; $\gamma_{32} = 0.4$; $\eta_1 = 0.4$; $\eta_2 = 0.2$; $\eta_3 = -0.03$; $\eta_4 = 0.03$; $\eta_5 = 0.30$; $\eta_6 = -0.20$; $\delta = 0.20$,
- Outcome: $\lambda = 0.05$; $\beta = 0.50$; $\sigma_\epsilon^2 = 0.50$,

into these two equations and use the exponential distribution of Eq. (7), we can then simulate the networks and the activity outcomes: using the designed parameters and those generated exogenous components, each artificial network W is simulated starting from an empty network. Each entry of W except the ones on the diagonal is updated using the Gibbs sampling with the conditional probability of w_{ij} in Eq. (13). Outcome variables are simulated along with the network. The Gibbs sampling runs through the whole network for total 10,000 iterations and realizations of the network and outcomes from the last iteration are used for the data. Those generated networks

[34]The same simulation study has been done for the case of simultaneous Tobit outcomes based on Eq. (4)) and we obtain similar results. For the purpose of illustration without unnecessary repeats, we only choose to report the results based on continuous outcomes here.

[35]We follows Mundlak (1978)'s specification to generate the dependence between group effects and exogenous regressors.

have the average density equal to 0.083, the average outdegree equal to 2.498, and the average clustering coefficient equal to 0.125.

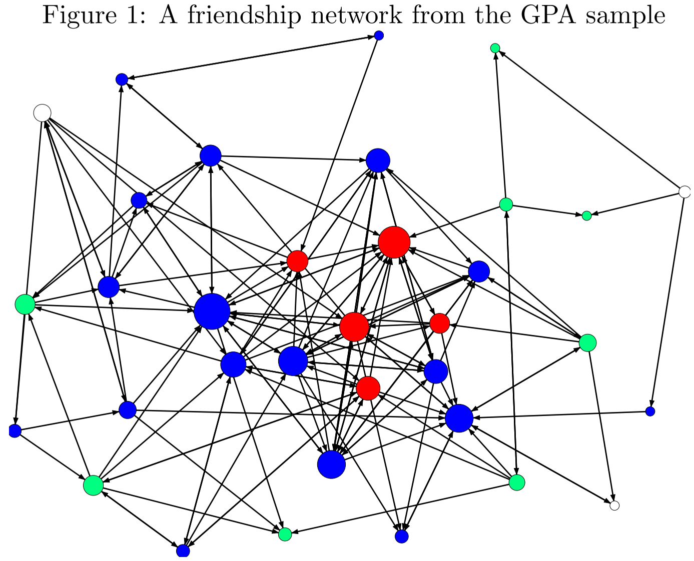
We estimate the true model using the algorithm provided in Appendix C-3. The idea is see whether the proposed algorithm can return the correct estimates of parameters. To check how well and how fast the posterior draws will converge to the target distribution, three sets of initial values are assigned to the sampling process, $\phi = (\gamma', \eta', \delta') =$

- (-2.00, 0.20, 0.20, 0.10, -0.01, 0.01, 0.20, -0.10, 0.10);
- (0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.10);
- (-5.00, 1.00, 1.00, 1.00, -0.05, -0.50, 0.50, -0.50, 0.50),

and initial values for all other parameters are set to 0. 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 100,000 draws of $(\gamma, \eta, \delta, \lambda)$ based on these three initial values are plotted separately in Figure 3 to Figure 5.

