

Hsieh, Chih-Sheng; Lee, Lung-fei; Boucher, Vincent

## Article

# Specification and estimation of network formation and network interaction models with the exponential probability distribution

Quantitative Economics

## Provided in Cooperation with:

The Econometric Society

*Suggested Citation:* Hsieh, Chih-Sheng; Lee, Lung-fei; Boucher, Vincent (2020) : Specification and estimation of network formation and network interaction models with the exponential probability distribution, Quantitative Economics, ISSN 1759-7331, The Econometric Society, New Haven, CT, Vol. 11, Iss. 4, pp. 1349-1390,  
<https://doi.org/10.3982/QE944>

This Version is available at:

<https://hdl.handle.net/10419/253562>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc/4.0/>

# Specification and estimation of network formation and network interaction models with the exponential probability distribution

CHIH-SHENG HSIEH

Department of Economics, National Taiwan University

LUNG-FEI LEE

Department of Economics, The Ohio State University

VINCENT BOUCHER

Department of Economics, Université Laval, CRREP, and CREATE

We model network formation and interactions under a unified framework by considering that individuals anticipate the effect of network structure on the utility of network interactions when choosing links. There are two advantages of this modeling approach: first, we can evaluate whether network interactions drive friendship formation or not. Second, we can control for the friendship selection bias on estimated interaction effects. We provide microfoundations of this statistical model based on the subgame perfect equilibrium of a two-stage game and propose a Bayesian MCMC approach for estimating the model. We apply the model to study American high school students' friendship networks using the Add Health dataset. From two interaction variables, GPA and smoking frequency, we find that the utility of interactions in academic learning is important for friendship formation, whereas the utility of interactions in smoking is not. However, both GPA and smoking frequency are subject to significant peer effects.

---

Chih-Sheng Hsieh: [cshsieh@ntu.edu.tw](mailto:cshsieh@ntu.edu.tw)

Lung-Fei Lee: [lee.1777@osu.edu](mailto:lee.1777@osu.edu)

Vincent Boucher: [vincent.boucher@ecn.ulaval.ca](mailto:vincent.boucher@ecn.ulaval.ca)

A previous version of this paper was circulated under the title, “A structural modeling approach for network formation and social interactions with applications to students' friendship choices and selectivity on activities.” The authors are grateful to Le-Yu Chen, Matthew Jackson, Xiaodong Liu, Angelo Mele, Ching-Jen Sun, anonymous referees, as well as conference participants at 2013 AMES in Singapore and seminar participants at Academia Sinica, Chicago Booth, Chinese University of Hong Kong, Lingnan University, National Taiwan University, Ohio State University, Shanghai University of Finance and Economics, and Xiamen University for helpful comments. Thanks also to Murray Hay for proofreading the manuscript. Vincent Boucher gratefully acknowledges financial support from SSHRC (430-2016-00777). This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

**KEYWORDS.** Social networks, social interactions, selectivity, spatial autoregressive model, Bayesian estimation.

**JEL CLASSIFICATION.** C21, C25, I21, J13.

## 1. INTRODUCTION

Economic research on social networks and interactions has grown rapidly over the past two decades. In many economic contexts, social networks have been found to be an important channel to disseminate information or facilitate activities.<sup>1</sup> Due to the importance of social networks for a wide range of applications, both academic researchers and practitioners have been and are still interested in understanding how network links are formed. Indeed, the question is not only interesting in its own right, but it is also important for understanding the role of network structures on economic outcomes.

For example, in the context of social interactions, we would like to understand how individuals choose their friends to benefit from peer effects on economic outcomes. In particular, friendship networks may be formed to produce favorable economic consequences; for example, students may prefer choosing high achieving friends who can help them study. Then, if one is interested in measuring peer effects on academic achievement, they need to correct for possible endogeneity bias due to friendship selection, as itself might also be based on academic achievement.

Moreover, endogenous friendship formation may amplify observed peer interactions due to unobserved factors that affect both friendship selection and economic outcomes (Weinberg (2007)). For example, Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Johnsson and Moon (forthcoming), and Auerbach (2019) study unobserved driving factors and use them to link network formation and network interactions to economic activities.

In this paper, we propose an unified modeling approach for individuals who form a friendship network inside a group and have their economic behaviors influenced by their friends' behaviors once the network is formed. In particular, we focus on a static model and present a novel approach for examining whether utilities of interactions in certain economic activities (e.g., peer influence on tobacco consumption) play any role in the formation of friendship networks for high school students.<sup>2</sup> Specifically, we allow

<sup>1</sup>For example, job finding and labor force participation (Calvo-Armengol and Jackson (2004, 2007), Bayer, Ross, and Topa (2008)); social learning and knowledge diffusion (Conley and Udry (2001), Conley and Udry (2010)); risk sharing and insurance (Fafchamps and Gubert (2007a, 2007b)); obesity transmission (Christakis and Fowler (2007), Fowler and Christakis (2008)); peer effects on students' academic achievement (Calvo-Armengol, Patacchini, and Zenou (2009)); sport and club participation (Bramoullé, Djebbari, and Fortin (2009), Liu, Patacchini, and Zenou (2014)); and juvenile delinquencies or criminal activities (Patacchini and Zenou (2008), Bayer, Hjälmarsson, and Pozen (2009), Ballester, Zenou, and Calvo-Armengol (2010), Patacchini and Zenou (2012)).

<sup>2</sup>A static network refers to a cross-sectional case in which only one observation of a network is available. We focus on a static setting because most widely used social network data are cross-sectional, for example, Add Health data (Udry (2003)) and Indian rural village data (Banerjee, Chandrasekhar, Duflo, and Jackson (2013)). Limited network data with a panel structure can be found in the literature dealing with stochastic actor-based dynamic network modeling; for example, see Snijders (2001) and Snijders, Van De Bunt, and Steglich (2010).

for economic choices that are subject to peer effects, for example, smoking decisions, to impact individuals' utilities of forming network links.

Formally, we present a two-stage game. A network is formed in the first stage, and then in the second stage, individuals choose the intensity of their involvement in some economic activities (e.g., tobacco consumption). We focus on a "subgame perfect" equilibrium; we consider individuals who anticipate the second stage of the game when choosing network links in the first stage.

The formation of the network, that is, the first stage, follows the literature regarding the stability and efficiency of social networks (Jackson and Wolinsky (1996), Dutta and Jackson (2000), Jackson (2005), Caulier, Mauleon, and Vannetelbosch (2015)). Accordingly, we adopt a *transferable* utility framework that allows individuals to make side payments. Indeed, we present a model in which individuals may have preferences over global network features (e.g., popularity, transitive triads, etc.). As such, individuals have a strong incentive to make side payments.

Once a network is formed, individuals choose the intensity of their activities in the second stage. We therefore follow the literature dealing with games on networks (e.g., Ballester, Calvó-Armengol, and Zenou (2006), Calvó-Armengol, Patacchini, and Zenou (2009), Bramoullé, Kranton, and D'amours (2014), Boucher (2016)) and focus on the Nash equilibrium. Specifically, individuals choose the intensity of each activity non-cooperatively, taking the network and the other individuals' choices as given.

The advantages of modeling both network formation and network interactions using an unified framework are twofold: first, we can evaluate the importance of individual incentives on forming friendship that stem from utilities of interactions in activity outcomes. That is, assessing how much individuals anticipate that they will be influenced once a network is formed. Second, the use of a jointly coherent model permits controlling for possible friendship selection biases on estimating peer effects on each activity.

A simple empirical approach to model network formation is to assume pairwise independence: the probability of forming each link is independent of other links, conditional on observables.<sup>3</sup> For example, Fafchamps and Gubert (2007a, 2007b) and Comola (2007) assume pairwise independence and focus on individual and dyad-specific variables to explain the formation of a network. Other examples include latent position models (Hoff, Raftery, and Handcock (2002), Handcock, Raftery, and Tantrum (2007), McCormick and Zheng (2015)) and models with unobserved (network) degree heterogeneity (Graham (2017), Jochmans (2018)). In these models, individuals are assumed to have unobserved positions (or fixed effects) in the network that reflect heterogeneity of their social or economic status. Those unobserved positions also allow researchers to control the homophily effect in terms of unobserved individual characteristics.

However, as noted by Bramoullé and Fortin (2009), pairwise independence is a strong assumption since it implies that individuals' utility functions are additively separable across links. Therefore, even if such models are flexible enough to replicate many network statistics observed in real data, their microeconomic foundations involve strong assumptions.

---

<sup>3</sup>Under pairwise independence, the likelihood of the entire network conditional on unobservables is the product of likelihoods from all pairs of agents.

In this paper, we go beyond pairwise independence specification and consider the exponential probability distribution to model network data. The idea is to treat an observed network as one of many possible configurations for links (represented by one or zero) among a population of  $m$  individuals. This idea matches the Exponential Random Graph model (ERGM) proposed by Frank and Strauss (1986) or, more generally, the  $p^*$  model of Wasserman and Pattison (1996) in the statistical literature.

In either an ERGM or a  $p^*$  model, several selected network statistics, such as the number of reciprocal links, the number of  $k$ -stars,  $k \geq 2$ , and the number of triangles, are specified using an exponential probability distribution as a way to measure how likely those network structures would appear in a network.<sup>4</sup> However, the coefficients of those network statistics in ERGMs and  $p^*$  models do not allow for causal interpretations.

Contrary to the standard literature on ERGMs, we motivate our model specification using a formal economic model where the probability of the observed network is given by the shape of the unique equilibrium of the game. Meanwhile, we show that previous methods used for controlling unobserved individual heterogeneity through latent variables (Hsieh and Lee (2016)) can still be used in our context. As a result, our proposed network formation model handles three distinguished features: observed and unobserved individual heterogeneity, global network dependence, and utilities of interactions from endogenous economic activities as incentives in link decisions. To our knowledge, this is the first paper to do so.

The drawback of using a very general and flexible specification is that it complicates the estimation. Indeed, the likelihood function of an ERGM involves an intractable normalizing term in the denominator, which requires the evaluation of the network statistics for all possible network realizations. To handle the intractable normalizing term during the estimation, many suggestions have been proposed; they include, for example, using simulations in a classical estimation setting (Geyer and Thompson (1992), Snijders (2002)) or in a Bayesian setting with auxiliary Markov chain Monte Carlo (MCMC) (Liang (2010), Murray, Ghahramani, and MacKay (2006), Mele (2017b)).

Due to its numerical efficiency, in this paper we adopt a Bayesian method based on a *double Metropolis–Hastings* (M–H) algorithm (Liang (2010)) to deal with the intractable normalizing term. We also implement the modification of the double M–H proposed by Mele (2017b) to improve convergence of the algorithm. We conduct an extensive simulation study to show that the proposed Bayesian MCMC sampler can successfully recover true model parameters from artificially generated network data. We also examine model misspecification issues in the simulation and provide new evidence of network endogeneity biases within network interaction studies.

We apply our model to the study of American high school students' friendship networks using the Add Health data. We focus on two activity variables: students' GPA and smoking habits. We find a significant impact of a student's GPA on the formation of the network, but we observe no effect from their smoking habits. However, we find peer effects for both activities. This suggests that the interaction in academic learning is a factor for building friendships, whereas the interaction in smoking is not.

<sup>4</sup>See Wasserman and Pattison (1996), Snijders (2002), and Snijders, Pattison, Robins, and Handcock (2006) for more complete lists of network statistics used in ERGMs.

Our results also reveal significant homophily effects from both the observed and unobserved characteristics in network formation. Unobserved characteristics in network formation have significant influence on activity outcomes. That is, peer effects on GPA and smoking are subject to selection biases due to unobserved characteristics linked to the formation of friendship relations.

This paper contributes mainly to two strands of the literature. First, it contributes to the empirical literature on network formation. [Graham \(2017\)](#), [Jochmans \(2018\)](#), and [Dzemski \(2019\)](#) introduced node specific parameters to capture degree heterogeneity in a pairwise independent link formation model; however, they ignore any possible network externality effect. We capture the network externality effects in our empirical model by allowing for global network effects such as popularity, transitive triads, etc.

[Sheng \(2014\)](#), [Miyauchi \(2016\)](#), and [De Paula, Richards-Shubik, and Tamer \(2018\)](#) specified strategic network formation models and characterize equilibria of their models by the pairwise stability condition ([Jackson and Wolinsky \(1996\)](#)). Instead of imposing equilibrium selection assumptions, they specify incomplete models and utilize a partial identification approach. We show that our model is complete as it corresponds to the unique equilibrium of our two-stage game. Importantly, the structure of the model gives us more flexibility in including relevant observed and unobserved individual characteristics, and network features.

[Christakis, Fowler, Imbens, and Kalyanaraman \(2010\)](#) and [Mele \(2017b\)](#) modeled network formation as a sequential process where in each period a single, randomly selected pair of agents has the opportunity to meet and decide to form or sever a link. This sequential process is equivalent to an equilibrium selection mechanism in the corresponding static model ([Jackson and Watts \(2002\)](#)). In contrast, our equilibrium concept, which allows for side payments, leads to a static random utility model.

More specifically, we contribute to the empirical literature on ERGMs of network formation (e.g., [Chandrasekhar and Jackson \(2014\)](#), [Boucher and Mourifié \(2017\)](#), [Mele \(2017b\)](#), [Mele and Zhu \(2017\)](#), [Mele \(2017c\)](#)). With an exception of [Mele \(2017c\)](#), the literature assumes that econometricians observe all of the payoff relevant variables. We contribute to the literature by allowing for unobserved heterogeneity using latent variables, following the strategy used in [Hsieh and Lee \(2016\)](#) and [Hsieh and Van Kippersluis \(2018\)](#). Also, our transferable utility setting allows us to study a wider range of preferences. In particular, we do not require the existence of a potential function.

Second, this paper contributes to the literature on peer effects in endogenous networks (e.g., [Goldsmith-Pinkham and Imbens \(2013\)](#), [Hsieh and Lee \(2016\)](#)). In particular, this paper contributes to the emerging empirical literature studying the impact of economic actions on the formation of a network.<sup>5</sup> To our knowledge, the only five existing papers that deal with this aspect are [Badev \(2013\)](#), [Boucher \(2016\)](#), [Lewis-Faupel \(2016\)](#), [Battaglini, Patacchini, and Rainone \(2019\)](#), and [Hsieh, König, and Liu \(2019\)](#).

[Badev \(2013\)](#) focused on a setup where individuals take a single binary action (affecting the preferences on the network structure). He presents a random utility model

<sup>5</sup>There is a consequent theoretical literature. We refer the interested reader to [Boucher \(2016\)](#) for a discussion and additional references.



and an original equilibrium concept—based on stability constraints—that nests a pairwise stable and pairwise Nash network. The equilibrium also follows an ERGM in which all payoff relevant variables are assumed to be observed and where existence is only guaranteed for potential games. Lewis-Faupel (2016) also focused on a setup where individuals take a single binary action, but this binary action is anticipated in the network formation stage, where individuals form heuristic expectations.

Boucher (2016) presented a model of conformism with respect to a single continuous action. Since the model features many equilibria, he assumes that the data is generated by the equilibrium that maximizes the potential function. Solving for such an equilibrium is not feasible, thus the estimation is performed using the maximum of first- and second-order approximations of the potential function. Hsieh, König, and Liu (2019) used a similar framework to Badev (2013) while considering a continuous action under the context of firm R&D collaboration network and R&D expenditure. The microfoundation of their model is based on the stochastic best response dynamics (Blume (1993)) and the equilibrium is also only guaranteed for potential games.

Battaglini, Patacchini, and Rainone (2019) studied the interaction between the intensity of a legislator's social connections and their legislative effectiveness. In particular, they allow for varying connections' strengths and present a novel equilibrium notion. They show that social connection are an important driver of legislators' effectiveness.

In contrast to this literature, the model presented in this paper simultaneously studies multiple activities. We also control for a wide range of unobserved heterogeneity and allow for a large class of preferences. In particular, characterization of our model equilibrium does not require the existence of a potential function.

The remainder of this paper is structured as follows. Section 2 presents a unified modeling approach for both network formation and network interactions on economic activities. A Bayesian estimation method for the proposed model is discussed in Section 3. Section 4 provides an application of the model to high school students' friendship networks and activities with the Add Health data. Section 5 concludes the paper. Appendix A provides a proof of the main proposition in this paper. Some technical details of estimation, a Monte Carlo simulation, and additional empirical results are relegated to the Online Supplementary Appendix (Hsieh, Lee, and Boucher (2020)).

## 2. ECONOMIC MODEL OF PEER EFFECTS IN AN ENDOGENOUS NETWORK

### 2.1 *Description of the economy*

Let  $W$  be an  $m \times m$  matrix representing the friendship network of  $m$  individuals.<sup>6</sup> The  $(i, j)$ th entry of  $W$ , denoted as  $w_{ij}$ , equals one if individual  $i$  has a link to individual  $j$  and zero otherwise. The notation  $w_i$  stands for the  $i$ th row of  $W$  and  $W_{-i}$  stands for  $W$  excluding  $w_i$ . We assume that the links are *directed*, so it is possible that  $i$  has a link to  $j$ , while  $j$  is not linked to  $i$  (i.e.,  $W$  is not symmetric).<sup>7</sup> We normalize diagonal elements

<sup>6</sup>It is called spatial weights matrix, adjacency matrix, or sociomatrix in the literature.

<sup>7</sup>This nonreciprocity is motivated by our empirical application. In fact, 54.36% of friendship links in our dataset (Add Health) are nonreciprocal. This assumption is also present in Hsieh and Lee (2016), Mele (2017b), Jochmans (2018), and others.

so that  $w_{ii} = 0$  for all  $i \in (1, \dots, m)$ . Importantly, we assume that individuals can form at most  $\bar{n}$  friendship relations. This assumption may reflect, for example, time constraint (e.g., Boucher (2015) and De Paula, Richards-Shubik, and Tamer (2018)).

Let  $X$  be the  $m \times k$  matrix of individual characteristics and  $x_i$  be the  $i$ th row of  $X$ . Individuals choose the intensity of their involvement in economic activities. For each activity  $d \in (1, \dots, \bar{d})$ , let  $y_{i,d}$  denote the intensity of individual  $i$  in activity  $d$ . Let also  $Y_d = (y_{1,d}, \dots, y_{m,d})'$  be the  $m$ -dimensional column vector of intensities for activity  $d$  and  $Y_{-i,d}$  be the  $m-1$  dimensional vector with  $y_{i,d}$  removed from  $Y_d$ . We assume that  $y_{i,d}$  is continuous, or continuous on strictly positive values, but we allow for left censoring at zero. Formally, we consider either  $y_{i,d} \in \mathbb{R}$  or  $y_{i,d} \in \mathbb{R}_+$ .<sup>8</sup> The impact of left censoring on equilibrium activity intensity is formally described in Proposition 1 in Section 2.3.

## 2.2 Preferences

As discussed above, we focus on a transferable utility framework in which individuals are allowed to make side payments. How those side payments are implemented is described in Section 2.3. In this section, we describe the individuals' preferences before side payments are allowed. The preference of any given individual  $i$  is represented by the utility function:

$$U_i(W, Y_1, \dots, Y_{\bar{d}}) = v_i(W) + \sum_{d=1}^{\bar{d}} \delta_d u_{i,d}(y_{i,d}, Y_{-i,d}, W), \quad (1)$$

where  $v_i(W)$  represents an explicit preference over the structure of network  $W$ , while  $u_{i,d}(y_{i,d}, Y_{-i,d}, W)$  is the utility derived from activity  $d$  when individuals are choosing  $Y_d = (y_{i,d}, Y_{-i,d})$  and the network structure is given by  $W$ . The coefficient  $\delta_d \geq 0$  captures the relative importance (or weight) of the utility of activity  $d$  with respect to the utility of the network  $v_i(W)$ . We call this the *incentive effect* of activity  $d$  on network formation.<sup>9</sup>

It is worth noting that, conditional on  $W$ , the utility in equation (1) is separable across activities, so we assume no complementarity across activities. This assumption is quite common in the literature, with the notable exception of Cohen-Cole, Liu, and Zenou (2018).<sup>10</sup> That being said, since the network structure is endogenously determined in our context, the optimal choice for each activity will, at equilibrium, be a function not only of the individual's preference for this particular activity but also of preferences for the other activities, and their explicit network preferences.<sup>11</sup> Indeed, the equilibrium value for  $(W, Y_1, \dots, Y_{\bar{d}})$  is the result of a complex interplay between individuals' preference over the network structure and their preferences regarding each

<sup>8</sup>We abstract from the activity outcome of discrete choices in this paper as it generally involves multiple equilibria (Krauth (2006), Soetevent and Kooreman (2007)). This is left for future research.

<sup>9</sup>Since this is a two-stage game (see Section 2.3), the coefficient  $\delta_d$  can also be interpreted as an activity specific discount factor.

<sup>10</sup>While our model could be extended to such a setting, it would raise nontrivial identification issues (Cohen-Cole, Liu, and Zenou (2018)), which are left for future research.

<sup>11</sup>We also assume that the unobserved part (for the econometrician) of the utility functions for each activity may be correlated. See details in Section 4.



type of activity. This has important consequences for estimation, as will be discussed in Section 4.

For tractability, we follow the literature (e.g., Ballester, Calvó-Armengol, and Zenou (2006), Calvó-Armengol, Patacchini, and Zenou (2009), Bramoullé, Kranton, and D'amours (2014), Boucher (2016)) and assume a linear quadratic specification for the utility of activity conditional on network structure:

$$u_{i,d}(y_{i,d}, Y_{-i,d}, W) = \mu_{i,d}y_{i,d} - \frac{1}{2}y_{i,d}^2 + \lambda_d y_{i,d} \sum_{j=1}^m w_{ij}y_{j,d}, \quad (2)$$

where  $\mu_{i,d}$  captures individual exogenous heterogeneity. The first and second terms of equation (2) represent the private benefit and cost of increasing the intensity of the activity  $y_{i,d}$ . The third term reflects an additional social benefit (or cost) of increasing the intensity of the activity for the individual, that is, a complementary (or competitive) effect from peers' activity intensities whenever  $\lambda_d \geq 0$  ( $\lambda_d < 0$ ).

We assume that individual  $i$ 's (explicit) preference for the network structure is given by

$$v_i(W) = \underbrace{\sum_{j=1}^m w_{ij}\psi_{ij}}_{\text{local network effects}} + \underbrace{\varpi_i(w_i, W_{-i})\eta}_{\text{global network effects}} + \tau_{i,W}, \quad (3)$$

where  $\tau_{i,W}$  is an idiosyncratic shock on the value of the network  $W$  for individual  $i$ . In equation (3), the local network effects captured by  $\psi_{ij}$  give the intrinsic bilateral value (for individual  $i$ ) of a link between  $i$  and  $j$ . This value is assumed to be independent from the position of the individual in the network. However, as argued by Bramoullé and Fortin (2009), individuals may also have preferences over the entire network structure. These preferences are captured by the global network effects  $\varpi_i(w_i, W_{-i})\eta$  and allow for preferences regarding many features of the networks (e.g., popularity, clustering, etc.). Specifically,  $\varpi_i(w_i, W_{-i})$  is an  $\bar{h}$ -dimensional row vector of network statistics that are relevant to individual  $i$ 's utility, and  $\eta$  is the corresponding vector of coefficients. Note that by considering these global network effects, our network model differs substantially from the pairwise independent network link case (Bramoullé and Fortin (2009)) and connects to ERGMs in the statistical literature. The empirical specification of global network effects used in this paper is discussed in Section 2.4. We now discuss our equilibrium concepts for  $(W, Y_1, \dots, Y_{\bar{d}})$  and the timing of the game.