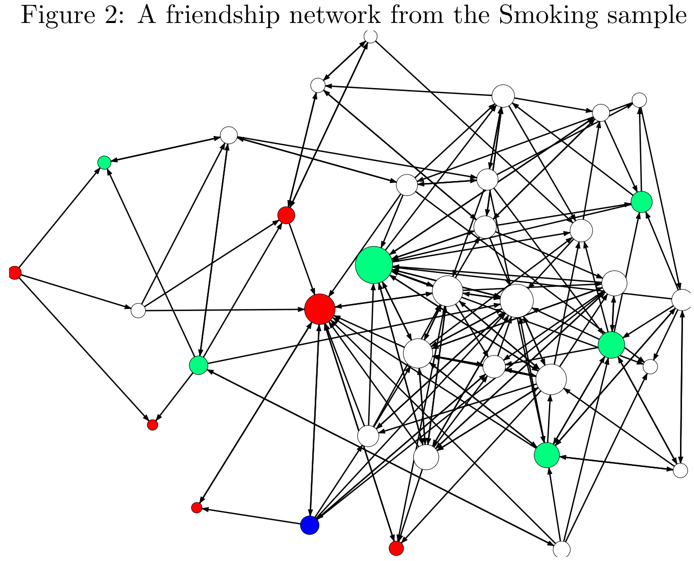
From Figure 3 one can observe, when starting with initial values which are close to the true parameters, the convergence of the Markov chain achieves in a timely fashion. In this case it costs about the first 10,000 draws to achieve convergence. Once the Markov chain converges and is stable at a certain level, the following draws show a nice variation which represents the variance of the target posterior distribution. In the following Figure 4 and 5, bad initial values cause the Markov chain to take longer for the convergence. Since the draws approach the true parameters gradually and hence, exhibit certain degree of dependence. In these two cases, more draws would be needed to make sure that the posterior distributions of parameters are characterized properly. From this simulation, it shows that the double M-H algorithm can handle the estimation of our model without any problem.

Figure 1: A friendship network from the GPA sample



Note: White: GPA less than 2. Green: GPA between 2 and 3. Blue: GPA between 3 and 4. Red: GPA equal to 4. Nodes with a larger size means they have higher indegrees.

Figure 2: A friendship network from the Smoking sample



Note: White: Do not smoke. Green: Smoke once. Blue: Smoke 3 to 5 times. Red: Smoke Every day. Nodes with a larger size means they have higher indegrees.

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 3: plot of MCMC draws from the continuous case – 1^{st} set of initial values

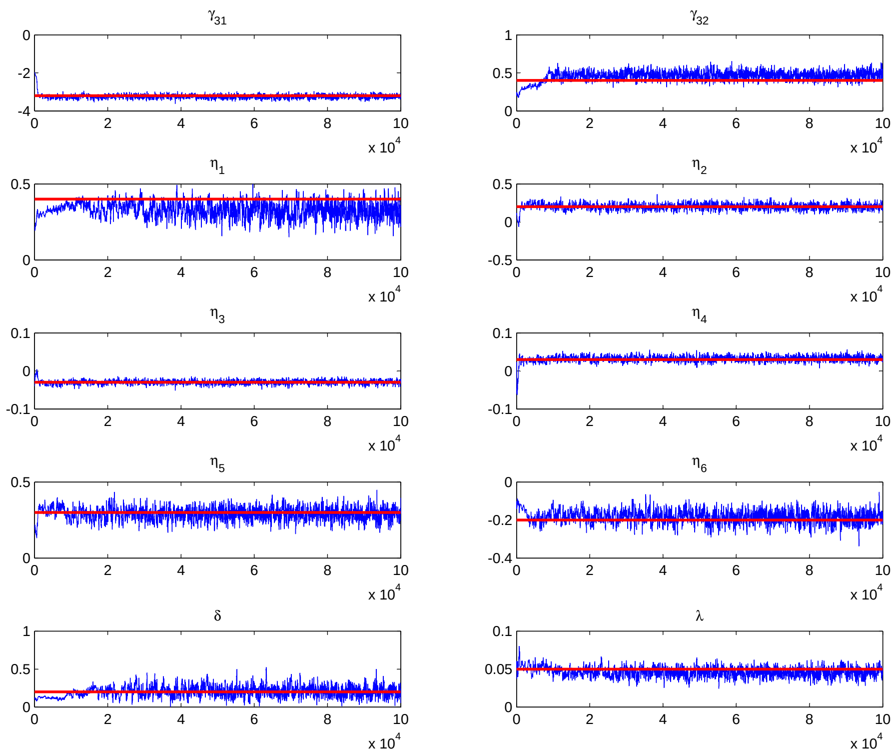


Figure 4: plot of MCMC draws from the continuous case – 2^{nd} set of initial values