### 2.3 Game

The game occurs in two stages. In the first stage, network links are determined. In the second stage, individuals play a noncooperative game for the choice of the activity intensities  $(Y_1, \dots, Y_{\bar{d}})$ , conditional on the network structure.

Our equilibrium (stability) concept for the first stage of the game is based on the literature focused on the stability and efficiency of network formation games (e.g., Jackson and Wolinsky (1996), Dutta and Jackson (2000), Jackson (2005), Caulier, Mauleon,



FIGURE 1. Example of a simple global network effect.

and Vannetelbosch (2015)). In this *transferable* utility setting, individuals are allowed to make side payments. For example, individuals may be willing to spend time or resources so that other individuals want to be linked to them. Although these side payments are not observed, they play an important role in the efficiency of the equilibrium network.

The focus on a transferable utility framework in this paper contrasts with the existing economic literature based on ERGMs (i.e., as in equation (6) below). Indeed, the usual microeconomic foundation is based on Christakis et al. (2010), Badev (2013), and Mele (2017b) where individuals are assumed to have the opportunity periodically to meet another and revise their friendship status (in a nontransferable utility framework). Importantly, it is assumed that the revision of that friendship relation is done myopically, taking the rest of the network structure as given. The meeting process runs through time, and it is assumed that the observed network is drawn from its steady state distribution.

In contrast, we focus on a static random utility model and do not assume any specific meeting process.<sup>12</sup> This implicitly allows for a richer variety of meeting processes. Under the assumptions of Section 2.2, individuals' preferences not only depend on which friends they have, but also on the the global network effects (see equation (3)). Consequently, we might expect individuals to be willing to spend resources to promote certain friendship relations.

To clarify the intuition, consider a very simple example of a population composed of only three individuals ( $i$ ,  $j$ , and  $k$ ). Assume that the global network effect for  $i$  is given by  $\varpi_i(w_i, W_{-i})\eta = w_{ij}w_{kj}\eta > 0$  so that for  $i$  the value of a link with  $j$  is greater when there exists a link between  $k$  and  $j$ , as Figure 1 illustrates.

Individual  $i$  would therefore be willing to spend resources to compensate individual  $k$  for creating a link with  $j$  (provided they would not do so otherwise). Of course, this is just a simple example, but the general intuition is the same: global network effects introduce strong incentives for side payments, irrespective of the specification used.

Then, following the literature (see above), we focus on the set of networks that are both efficient and individually stable (or rational)—allowing for side payments—where efficiency is defined with respect to the *network value* denoted by  $T(W)$ . Although this quantity can be defined in many ways, we assume (as in Dutta and Jackson (2000)) that the network value is given by the sum of the individuals' utilities, that is,  $T(W) = \sum_i U_i(W)$ .<sup>13</sup> This definition allows to see side payments as being made in “utility units.” We focus on *strongly efficient* networks (Dutta and Jackson (2000)), that is, networks  $W^*$  such that  $T(W^*) \geq T(W)$  for all  $W$ .

Whether or not strongly efficient networks are individually stable depends on how the network value is shared among individuals, that is, how side payments are made.

<sup>12</sup>The randomness is due to the preference shocks  $\tau_{i,W}$ , which would be unobservable to an econometrician.

<sup>13</sup>The dependence of  $U_i$  on  $Y$  is omitted on purpose. The formal definition is presented in Definition 1 below.

This is formally described by the *allocation rule*  $\{\Lambda_i(W, T) \mid \sum_i \Lambda_i(W, T) = T(W)\}$ , which is the utility received after the side payments have been paid.<sup>14</sup> If no side payments are allowed, the allocation rule is simply given by the individuals' utility:  $\Lambda_i(W, T) = U_i(W)$ . If there are side payments, the value of the side payments received by  $i$  is simply given by the difference between the allocation rule and the utility:  $\text{sidepayments}_i = \Lambda_i(W, T) - U_i(W)$ . For example, if the value of the network is shared equally among individuals, we get:  $\Lambda_i(W, T) = T(W)/m$ . Note that the balance requirement of  $\sum_i \Lambda_i(W, T) = T(W)$  imposes that no outside resources are used in making side payments. Indeed, using our definition for the network value, we obtain  $\sum_i \Lambda_i(W, T) = \sum_i U_i(W)$ .

Let  $\Omega_{\bar{n}}$  be the set of networks structures such that individuals have at most  $\bar{n}$  friends. We say that a network  $W \in \Omega_{\bar{n}}$  is *individually stable*, relative to  $\{\Lambda_i\}_{i=1}^m$  and  $T$ , if for all  $i$ ,  $\Lambda_i(W, T) \geq \Lambda_i(\tilde{W}, T)$  for all networks  $\tilde{W} \in \Omega_{\bar{n}}$  such that  $\tilde{w}_{jk} = w_{jk}$  for all  $k$  and all  $j \neq i$ .<sup>15</sup> In essence, the network is individually stable if no individual wants to create or remove links, given the side payments. In what follows, we do not make any assumptions regarding the specific allocation rule used. We merely assume that the allocation rule is selected among the (nonempty) set of allocation rules compatible with both strong efficiency and individual stability.

Once the network is formed, individuals are free to select intensities of activities in which they are involved, conditional on network structure. We follow the extensive literature for games on networks (e.g., Ballester, Calvó-Armengol, and Zenou (2006), Calvó-Armengol, Patacchini, and Zenou (2009), Bramoullé, Kranton, and D'amours (2014), Boucher (2016)) and assume that activity intensity choices are part of a Nash equilibrium.<sup>16</sup> Formally, our equilibrium concept for the two-stage game is defined as follows.

**DEFINITION 1.** An (subgame perfect) equilibrium of the two-stage game is a collection  $(W, Y_1, \dots, Y_{\bar{d}})$  such that:

1.  $(Y_1, \dots, Y_{\bar{d}})$  is in a Nash equilibrium, conditional on  $W \in \Omega_{\bar{n}}$ . We denote such an equilibrium by  $(Y_1^*(W), \dots, Y_{\bar{d}}^*(W))$ .
2. The network value

$$T_{Y^*}(W) = \sum_i U_i(W, Y_1^*(W), \dots, Y_{\bar{d}}^*(W))$$

is strongly efficient and individually stable for networks in  $\Omega_{\bar{n}}$  under some allocation rules.

Note that the definition of the value of the network in the first stage of the game (i.e.,  $T_{Y^*}(W)$ ) is given by the sum of the individuals' utilities, anticipating that individuals will play the Nash equilibrium  $Y_d^*(W)$  in the second stage. In this sense, therefore, the equilibrium is subgame perfect since it is solved by backward induction. The next proposition follows.

<sup>14</sup>This definition assumes implicitly that the allocation rule is *balanced*, as in Dutta and Jackson (2000).

<sup>15</sup>This is equivalent to saying that  $W$  is individually stable if  $W$  is a Nash equilibrium of the game where individuals' payoffs are given by the allocation rule, under the network value  $T(W)$ .

<sup>16</sup>That is,  $(Y_1^*, \dots, Y_{\bar{d}}^*)$  such that  $y_{i,d}^* \in \arg \max_{y_{i,d}} U_i(W, Y_1^*, \dots, y_{i,d}, Y_{-i,d}^*, \dots, Y_{\bar{d}}^*)$  for all  $i$  and  $d$ .

PROPOSITION 1. Assume that  $|\lambda_d| < 1/\bar{n}$  for all  $d = 1, \dots, \bar{d}$  and that  $\tau_W \equiv \sum_i \tau_{i,W}$  is distributed according to a Type I extreme value distribution. Then there exists a (generically) unique equilibrium of the two-stage game. Moreover,

(i) for all  $d$ , such that  $y_{i,d} \in \mathbb{R}$ , we have

$$Y_d^*(W) = (I_m - \lambda_d W)^{-1} \mu_d, \quad (4)$$

where  $I_m$  is an  $m \times m$  identity matrix and  $\mu_d = (\mu_{1,d}, \dots, \mu_{m,d})'$ . While for all  $d$  such that  $y_{i,d} \in \mathbb{R}_+$ , the unique equilibrium is determined as

$$y_{i,d}^*(W) = \max \left\{ 0, \mu_{i,d} + \lambda_d \sum_{j=1}^m w_{ij} y_{j,d}^*(W) \right\}. \quad (5)$$

(ii) The equilibrium value is given by  $T_{Y^*}(W) = V(W) + \tau_W$ , where

$$\begin{aligned} V(W) = & \sum_{i=1}^m \sum_{j=1}^m w_{ij} \psi_{ij} + \sum_{i=1}^m \varpi_i(w_i, W_{-i}) \eta \\ & + \sum_{d=1}^{\bar{d}} \delta_d \left[ \mu_d' Y_d^*(W) - \frac{1}{2} Y_d^{*'}(W) Y_d^*(W) + \lambda_d Y_d^{*'}(W) W Y_d^*(W) \right]. \end{aligned}$$

Therefore, the probability of  $W$  at equilibrium is given by

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\tilde{W} \in \Omega_{\bar{n}}} \exp\{V(\tilde{W})\}}. \quad (6)$$

The uniqueness of the Nash equilibrium uses standard arguments (e.g., Ballester, Calvo-Armengol, and Zenou (2006)). Whenever  $|\lambda_d| < 1/\bar{n}$  for all  $d \in (1, \dots, \bar{d})$ , the best response functions are contraction mappings, leading to a unique fixed point.<sup>17</sup> Note that the contracting property also has the important numerical advantage of providing an iterative procedure to solve for the equilibrium when  $y_{i,d} \in \mathbb{R}_+$ . See the proof of Proposition 1 in Appendix A.

Also, since the Nash equilibrium in the second stage is unique, the value of the network is also uniquely determined, that is,  $T_{Y^*}(W) = T(W)$ . Since  $\tau_W \equiv \sum_i \tau_{i,W}$  is distributed according to a Type I extreme value distribution, standard arguments show that the distribution of the maximum of  $T(W)$  follows a logistic form. Brock and Durlauf (2001) used the same assumption when specifying the social welfare function. Indeed, while one might prefer to assume that each  $\tau_{i,W}$  follow a Type I extreme value distribution, this would make the model intractable since the sum of Type I extreme value distributions is not Type I extreme value distributed. Assuming that  $\tau_W$  is distributed according to a Type I extreme value distribution guarantees that the unique strongly efficient network is given by (ii) of Proposition 1.

<sup>17</sup>See the proof of Proposition 1 for a discussion of the strength of this assumption.

An example of an allocation rule for which the strongly efficient network is also individually stable is  $\Lambda_i(W, T) = T/m$ . Of course, this is not the only admissible allocation rule. In particular, any allocation rule that can be written as a nondecreasing function of the network value is individually stable. More generally, it is also possible to impose additional normative properties on the admissible allocation rules. We refer the interested reader to [Dutta and Jackson \(2000\)](#) for further discussions and results. [Jackson \(2005\)](#) also presented an extensive discussion of the type of allocation rules compatible with both strong efficiency and individual stability.

It is worth noting that the expression  $\mu'_d Y_d^*(W) - \frac{1}{2} Y_d^{*'}(W) Y_d^*(W) + \lambda_d Y_d^{*'}(W) \times W Y_d^*(W)$  in the equilibrium value for  $V(W)$  reduces to  $\frac{1}{2} Y_d^{*'}(W) Y_d^*(W)$  when activity  $d$ 's intensity is uncensored since we can exploit the closed-form solution in equation (4). In that case, the incentive effect of activities is always non-negative (recall that  $\delta_d \geq 0$  in equation (1)). Indeed, as noted by [Ballester, Calvó-Armengol, and Zenou \(2006\)](#), this property holds whenever the choice of activities features complementarities.<sup>18</sup> As such, without explicit network preferences, that is,  $v_i(W)$ , individuals would want to have as many links as possible, leading to a regular network in which all individuals have  $\bar{n}$  links. In Section 2.4, we discuss the specific parametric assumptions of  $v_i(W)$  and how they prevent the model from generically producing degenerated network structures.

## 2.4 Parametric specification

We specify individual exogenous heterogeneity in equation (2) via  $\mu_{i,d} = x_i \beta_{1d} + \sum_{j=1}^m w_{ij} x_j \beta_{2d} + \alpha_d + \epsilon_{i,d}$ , where  $\alpha_d$  is a constant term, and  $\epsilon_{i,d}$  is the unobserved heterogeneity of  $i$ 's preference regarding activity  $d$ . Following Proposition 1, the equilibrium for uncensored activities is given by

$$Y_d^*(W) = (I_m - \lambda_d W)^{-1} (X \beta_{1d} + W X \beta_{2d} + l_m \alpha_d + \epsilon_d), \quad (7)$$

where  $l_m$  is the  $m$ -dimensional vector of ones, and  $\epsilon_d = (\epsilon_{1,d}, \dots, \epsilon_{m,d})'$ . Equation (7) matches the reduced form of a spatial autoregressive (SAR) model ([Bramoullé, Djebbari, and Fortin \(2009\)](#), [Lee, Liu, and Lin \(2010\)](#), [Lin \(2010\)](#)) for studying social interactions. The coefficient  $\lambda_d$  in equation (7) represents the endogenous (peer) effect, which has been the focus of recent literature due to its policy implications ([Glaeser, Sacerdote, and Scheinkman \(2003\)](#)). The vector of coefficients  $\beta_d = (\beta'_{1d}, \beta'_{2d})'$  captures the own and contextual effects of individuals' and friends' exogenous characteristics on  $Y_d$ .

However, a notable departure of our model from the literature is that the network structure  $W$  in equation (7) is explicitly endogenous. To understand the source of this endogeneity, recall that from Proposition 1, we have

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\tilde{W} \in \Omega_{\bar{n}}} \exp\{V(\tilde{W})\}}.$$

<sup>18</sup>[Boucher \(2016\)](#) studied a model of conformism having an endogenous network. He shows that conformism has non-monotonic effects on the value of network links.

For the sake of the discussion, assume for the moment that  $y_{i,d} \in \mathbb{R}$ , so that we have

$$V(W) = \sum_{i=1}^m \sum_{j=1}^m w_{ij} \psi_{ij} + \sum_{i=1}^m \varpi_i(w_i, W_{-i}) \eta + \frac{1}{2} \sum_{d=1}^{\bar{d}} \delta_d Y_d^*(W) Y_d^*(W).$$

We see immediately that  $V(W)$  is also a function of  $\epsilon_d$ . Then this implies that any shock  $\epsilon_d$  has three conceptually distinct effects. First, it directly affects  $Y_d^*$ , conditional on  $W$ . Second, through its effect on  $Y_d^*$ , it indirectly affects network structure  $W$  through its effect on  $V(W)$ . Third, through its indirect effect on  $W$ , it also affects  $Y_{\tilde{d}}^*$  for the other activities  $\tilde{d} \neq d$ .

This endogeneity is exacerbated if there exist unobserved variables that are directly affecting the network formation as well as intensities of activities (e.g., Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016)). We also allow for such unobserved variables to affect individuals' preferences over links in our model. Specifically, we follow Hsieh and Lee (2016) to introduce the multidimensional individual latent variables  $z_i = (z_{i1}, \dots, z_{i\bar{\ell}})'$  in the network formation process through the local network effect as follows:

$$\psi_{ij} = \gamma_0 + c_i \gamma_1 + c_j \gamma_2 + c_{ij} \gamma_3 + \sum_{\ell=1}^{\bar{\ell}} \gamma_{4\ell} |z_{i\ell} - z_{j\ell}|. \quad (8)$$

The variables  $c_i$  and  $c_j$  in equation (8) are *observed*  $\bar{s}$ -dimensional row vectors of individual specific characteristics, and the variable  $c_{ij}$  is an *observed*  $\bar{q}$ -dimensional row vector of dyad-specific characteristics, such as whether  $i$  and  $j$  have the same age, sex, or race.<sup>19</sup> In particular, the individual and dyadic characteristics  $C = \{(c_i, c_j, c_{ij}) : i = 1, \dots, m, j = 1, \dots, m, i \neq j\}$  control for observed homophily in the network formation process (see, e.g., Fafchamps and Gubert (2007a, 2007b) in the context of risk sharing network formation). The variables  $|z_{i\ell} - z_{j\ell}|$  for  $\ell = 1, \dots, \bar{\ell}$  in equation (8) are meant to capture the homophily on *unobserved* characteristics. We expect the coefficients  $\gamma_4$ 's to be negative, reflecting the fact that larger differences between individual unobserved characteristics reduce the likelihood that two individuals become friends. For exposition purposes, we denote  $\gamma = (\gamma_0, \gamma_1', \gamma_2', \gamma_3', \gamma_4')'$ .

The latent variables also affect intensities of activities through the unobserved heterogeneity:  $\epsilon_d = Z\rho_{1d} + WZ\rho_{2d} + \xi_d$ , where  $Z = (z_1', \dots, z_m')'$  is a  $m \times \bar{\ell}$  matrix of unobserved (latent) variables. Note that  $Z$  is not specific to any activity  $d$ . Correspondingly, the activity intensity of equation (7) can be rewritten as

$$Y_d^*(W) = (I_m - \lambda_d W)^{-1} (X\beta_{1d} + W X\beta_{2d} + Z\rho_{1d} + W Z\rho_{2d} + l_m \alpha_d + \xi_d), \quad (9)$$

where we assume that  $\xi_d \sim \mathcal{N}(0, \sigma_{\xi_d}^2 I_m)$ . Given the role played by the latent variables  $Z$  in equation (8) for network formation,  $Z$  and  $WZ$  that appear in equation (9) can also be

<sup>19</sup>It is possible to specify  $c_{ij} = |c_i - c_j|$  if  $c_i$  is continuous. For binary  $c_i$ , however, we prefer the use of dummy variables  $c_{ij}$  taking a value of 1 if  $i$  and  $j$  have the same value. Of course, this is fully equivalent to taking the distance, which would take a value of 1 if  $i$  and  $j$  have different values, and 0 otherwise.



interpreted as control functions for solving endogeneity due to individual and contextual unobserved correlated effects (Arcidiacono, Foster, Goodpaster, and Kinsler (2012), Fruehwirth (2014), Hsieh and Van Kippersluis (2018)).<sup>20</sup> Note that latent variables  $Z$  are treated as random effects and, therefore, they should be independent of  $X$  in equation (9). Besides, although both  $WZ$  and  $\alpha_d$  capture the correlated effects due to unobservables, they are differentiated by the fact that  $WZ$  reflect individual variations while  $\alpha_d$  is a constant for all group members.

In our empirical application, we consider the following specification of global network effects (see equation (3)):

$$\varpi_i(w_i, W_{-i})\eta = \sum_{j=1}^m w_{ij} \left\{ \underbrace{\eta_1 w_{ji}}_{\text{reciprocity effect}} + \underbrace{\eta_2 \sum_{k \neq j} w_{ik} + \eta_3 \left( \sum_{k \neq j} w_{ik} \right)^2}_{\text{congestion effect}} + \underbrace{\eta_4 \sum_{k \neq i} w_{kj}}_{\text{popularity effect}} + \underbrace{\sum_k (\eta_{51} w_{ik} w_{kj} + \eta_{52} w_{ki} w_{kj} + \eta_{53} w_{ik} w_{jk})}_{\text{transitive triads effects}} + \underbrace{\eta_6 \sum_k w_{jk} w_{ki}}_{\text{three-cycle effect}} \right\}. \quad (10)$$

We discuss the interpretation of each term in equation (10) in turn and provide a visualization of each effect in Figure 2.<sup>21</sup>

The reciprocity effect implies that (provided that  $\eta_1 > 0$ )  $i$  enjoys more utility from a link with  $j$  if  $j$  also has a link with  $i$ . The congestion effect (provided that either  $\eta_2 < 0$  or  $\eta_3 < 0$ ) implies that the value for  $i$  of a link with  $j$  decreases with the number of links that  $i$  has. This may represent the fact that  $i$  has limited resources, for example, limited time, energy, or money (Boucher (2015)). Note that we capture this effect by individual  $i$ 's out-degree and out-degree square to allow this cost to be concave or convex. The popularity effect captures the fact that  $i$  may enjoy more utility (if  $\eta_4 > 0$ ) from their link with  $j$  if  $j$  is popular, that is, receives many links.

The transitive triads effects include preferences for cliques, that is, explicit preferences for the transitivity of the network. When  $i$  is considering a link to  $j$ , he may take into account that he has a link to  $k$ , and  $k$  has a link to  $j$ . Therefore, the creation of a link between  $i$  and  $j$  would close the triad between  $i$ ,  $j$ , and  $k$ . There are, of course, other types of transitive triads effects, displayed at the bottom left of Figure 2.

A similar intuition holds for the three-cycle effect, although as noted by Snijders, Van De Bunt, and Steglich (2010), the emergence of more three cycles in a network implies fewer hierarchical relationships among individuals. Indeed, note that the three cycle effect differs from the transitive triads effects by having a circular links between  $i$ ,  $j$ , and

<sup>20</sup>Fruehwirth (2014) argued that the rationale of including peer activity outcome into the social interactions model is to proxy for the unobserved peer inputs or characteristics. Therefore, even if the model is identified, the estimated effects of peer characteristics from the social interactions model with the endogenous effect is hard to interpret and used for policy suggestion. Thus, an intuitive way to alleviate such a concern is to distinguish the effect of unobserved peer characteristics from the endogenous effect of peer activity outcome by directly specifying  $WZ$  in the model.

<sup>21</sup>We also provide additional discussion in the Online Supplementary Appendix B.

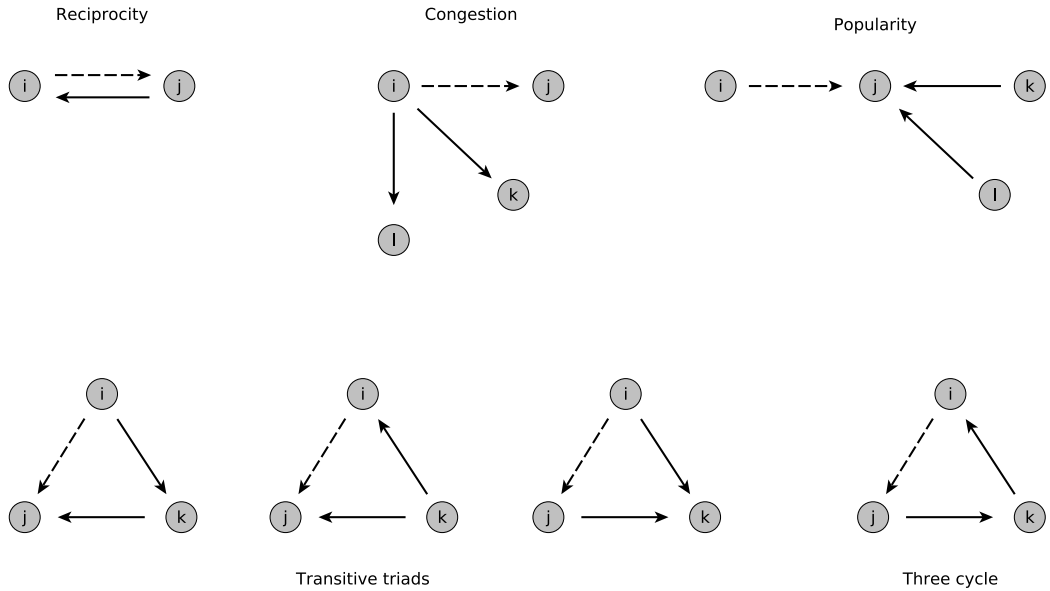


FIGURE 2. Global network effects.