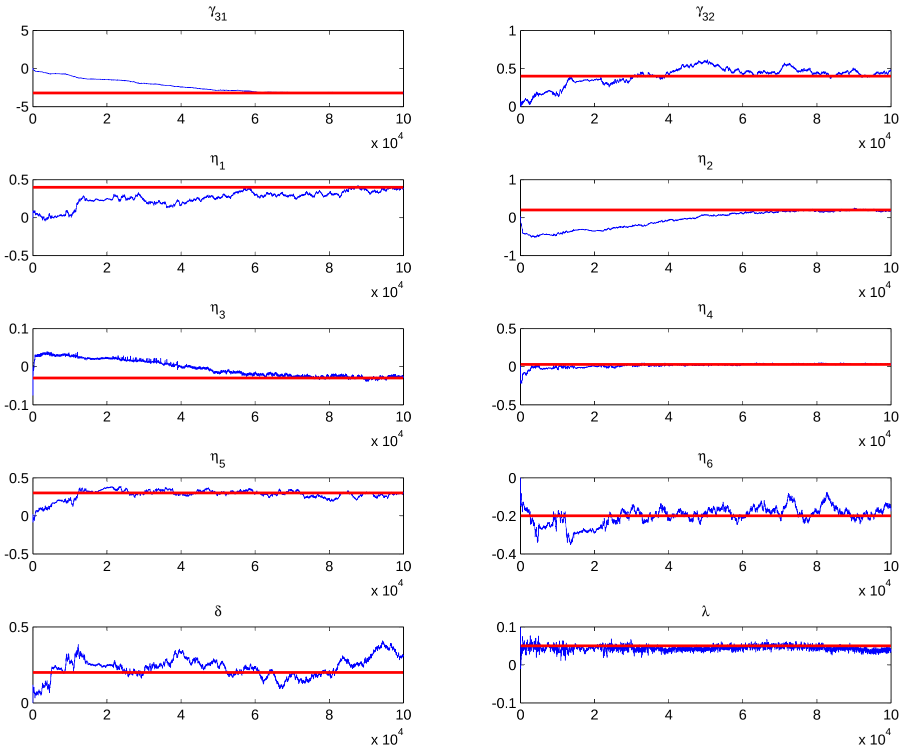


Figure 5: plot of MCMC draws from the continuous case – 3^{rd} set of initial values