$k$ , see the bottom right of Figure 2. As such, the three cycle is less hierarchical since all individuals have one inward link and one outward link. As discussed by Davis (1970), social networks usually feature local hierarchy, that is, fewer three cycles. We therefore expect  $\eta_6$  to be negative.<sup>22</sup>

Given the parametric assumption in equation (10), we can write

$$\begin{aligned}
 & \sum_{i=1}^m \varpi_i(w_i, W_{-i}) \eta \\
 &= \eta_1 \text{tr}(W^2) + \eta_2 (l'_m W' W l_m - l'_m W l_m) \\
 &+ \eta_3 (l'_m W' \text{Diag}(W l_m) W l_m - 2 l'_m W' W l_m + l'_m W l_m) \\
 &+ \eta_4 (l'_m W W' l_m - l'_m W l_m) + (\eta_{51} + \eta_{52} + \eta_{53}) \text{tr}(W^2 W') + \eta_6 \text{tr}(W^3), \quad (11)
 \end{aligned}$$

where in general for an  $m \times 1$  vector  $A$ ,  $\text{Diag}(A)$  is an  $m \times m$  diagonal matrix with its diagonal elements formed by the entries of the vector  $A$ . One can see that parameters  $\eta_{51}$ ,  $\eta_{52}$ , and  $\eta_{53}$  are not identified separately from equation (11) via the (first stage) game. Hence, we will use  $\eta_5 = \eta_{51} + \eta_{52} + \eta_{53}$  hereafter. We further denote  $\eta = (\eta_1, \dots, \eta_6)'$  for the purposes of exposition.

It is important to note that by including the possible cost captured by the congestion effect, some of the global network effects are expected to produce sufficiently large negative externalities on link formation so that not all individuals will form  $\bar{n}$  links (resulting in a regular graph). Moreover, as discussed by Snijders et al. (2006), Bhamidi et al. (2011), Chatterjee, Diaconis et al. (2013), and Mele (2017b), we need negative externalities in

<sup>22</sup>This is indeed what we find in our empirical study; see Table 2.

ERGMs to produce sparser graphs so that they can be distinguished from Erdős–Rényi random graphs when the number of individuals increases. In our case,  $\eta_2$  (or  $\eta_3$ ) and  $\eta_6$  are expected to create such negative externalities, and we confirm that these parameter estimates are significantly negative in our empirical study in Section 4. Lastly, in order to summarize our empirical model and provide an intuitive illustration, we use Figure 3 to present the model with two activities in which variables and interactions between them are displayed.

### 3. MODEL ESTIMATION

#### 3.1 Group heterogeneity

Prior to this section, we assumed that individuals belong to one, potentially large population. However in many contexts—such as the high school students in our empirical application—the population can be partitioned into groups, such that individuals can only form links within each group. In this context, it is important to capture the heterogeneity across these groups. We therefore expand the model assuming that the population is partitioned into  $G$  groups, and we use the subscript  $g \in \{1, \dots, G\}$  to indicate explicit group heterogeneity.

#### 3.2 Likelihood function of the model

To clarify the intuition of the estimation procedure—and for the purposes of exposition—we assume that  $\bar{d} = 1$  and that  $y_{i,d} = y_i$  is uncensored. Accordingly, we drop the subscript  $d$  for clarity.<sup>23</sup> Although our preferred specification for the empirical application is for two activities (one censored and one uncensored), its formal description involves additional notations and steps but not any new conceptual issue. We refer the interested reader to Online Supplementary Appendix D for details.

Since both  $Y_g$  and  $W_g$  are endogenous variables, we focus on the joint likelihood, that is,<sup>24</sup>

$$P(W_g, Y_g | \theta_g, \alpha_g, Z_g) = P(Y_g | W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g | \theta_g, \alpha_g, Z_g),$$

where  $\theta_g = (\gamma', \eta', \delta, \lambda, \beta', \rho', \sigma_{\xi_g}^2)$ . Note that to describe group heterogeneity on activity outcomes at both the mean and variance levels, we use  $\alpha_g$  and  $\sigma_{\xi_g}^2$  to capture respectively the group fixed effect and group heteroskedasticity in equation (9). To adhere to the principle of model parsimony, the other coefficients are assumed to be common across groups. Using the parametric forms assumed in Section 2.4, we can write the joint mixed density function of  $W_g$  and  $Y_g$  conditional on the latent variable  $Z_g$  as

$$P(W_g, Y_g | \theta_g, \alpha_g, Z_g)$$

<sup>23</sup>In the Online Supplementary Appendix C, we present the model likelihood function for the censored activity outcome case.

<sup>24</sup>Of course, assuming independence across groups, the joint likelihood (over all groups) can be written as  $\prod_{g=1}^G P(W_g, Y_g | \theta_g, \alpha_g, Z_g)$ .

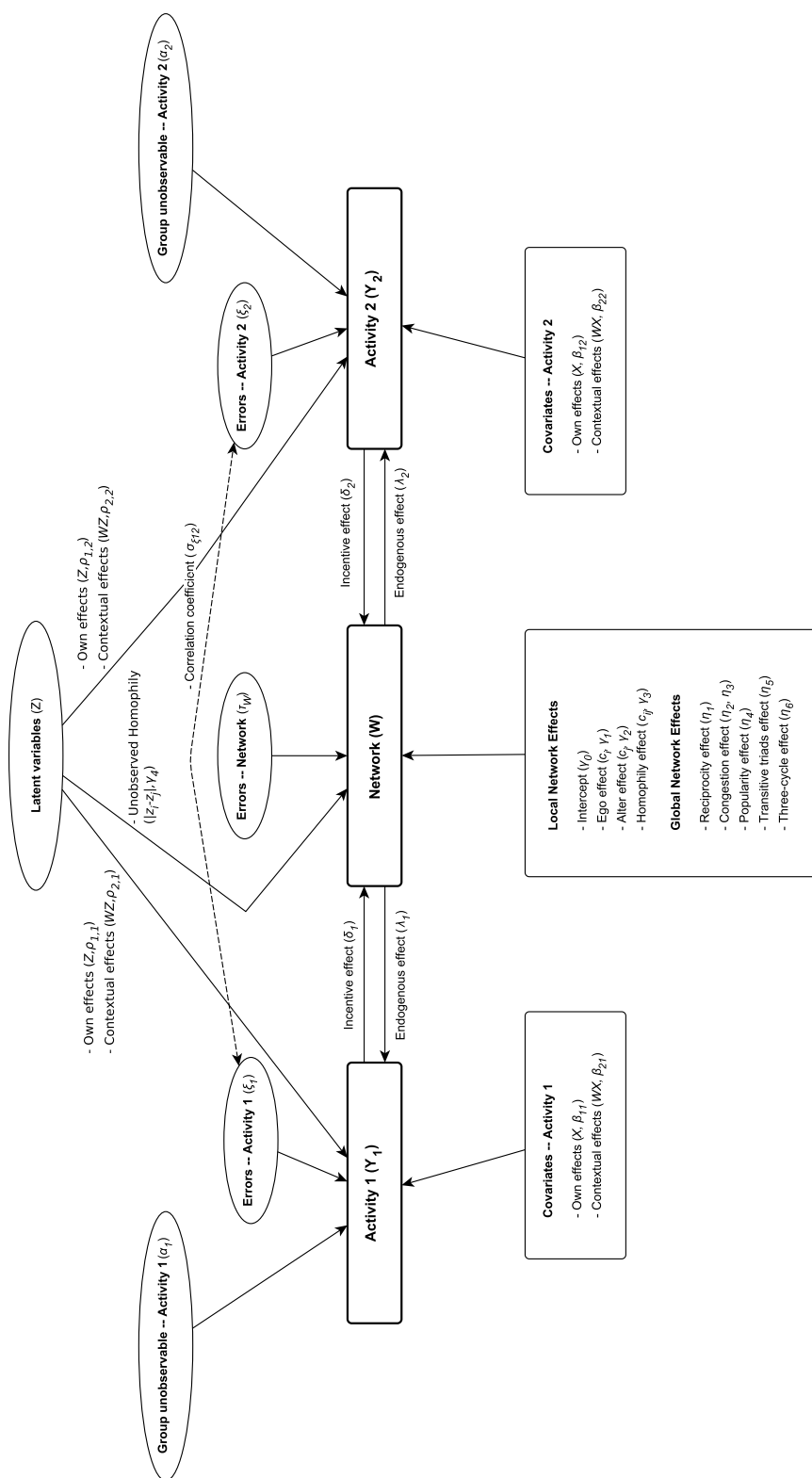


FIGURE 3. Summary of the empirical model. *Note:* The figure presents a summary of the empirical model with two activities. Observed variables are represented by rectangles, while unobserved variables are represented by ellipses. The three endogenous/chosen variables ( $Y_1, Y_2, W$ ) are displayed in the center of the figure. Note that omitting a variable has an impact on the variables directly and indirectly connected to it.

$$\begin{aligned}
&= P(Y_g|W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g, W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|\theta_g, \alpha_g, Z_g) \cdot P(W_g|\xi_g, \theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|\theta_g, \alpha_g, Z_g) \cdot \frac{\exp(V(W_g, \xi_g, \theta_g, \alpha_g, Z_g))}{\sum_{\tilde{W}_g \in \Omega_{n,g}} \exp(V(\tilde{W}_g, \xi_g, \theta_g, \alpha_g, Z_g))}, \tag{12}
\end{aligned}$$

where  $S_g(W_g) = I_{m_g} - \lambda W_g$ ,  $\xi_g = S_g(W_g)Y_g - X_g\beta_1 - W_gX_g\beta_2 - Z_g\rho_1 - W_gZ_g\rho_2 - l_{m_g}\alpha_g$ , and

$$f(\xi_g|\theta_g, \alpha_g, Z_g) = (2\pi)^{-\frac{m_g}{2}} (\sigma_{\xi_g}^2)^{-\frac{m_g}{2}} \exp\left(-\frac{1}{2\sigma_{\xi_g}^2} \xi_g' \xi_g\right).$$

The probability of  $W_g$  in equation (12) is the logit probability from Proposition 1. As  $Z_g$  consists of latent variables, the joint density of the observed  $(W_g, Y_g)$  would be  $P(W_g, Y_g|\theta_g, \alpha_g) = \int_{Z_g} P(W_g, Y_g|\theta_g, \alpha_g, Z_g) dF(Z_g)$ , where  $F(\cdot)$  is the joint distribution of  $Z_g$ .

Before discussing the estimation strategy, we would like to briefly comment on the identification of the model. As discussed below, most of these formal identification results follow the literature. The intuition is also discussed in Figure 3.

First, as seen from the likelihood function of equation (12), the vast majority of parameters in our model are identified from  $P(W_g|\theta_g, \alpha_g, Z_g)$ , which is an ERGM. Intuitively, our model is identified under the *many networks asymptotics* (Mele (2017b)), which assumes that the number of groups grows with the sample size. In such a case, the identification of the parameters is standard and follows the theory for exponential families under the usual regularity conditions as in Lehmann and Casella (2006).

The additional identification issue beyond ERGM is linked to the latent variables  $Z$  in equations (8) and (9). We follow Hsieh and Lee (2016) and Hsieh and Van Kippersluis (2018) to impose some assumptions on  $z_{i,\ell}$ : (1) the variance of  $z_{i,\ell}$  is normalized to one; (2)  $z_{i,\ell}$  is independent across  $i$  and  $\ell$ ; (3)  $z_{i,\ell}$  follows a known distribution, in our case a normal distribution; (4) to distinguish different dimension of  $z_{i,\ell}$ , we further restrict  $|\gamma_{41}| \geq |\gamma_{42}| \geq \dots \geq |\gamma_{4\bar{\ell}}|$  in equation (8).<sup>25</sup>

Also note that while the individual specific variables  $c_i$  in the local network effect of equation (8) can overlap with  $x_i$  in equation (7), the dyad-specific variables  $c_{ij}$  used in equation (8) are naturally excluded from equation (7) because of mismatched dimensions. Thus, the model satisfies the exclusion restriction condition in identifying the effects of  $Z$  and  $WZ$  in equation (9).

The main challenge in estimating this model is to compute the denominator of a network probability in equation (12). As it sums all possible network structures, its direct evaluation is impossible, even for small sized networks. For example, in a network

<sup>25</sup>The justification of these identification restrictions can be found in the Online Supplementary Appendix of Hsieh and Van Kippersluis (2018).

with only five individuals in which individuals can form up to five friends, the number of possible network structures is  $2^{5(5-1)} = 2^{20}$ . Hence, any estimation method involving a direct likelihood evaluation is not feasible. This problem applies to all ERGMs for networks (e.g., [Badev \(2013\)](#), [Chandrasekhar and Jackson \(2014\)](#), [Mele \(2017b\)](#), [Boucher and Mourifié \(2017\)](#)), and that can be traced back to the spatial analysis in [Besag \(1974\)](#).

In this paper, we implement a Bayesian estimation approach using an effective MCMC technique developed to handle the intractable normalizing term in the posterior density function (e.g., [Mele \(2017b\)](#)).<sup>26</sup> We start by reviewing the intuition behind the general technique.

### 3.3 General intuition: Double M–H algorithm

To clarify the intuition, we will (abusively) use the following simplifying notation: for any variable  $A_g$ , we use the notation  $\{A_g\}$  to represent the collection of variables  $A_g$  across  $G$  groups, that is,  $\{A_g\} := (A_1, \dots, A_G)$ . Let  $y = (\{Y_g\}, \{W_g\})$  and  $\vartheta = (\{\theta_g\}, \{\alpha_g\}, \{Z_g\})$ . From Section 3.2, the likelihood function of  $y$ , given  $\vartheta$ , has the following form:  $P(y|\vartheta) = f(y; \vartheta)/D(\vartheta)$ , where  $D(\vartheta)$  is an intractable normalizing term.

The standard M–H algorithm to simulate random draws of  $\vartheta$  operates as follows: given an old draw  $\vartheta_{\text{old}}$ , one proposes a new draw  $\vartheta_{\text{new}}$  from a proposal distribution  $q(\cdot|\vartheta_{\text{old}})$ , and then updates the old draw to the new draw with an acceptance ratio  $\alpha_{\text{MH}}(\vartheta_{\text{new}}, \vartheta_{\text{old}})$  given by

$$\begin{aligned} \alpha_{\text{MH}}(\vartheta_{\text{new}}, \vartheta_{\text{old}}) &= \min \left\{ 1, \frac{P(\vartheta_{\text{new}}|y)q(\vartheta_{\text{old}}|\vartheta_{\text{new}})}{P(\vartheta_{\text{old}}|y)q(\vartheta_{\text{new}}|\vartheta_{\text{old}})} \right\} \\ &= \min \left\{ 1, \frac{\pi(\vartheta_{\text{new}})f(y; \vartheta_{\text{new}})q(\vartheta_{\text{old}}|\vartheta_{\text{new}})}{\pi(\vartheta_{\text{old}})f(y; \vartheta_{\text{old}})q(\vartheta_{\text{new}}|\vartheta_{\text{old}})} \cdot \frac{D(\vartheta_{\text{old}})}{D(\vartheta_{\text{new}})} \right\}, \end{aligned} \quad (13)$$

where  $\pi(\vartheta)$  is the prior distribution of  $\vartheta$ . One can see that in equation (13), the normalizing terms  $D(\vartheta_{\text{old}})$  and  $D(\vartheta_{\text{new}})$  do not cancel out; thus the evaluation of the acceptance-rejection criterion in equation (13) is intractable.

To solve this problem, [Murray, Ghahramani, and MacKay \(2006\)](#) consider including auxiliary variables  $\tilde{y}$  into the acceptance probability, that is, the acceptance probability can be written as

$$\begin{aligned} \alpha_{\text{MH}}(\tilde{y}, \vartheta_{\text{new}}, \vartheta_{\text{old}}) &= \min \left\{ 1, \frac{\pi(\vartheta_{\text{new}})P(y|\vartheta_{\text{new}})q(\vartheta_{\text{old}}|\vartheta_{\text{new}})}{\pi(\vartheta_{\text{old}})P(y|\vartheta_{\text{old}})q(\vartheta_{\text{new}}|\vartheta_{\text{old}})} \cdot \frac{P(\tilde{y}|\vartheta_{\text{old}})}{P(\tilde{y}|\vartheta_{\text{new}})} \right\} \\ &= \min \left\{ 1, \frac{\pi(\vartheta_{\text{new}})f(y; \vartheta_{\text{new}})q(\vartheta_{\text{old}}|\vartheta_{\text{new}})}{\pi(\vartheta_{\text{old}})f(y; \vartheta_{\text{old}})q(\vartheta_{\text{new}}|\vartheta_{\text{old}})} \cdot \frac{f(\tilde{y}; \vartheta_{\text{old}})}{f(\tilde{y}; \vartheta_{\text{new}})} \right\}, \end{aligned} \quad (14)$$

where  $\tilde{y}$  are simulated from the likelihood function  $P(\tilde{y}|\vartheta_{\text{new}}) = f(\tilde{y}; \vartheta_{\text{new}})/D(\vartheta_{\text{new}})$  with the exact sampling ([Propp and Wilson \(1996\)](#)).

<sup>26</sup>Alternative estimation approaches include the maximum pseudo likelihood approach ([Besag \(1974\)](#), [Strauss and Ikeda \(1990\)](#), [Boucher and Mourifié \(2017\)](#)) and Monte Carlo maximum likelihood ([Geyer and Thompson \(1992\)](#)).



In the conditional acceptance probability of equation (14), all normalizing terms cancel out, and the remaining terms are computable. This algorithm is called the “exchange algorithm” because a swapping operation between  $(\vartheta_{\text{old}}, y)$  and  $(\vartheta_{\text{new}}, \tilde{y})$  is involved (Geyer (1991)). The exchange algorithm differs from the conventional M–H algorithm by adding a randomization component into the proposal density; this changes  $q(\vartheta_{\text{new}}|\vartheta_{\text{old}})$  into  $q(\vartheta_{\text{new}}|\vartheta_{\text{old}})P(\tilde{y}|\vartheta_{\text{new}})$ .

The exchange algorithm defines a valid Markov chain for simulating from  $P(\vartheta|y)$  (Murray, Ghahramani, and MacKay (2006), Liang (2010), Liang, Jin, Song, and Liu (2016)). However, implementing the exchange algorithm is time consuming because it requires the exact sampling of  $\tilde{y}$  from  $P(\tilde{y}|\vartheta_{\text{new}})$ . To save computation time, Liang (2010) proposed a “double M–H algorithm” that utilizes the reversibility condition and shows that when  $\tilde{y}$  is simulated by the M–H algorithm—starting from  $y$  with  $R$  iterations—the conditional acceptance probability in equation (14) can be obtained regardless of the value of  $R$ . This gives the double M–H algorithm an advantage as a small value of  $R$  can be used, thereby removing the need for exact sampling.

Also note that Mele (2017b) suggests similarly the use of the double M–H algorithm in estimating ERGMs; however, he improves the convergence of the double M–H algorithm further by mixing the conventional random walk proposal with other proposals, such as random block techniques (Chib and Ramamurthy (2010)), to improve the mixing and convergence of the network simulation.<sup>27</sup> With this mixed proposal, Mele (2017b) showed that simulation of the network can escape from the local maxima in the “low temperature regime” of the ERGM (Bhamidi et al. (2011)) where the mixing is problematic. Given this greater computational efficiency compared to exact sampling, we adopt the double M–H algorithm, combined with Mele (2017b)’s improvement for network simulation.

We also provide a technical contribution for the computation of the double M–H algorithm for our model. Using the double M–H algorithm to update  $\vartheta$  from  $P(\vartheta|y)$  requires simulating the auxiliary variable  $\tilde{y}$ . In our context, however, the auxiliary activity variables  $\{\tilde{Y}_g\}$  in  $\tilde{y}$  are redundant during simulation as they can be replaced by either a closed-form function (as in equation (4)) or a contraction mapping (as in equation (5)) of auxiliary networks and estimated individual heterogeneity. Thus, we can simplify  $\tilde{y}$  to  $\tilde{w} = \{\tilde{W}_g\}$  and modify the conditional acceptance probability in equation (14) to

$$\begin{aligned} \alpha_{\text{MH}}(\tilde{w}, \vartheta_{\text{new}}, \vartheta_{\text{old}}) &= \min \left\{ 1, \frac{\pi(\vartheta_{\text{new}})P(y|\vartheta_{\text{new}})q(\vartheta_{\text{old}}|\vartheta_{\text{new}})}{\pi(\vartheta_{\text{old}})P(y|\vartheta_{\text{old}})q(\vartheta_{\text{new}}|\vartheta_{\text{old}})} \cdot \frac{P(\tilde{w}|\vartheta_{\text{old}})}{P(\tilde{w}|\vartheta_{\text{new}})} \right\} \\ &= \min \left\{ 1, \frac{\pi(\vartheta_{\text{new}})f(y; \vartheta_{\text{new}})q(\vartheta_{\text{old}}|\vartheta_{\text{new}})}{\pi(\vartheta_{\text{old}})f(y; \vartheta_{\text{old}})q(\vartheta_{\text{new}}|\vartheta_{\text{old}})} \cdot \frac{f(\tilde{w}; \vartheta_{\text{old}})}{f(\tilde{w}; \vartheta_{\text{new}})} \right\}. \end{aligned} \quad (15)$$

To evaluate  $\alpha(\tilde{w}, \vartheta_{\text{new}}, \vartheta_{\text{old}})$  in equation (15), we need only to simulate the auxiliary networks  $\tilde{w}$  from their probability density function  $P(\tilde{w}|\vartheta_{\text{new}}) = f(\tilde{w}; \vartheta_{\text{new}})/D(\vartheta_{\text{new}})$  that shares the same normalizing term, that is,  $D(\vartheta_{\text{new}})$  in  $P(\tilde{y}|\vartheta_{\text{new}})$ .

<sup>27</sup>Mele (2017b) used the term “approximate exchange algorithm” instead of double M–H algorithm in his paper. He also provides the formal statement of the convergence of the algorithm in Appendix B of his paper.

### 3.4 Posterior distributions and the MCMC

We now present the MCMC procedure. As we regard unobserved latent variables  $\{Z_g\}$  as individual random effects, we follow Zeger and Karim (1991), Hoff, Raftery, and Handcock (2002), and Handcock, Raftery, and Tantrum (2007) to basically treat  $\{Z_g\}$  as parameters to be estimated. By Bayes' theorem, the joint posterior distribution of the parameters and unobservables in the model based on the likelihood function of equation (12) can be written as

$$P(\{\theta_g\}, \{\alpha_g\}, \{Z_g\} | \{Y_g\}, \{W_g\}) \\ \propto \pi(\{\theta_g\}, \{\alpha_g\}, \{Z_g\}) \cdot \prod_{g=1}^G P(Y_g, W_g | \theta_g, \alpha_g, Z_g), \quad (16)$$

where  $\pi(\cdot)$  represents the density function of the prior distribution, and as earlier we suppress the dependence of the likelihood function on  $\{X_g\}$  and  $\{C_g\}$  for notational clarity. We discuss the choice of prior distributions in the Online Supplementary Appendix D.2.

Obtaining draws directly from the joint posterior distribution of equation (16) is challenging. Thus, we block the unknown parameters and latent variables into subgroups and proceed with the Gibbs sampling. We provide the list of conditional posterior distributions used by the Gibbs sampler in the Online Supplementary Appendix D.3. A subset of the parameters admit closed-form conditional posterior distributions; therefore, they can be drawn directly to improve convergence. However, this is not true for remaining parameters, and therefore we use the double M–H algorithm in Section 3.3 to draw from those relevant conditional posterior distributions. Tierney (1994) and Chib and Greenberg (1996) have shown that the combination of Markov chains (Metropolis-within-Gibbs) remains a Markov chain with the invariant distribution being the correct objective distribution.