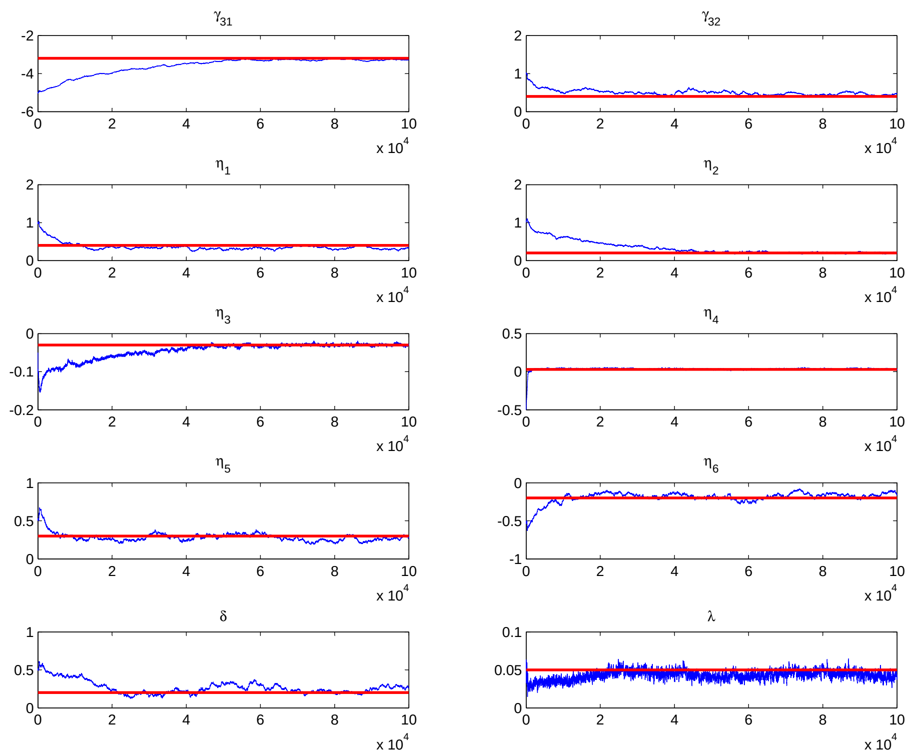


Table 2: Estimation result based on GPA

	[Full Model]				[Activity Outcome Only]				[Full Model with missing outcomes]			
Network Model												
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)	1.118	(0.225)			-				1.056	(0.218)		
Outcome Equation												
Endogenous	0.021	(0.007)			0.040	(0.012)			0.023	(0.006)		
σ_ϵ^2	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)
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)

Continued on Next Page

Table – Continued

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 Model]				[Activity Outcome Only]			
Network Model								
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)	0.023	(0.017)			-			
Outcome Equation								
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.

Table 4: Estimation result based on both GPA and Smoking

	[Full Model]				[Activity Outcome Only]			
Network Model								
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)				-			
Outcome Equation	[GPA]		[Smoking]		[GPA]		[Smoking]	
Endogenous	0.025 (0.005)		0.079 (0.013)		0.049 (0.015)		0.085 (0.013)	
σ^2_ϵ	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.

References

[1] Albert, J. H. and S. Chib (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 Y. Zenou (2006). Who's Who in Netowrks. Wanted: the Key Player. *Econometrica* 74, 1403-1417.

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

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

[6] Bayer, P., S. L. Ross, and G. Topa (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., R. Hjalmarsson, and D. Pozen (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., Y. Bramoullé, H. Djebbari and B. Fortin (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., H. Djebbari and B. Fortin (2009). Identification of Peer Effects through Social Networks. *Journal of Econometrics* 150, 41-55.

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

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

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

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

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

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

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

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

[20] Clark, A. and Y. Loheac (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 C. R. Udry (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 C. R. Udry (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] Cressie N. A.C. (1993). Statistics for Spatial Data, John Wiley and Sons, Inc., New York.

[25] Currarini, S., M. O. Jackson, and P. Pin (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.

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

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

[28] 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.

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

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

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

[32] Gelman, A. (1996). Inference and Monitoring Convergence. In Gilks, W. R., S. Richardson and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*.

[33] Gelman, A., G. O. Roberts, and W. R. Gilks (1996). Efficient Metropolis Jumping Rules. *Bayesian Statistics* 5, 599-607.

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

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

[37] Haario, H., E. Saksman, and J. Tamminen (2001). An Adaptive Metropolis Algorithm. *Bernoulli* 7, 223-242.

[38] Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based Clustering for Social Networks. *Journal of the Royal Sataistical Society* 170, 301-54.

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

[40] Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97, 1090-98.

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

[42] Hsieh, C. S. and L. F. Lee (2012). 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] Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*, 103, 248-58.

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

[46] 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.

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

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

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

[50] Kovářík, J. and M. J. van der Leij (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 X. Lin (2007). Binary Choice Models with Social Interactions under Heterogeneous Rational Expectations. Unpublished working paper.

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

[54] LeSage, J. P. (1997). Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review* 20, 113-29.

[55] LeSage, J. P. (2000). Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis* 32, 19-35.

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

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

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

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

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

[61] Monderer, D. and L. S. Shapley (1996). Potential Games. *Games and Economic Behavior* 14, 124-143.

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

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

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

[65] McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. in P. Zarembka (ed), *Frontiers in Econometrics* Academic Press, New York 105-142.

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

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

[68] Møller, J., A. N. Pettitt, R. Reeves, and K. K. Berthelsen (2006). Miscellanea: An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalizing Constants. *Biometrika*, 93, 451-58.

[69] Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica* 46, 69-85.

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

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

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

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

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

[75] Roberts, G. and J. S. Rosenthal (2009). Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics* 18, 349-67.

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

[77] Ruelle, D. (1969). Statistical Mechanics: Rigorous Results. Benjamin, New York.

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

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

[80] 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.

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

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

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

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

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

[86] Wasserman, S. S. and K. Faust (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

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

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

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