To understand the general approach and to compare our algorithm with the literature, we first present a pseudo MCMC algorithm, highlighting the main double M–H steps of our formal and full MCMC algorithm. The presentation of the pseudo MCMC allows us to present the main steps of the formal algorithm without introducing heavy notation and computational details. The formal and full MCMC algorithm—with one left-censored and one uncensored activities—is presented in the Online Supplementary Appendix D.4.

**ALGORITHM 1 (Pseudo MCMC).** *In each iteration  $t$ , given  $\vartheta^{(t-1)} = \{\{\theta_g\}^{(t-1)}, \{\alpha_g\}^{(t-1)}, \{Z_g\}^{(t-1)}\}$  from the previous iteration, perform the following double M–H steps sequentially for each group  $g = 1, \dots, G$  and for any variable  $\Xi_g \in \vartheta$ :*

- (a) *Propose  $\tilde{\Xi}_g$  from  $q(\Xi_g | \Xi_g^{(t-1)})$ .*
- (b) *Compute the residuals  $\tilde{\xi}_g$  from the activity intensity equation conditional on the proposed  $\tilde{\Xi}_g$  and the value of the other unknown parameters and variables at iteration  $t - 1$ , that is,  $(\vartheta \setminus \Xi_g)^{(t-1)}$ .*

(c) *Simulate auxiliary network  $\tilde{W}_g$ . Set the initial auxiliary network  $(\tilde{W}_g^{(0)})$  equal to the observed network  $W_g$ . Conditional on  $\tilde{\xi}_g$ ,  $\tilde{\Xi}_g$ , and  $(\vartheta \setminus \Xi_g)^{(t-1)}$ , looping over all entries (except the diagonal ones) of  $\tilde{W}_g^{(0)}$  and for each entry performing either a local or a global update (with probability  $P_{\text{inv}}$ ) described below:*

(i) *local update: for entry  $w_{ij}$ , where  $j \neq i$ , propose  $\tilde{w}_{ij,g}^{(r)} = 1 - \tilde{w}_{ij,g}^{(r-1)}$ . Accept  $\tilde{w}_{ij,g}^{(r)}$  with the probability*

$$\alpha_{\text{MH,local}}(\tilde{w}_{ij,g}^{(r)}, \tilde{w}_{ij,g}^{(r-1)}) = \min \left\{ 1, \frac{P(\tilde{w}_{ij,g}^{(r)}, \tilde{W}_{-ij,g}^{(r-1)} | \tilde{\xi}_g, \tilde{\Xi}_g, (\vartheta \setminus \Xi_g)^{(t-1)})}{P(\tilde{w}_{ij,g}^{(r-1)}, \tilde{W}_{-ij,g}^{(r-1)} | \tilde{\xi}_g, \tilde{\Xi}_g, (\vartheta \setminus \Xi_g)^{(t-1)})} \right\},$$

*otherwise, set  $\tilde{w}_{ij,g}^{(r)} = \tilde{w}_{ij,g}^{(r-1)}$ .*

(ii) *global update: propose  $\tilde{W}_g^{(r)}$  which inverts the entire adjacency matrix, that is,  $\tilde{W}_g^{(r)} = l_{m_g} l'_{m_g} - I_{m_g} - \tilde{W}_g^{(r-1)}$ . Accept  $\tilde{W}_g^{(r)}$  with the probability*

$$\alpha_{\text{MH,global}}(\tilde{W}_g^{(r)}, \tilde{W}_g^{(r-1)}) = \min \left\{ 1, \frac{P(\tilde{W}_g^{(r)} | \tilde{\xi}_g, \tilde{\Xi}_g, (\vartheta \setminus \Xi_g)^{(t-1)})}{P(\tilde{W}_g^{(r-1)} | \tilde{\xi}_g, \tilde{\Xi}_g, (\vartheta \setminus \Xi_g)^{(t-1)})} \right\},$$

*otherwise, set  $\tilde{W}_g^{(r)} = \tilde{W}_g^{(r-1)}$ .*

*Repeat the described updating procedure  $R$  times and obtain the realization of  $\tilde{W}_g^{(R)}$  as the simulation result.<sup>28</sup> If the obtained auxiliary network does not belong to  $\Omega_{\tilde{n},g}$ , reject it and rerun the simulation.*

(d) *Set  $\Xi_g^{(t)}$  equal to  $\tilde{\Xi}_g$  with the probability*

$$\begin{aligned} & \alpha_{\text{MH},\Xi_g}(\tilde{w}, \tilde{\Xi}_g, \Xi_g^{(t-1)}) \\ &= \min \left\{ 1, \frac{\pi(\tilde{\Xi}_g) P(y | \tilde{\Xi}_g, (\vartheta \setminus \Xi_g)^{(t-1)}) q(\Xi_g^{(t-1)} | \tilde{\Xi}_g)}{\pi(\Xi_g^{(t-1)}) P(y | \Xi_g^{(t-1)}, (\vartheta \setminus \Xi_g)^{(t-1)}) q(\tilde{\Xi}_g | \Xi_g^{(t-1)})} \right. \\ & \quad \left. \times \frac{P(\tilde{w} | \Xi_g^{(t-1)}, (\vartheta \setminus \Xi_g)^{(t-1)})}{P(\tilde{w} | \tilde{\Xi}_g, (\vartheta \setminus \Xi_g)^{(t-1)})} \right\}, \end{aligned}$$

*otherwise, set  $\Xi_g^{(t)} = \Xi_g^{(t-1)}$ .*

Note that Step (c) of Algorithm 1 represents the simulation of the auxiliary variable  $\tilde{w} = \{\tilde{W}_g\}$  discussed in Section 3.3. It is worth noting that some authors have used similar estimation strategies. For example, Mele (2017b) used a special case of Algorithm 1 where  $\vartheta = \{\theta_g\}$  and Step (b) is left out. Note also that Algorithm 1 can be adapted to

<sup>28</sup>Similar local and global updates are suggested in Snijders (2002) and Mele (2017b) to improve the convergence of graph sampling, particularly when the graph distribution exhibits a bimodal shape, one mode having low and the other high graph densities. In practice, we set  $R = 2$  and the probability of global update  $P_{\text{inv}} = 0.01$  in the following simulation and empirical studies.

the case where  $y_{i,d}$  is left-censored. To do so, one only has to incorporate the additional latent variable  $\check{Y}_g$  to the list in  $\Xi_g$ .

To examine the finite sample performance of the Bayesian MCMC sampler, we carry out a Monte Carlo simulation experiment to demonstrate that the MCMC sampler can successfully recover the true parameters from the artificially generated network data. We also use this simulation to show the issue of model misspecification and the associated estimation biases. More details of this simulation study can be found in the Online Supplementary Appendix E.

#### 4. EMPIRICAL APPLICATIONS: FRIENDSHIP NETWORKS, ACADEMIC OUTCOMES, AND SMOKING BEHAVIORS

We present an empirical application of our model on American high school students' friendship networks within the Add Health data, a national survey based on 132 schools, covering grades 7 through 12 (Udry (2003)). Five waves of the survey were conducted between 1994 and 2018. In the wave I *in-school* survey, a total of 90,182 students were interviewed. Respondents answered questions regarding their demographic backgrounds, academic performances, and health related behaviors. Most uniquely, students were asked to nominate up to five male and five female friends. This provides detailed information related to their friendship networks.

The different waves' of *in-home* surveys of the Add Health project ask a greater amount of information about students' families and neighborhoods; however, this information is only recorded for a subset of individuals. To include most of the students' nominated friends and to mitigate as much as possible sampling biases (e.g., Chandrasekhar and Lewis (2011), Liu (2013)), we use the wave I *in-school* survey.

We consider two types of activities that may be subject to social interactions and that are relevant for friendship formation.<sup>29</sup> The first is the student's academic performance (measured by GPA), which is represented by a continuous (uncensored) variable. The second is the student's smoking habit, or more precisely, how frequently a student smokes in a typical week. The latter variable is represented by a censored variable as we observe a significant fraction of nonsmokers.

In the context of social interactions, students' academic performance and smoking behavior are studied extensively as they have important long term consequences on students' future lives and health. Studies of peer effects on students' academic performance, for example, Hoxby (2000), Sacerdote (2001), Hanushek, Kain, Markman, and Rivkin (2003), and Zimmerman (2003) used the linear-in-means model; whereas Calvó-Armengol, Patacchini, and Zenou (2009), Lin (2010), Boucher, Bramoullé, Djebbari, and Fortin (2014), and Liu, Patacchini, and Zenou (2014) use the network interactions model. For studies of peer effects on students' smoking behaviors, evidence of peer effects

<sup>29</sup>Discussions about how academic performance and smoking affect friendship selections can be found in, for example, Kiuru, Burk, Laursen, Salmela-Aro, and Nurmi (2010), Lomi, Snijders, Steglich, and Torló (2011), Flashman (2012), Schaefer, Haas et al. (2013). Other activities may also affect friendship choices. We focus on academic performance and smoking because they are the key subjects of interest discussed in the literature of social interactions.

can be found in Gaviria and Raphael (2001), Powell, Tauras, and Ross (2005), Lundborg (2006), Clark and Lohéac (2007), Fletcher (2010), and Hsieh and Van Kippersluis (2018).

When studying interaction (peer) effects, researchers face a difficulty in identifying correlated effects from group-level unobservables and endogenous selection into groups, as well as separating the endogenous interaction effect from contextual effects in a linear model (the reflection problem of Manski (1993)). Using various approaches (e.g., randomization, fixed effects, etc.) to avoid these difficulties, researchers generally produce evidence for the existence of peer effects.

Hsieh and Lee (2016) considered further the problem of endogenous friendship selection of peer effects on economic activities by modeling unobservables in both network interactions and network formation processes. They find that the endogenous effect on academic performance obtained from a SAR model without controlling for the endogeneity of the spatial weight matrix can be upward biased.

In this paper, we follow Hsieh and Lee (2016) in controlling individual unobservables in the formation of friendship networks and activity outcomes. Furthermore, we investigate the incentive effects of activities in network formation and find that the benefit of interactions from academic learning is an important factor for students to form friendships in a school environment.

#### 4.1 Data summary

We use the Add Health wave I *in-school* survey dataset in which all students in the sampled schools were expected to participate. We let each school be a group, and we ignore friendship relations between schools. Although there could still be network measurement errors due to students' absences, refusal to cooperate, etc., when compared to the strategies relying on *in-home* surveys or defining network groups at the grade level, the issue of missing links in our study is minimized. However, to ease the computation burden, we restrict our sample to small schools having student sizes less than 120.

This school-level sample is particularly well adapted to our study since it is very likely that students know each other. As discussed in Section 2, we assume that friendships are formed conditional on side payments. Students must therefore be sufficiently aware of one another as to be able to pay those transfers.

The final sample comprises a total of 1036 respondents from 15 schools (groups).<sup>30</sup> The school networks have an average size of 69.29, an average density of 0.076, an average out-degree of 3.752, and an average clustering coefficient of 0.095. Since the average out-degree is far below its top-coded value (at 10), the threat of missing links due to the fixed survey design (Kossinets (2006)) could be ignored.

To capture the local network effects in equation (8), we include an individual specific variable denoting how many years a student has spent in his or her school as well as three dummy variables: whether a pair of students are of the same age, sex, or race.

<sup>30</sup>To clarify, we do not use the Add Health saturation sample (Udry (2003)) which consists of 16 schools. In this saturation sample, all enrolled students in the schools were selected for in-home interviews; thus, it is an ideal sample if information from in-home interviews is needed. However, since we do not use variables from the in-home survey, we do not use the saturation sample.

TABLE 1. Summary statistics.

Variable	Min	Max	Mean	SD
GPA	1	4	3.059	0.742
Smoking	0 (73.26%)	7	0.543	1.715
Age	11	18	13.788	1.641
Male	0	1	0.457	0.498
<i>Female</i>	0	1	0.543	0.498
<i>White</i>	0	1	0.741	0.438
Black	0	1	0.128	0.335
Asian	0	1	0.017	0.131
Hispanic	0	1	0.050	0.218
Other race	0	1	0.063	0.243
Both parents	0	1	0.821	0.383
Less HS	0	1	0.069	0.254
<i>HS</i>	0	1	0.326	0.469
More HS	0	1	0.460	0.499
Edu missing	0	1	0.094	0.291
Professional	0	1	0.273	0.446
<i>Staying home</i>	0	1	0.230	0.421
Other jobs	0	1	0.375	0.484
Job missing	0	1	0.069	0.254
Welfare	0	1	0.002	0.044
Num. of other students at home	0	6	0.591	0.850
Network size	29	101	69.29	23.96
Network density	0.016	0.136	0.076	0.062
Out-degree	0.000	10.000	3.752	2.688
Clustering coefficient	0.025	0.186	0.095	0.048
Sample size		1036		
Num. of networks		15		

*Note:* “Both parents” means living with both parents. “Less HS” means mother’s education is lower than a high-school level, “HS” means mother’s education level is high school. “More HS” means mother’s education is above a high-school level. “Edu missing” means mother’s education level is missing. “Professional” means mother’s employment is as either a scientist, teacher, executive, director, and the like. “Other jobs” means mother’s occupation is not among Professional or Staying home categories. “Job missing” means the mother’s occupation information is missing. “Welfare” means the mother participates in social welfare programs. “Num. of other students at home” means the number of other students from grades 7 to 12 living in the same household with the student. The variables in italics are the omitted categories during the estimation.

For the variables used in the activity intensity equation of equation (9), the continuous (uncensored) activity outcome, GPA, is calculated using the average of a respondent’s reported grades from several subjects, including language, social science, mathematics, and science (each of which each has a value between 1 and 4). The average GPA in the sample is 3.059. The censored activity variable, smoking, is obtained from the student’s response to the survey question, “*During the past twelve months, how often did you smoke cigarettes?*”; the response has a value between 0 and 7. The average smoking frequency is 0.543 with 73.26% of observations censored at zero. We follow [Lin \(2010\)](#), [Lee, Liu, and Lin \(2010\)](#), and [Hsieh and Lee \(2016\)](#) to choose independent variables. A complete list of variables and their summary statistics are in Table 1.



## 4.2 Estimation results

In this empirical study, we specify our full model with the incentive effects from both activity outcomes—GPA and smoking. As discussed in Section 2.2, despite our assumption of conditional separability, omitting a relevant activity would likely bias the estimation. We therefore proceed to estimate the multiple-activity model. Separate estimations for the two single-activity models are provided in Supplementary Appendix Tables E3 and E4.<sup>31</sup> In addition, modeling multiple activity outcomes allows us to explicitly control the correlation of error terms between activities.

Similar to the simulation study in the Online Supplementary Appendix E, we compare the estimation results of the full model with those of several possibly misspecified models in order to see how each model misspecification affects estimates of the full model, particularly for the estimate of endogenous peer effect ( $\lambda$ ) on activity outcomes in equation (9). We present the estimation results in Table 2. From the first to the fifth columns, they are respectively the results of the full model, the model without the latent variables, the model without the global network effects, the model with only latent variables, and lastly the activity intensity equation alone like equation (7) assuming exogenous network links.<sup>32</sup>

**4.2.1 Network Formation** The results on local network effects in the full model are as follows. Staying in a same school for a longer time has a significant negative effect on sending out ( $\gamma_1$ ) but a positive effect ( $\gamma_2$ ) on receiving friendship nominations. The exogenous dyad-specific effects are all positive and significant, where the effect of the same age ( $\gamma_{31}$ ) is strongest, followed by the effect of the same sex ( $\gamma_{32}$ ), and then the effect of the same race ( $\gamma_{33}$ ). We find that the distances of latent variables ( $\gamma_{41}$  to  $\gamma_{43}$ ) have significant negative effects on network formation, confirming the existence of homophily with respect to unobservables (Hoff, Raftery, and Handcock (2002), Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016)).<sup>33</sup>

<sup>31</sup>Comparing the results of the multiple-activity model in Table 2 with those of the single-activity models in the Online Supplementary Appendix Tables E3 and E4, we find that the latent variables in the local network effects in single-activity models have smaller estimated coefficients. Also, the estimated incentive effect from smoking is higher, and the estimated endogenous effect on smoking is lower in the single-activity model compared to those in Table 2. These differences illustrate the potential concern of omitted-variable biases when activity outcomes are modeled separately.

<sup>32</sup>In Table 2, the mean and the standard deviation (in parentheses) of the MCMC posterior draws are reported as point estimates for each parameter. We set the hyperparameters in the prior distributions to be identical to those used in the simulation study presented in the Online Supplementary Appendix E.

<sup>33</sup>Because the exact likelihood value for the full model in equation (12) is unavailable due to the intractable denominator, we cannot directly apply the likelihood-based model selection criteria to choose the number of latent dimensions in the full model. Alternatively, we determine the latent dimensions based on the model in which the global network effects and the incentive effects are taken away. When there are no global network effects or incentive effects in the network formation model, each link becomes conditionally independent given the latent variables. In that case, the likelihood value can be computed, and we can apply the Akaike's information criterion–Monte Carlo (AICM) as proposed by Raftery, Newton, Satagopan, and Krivitsky (2007) to choose the latent dimension (Hsieh and Lee (2016)). We report the estimation results for that model having one to four latent variable dimensions and the corresponding AICM values in the Online Supplementary Appendix Table E5. Dimension three is chosen due to it having the smallest AICM value.

TABLE 2. Estimation results based on both GPA and smoking.

	Full	No Latent	No Global	Latent Only	Activity Alone
<i>Local &amp; global &amp; incentive effects</i>					
Constant ( $\gamma_0$ )	-4.8531 (0.0366)	-5.7071 (0.1194)	-3.9329 (0.1134)	1.1101 (0.0958)	-
Experience in school (sender) ( $\gamma_1$ )	-0.0405 (0.0114)	-0.0159 (0.0212)	0.0892 (0.0211)	0.0528 (0.0148)	-
Experience in school (receiver) ( $\gamma_2$ )	0.0374 (0.0143)	0.0515 (0.0198)	0.1586 (0.0247)	0.1301 (0.0147)	-
Same age ( $\gamma_{31}$ )	0.3675 (0.0431)	0.5340 (0.0496)	1.0300 (0.0773)	0.4450 (0.0586)	-
Same sex ( $\gamma_{32}$ )	0.3471 (0.0441)	0.4889 (0.0608)	0.3454 (0.0686)	0.5477 (0.0517)	-
Same race ( $\gamma_{33}$ )	0.3116 (0.0430)	0.2152 (0.0789)	0.3716 (0.0839)	0.3961 (0.0595)	-
Latent distance ( $\gamma_{41}$ )	-0.1469 (0.0173)	-	-0.3945 (0.0671)	-3.9387 (0.1214)	-
Latent distance ( $\gamma_{42}$ )	-0.0966 (0.0414)	-	-0.2231 (0.0662)	-2.5592 (0.1000)	-
Latent distance ( $\gamma_{43}$ )	-0.0125 (0.0528)	-	-0.0868 (0.0594)	-2.3793 (0.0887)	-
Reciprocity ( $\eta_1$ )	1.4309 (0.0476)	1.4080 (0.0552)	-	-	-
Congestion ( $\eta_2$ )	0.2699 (0.0151)	0.3521 (0.0281)	-	-	-
Congestion ( $\eta_3$ )	-0.0247 (0.0017)	-0.0304 (0.0022)	-	-	-
Popularity ( $\eta_4$ )	0.0049 (0.0068)	0.0015 (0.0061)	-	-	-
Trans. triads ( $\eta_5$ )	0.4715 (0.0220)	0.4767 (0.0189)	-	-	-
Three cycles ( $\eta_6$ )	-0.2071 (0.0171)	-0.2083 (0.0166)	-	-	-

(Continues)

TABLE 2. Continued.

	Full		No Latent		No Global		Latent Only		Activity Alone	
	Own	Contex.	Own	Contex.	Own	Contex.	Own	Contex.	Own	Contex.
Incentive from GPA ( $\delta_1$ )	0.2145 (0.0956)		0.1825 (0.0344)		0.2921 (0.1122)		-		-	
Incentive from smoking ( $\delta_2$ )	0.0197 (0.0134)		0.0083 (0.0053)		0.0214 (0.0131)		-		-	
Activity intensity—GPA										
Endogenous ( $\lambda$ )	0.0177 (0.0063)		0.0189 (0.0070)		0.0162 (0.0071)		0.0239 (0.0091)		0.0330 (0.0105)	
Age	-0.0327 (0.0164)	0.0018 (0.0024)	-0.0368 (0.0136)	0.0013 (0.0024)	-0.0382 (0.0148)	0.0009 (0.0025)	-0.0388 (0.0202)	0.0009 (0.0025)	-0.0366 (0.0167)	-0.0013 (0.0026)
Male	-0.1984 (0.0307)	-0.0152 (0.0215)	-0.1926 (0.0381)	-0.0166 (0.0239)	-0.1854 (0.0335)	-0.0143 (0.0212)	-0.1863 (0.0381)	-0.0166 (0.0232)	-0.1906 (0.0359)	-0.0161 (0.0244)
Black	-0.0014 (0.0473)	-0.0636 (0.0247)	-0.0564 (0.0613)	-0.0564 (0.0234)	-0.0265 (0.0433)	-0.0608 (0.0226)	-0.0109 (0.0716)	-0.0489 (0.0219)	-0.0607 (0.0626)	-0.0577** (0.0212)
Asian	0.0028 (0.0723)	-0.1957 (0.0566)	-0.0791 (0.1004)	-0.2240 (0.0659)	-0.0574 (0.0501)	-0.1647 (0.0393)	-0.0384 (0.0902)	-0.1824 (0.0567)	-0.0309 (0.0826)	-0.2251*** (0.0643)
Hispanic	-0.2332 (0.0360)	-0.0362 (0.0485)	-0.2241 (0.0776)	-0.0381 (0.0576)	-0.2406 (0.0442)	-0.0644 (0.0337)	-0.2433 (0.0737)	-0.0502 (0.0500)	-0.1990 (0.0680)	-0.0278 (0.0526)
Other race	-0.0689 (0.0418)	-0.0967 (0.0449)	-0.1168 (0.0673)	-0.0873 (0.0483)	-0.0671 (0.0524)	-0.1137 (0.0385)	-0.0555 (0.0575)	-0.0838 (0.0473)	-0.0914 (0.0556)	-0.1020** (0.0491)
Both parents	0.1149 (0.0368)	-0.0701 (0.0213)	0.1140 (0.0510)	-0.0598 (0.0271)	0.1147 (0.0382)	-0.0525 (0.0268)	0.1145 (0.0427)	-0.0602 (0.0258)	0.1126 (0.0449)	-0.0637** (0.0293)
Less HS	-0.3424 (0.0586)	-0.0889 (0.0392)	-0.3466 (0.0702)	-0.0895 (0.0450)	-0.3074 (0.0507)	-0.1027 (0.0333)	-0.2972 (0.0623)	-0.0924 (0.0435)	-0.3070 (0.0608)	-0.0839** (0.0378)
More HS	0.1593 (0.0334)	0.0485 (0.0211)	0.1561 (0.0397)	0.0419 (0.0236)	0.1686 (0.0324)	0.0397 (0.0223)	0.1536 (0.0435)	0.0261 (0.0248)	0.1553 (0.0367)	0.0272 (0.0256)
Edu missing	-0.1739 (0.0532)	0.0634 (0.0334)	-0.1751 (0.0588)	0.0457 (0.0391)	-0.1696 (0.0362)	0.0355 (0.0341)	-0.1624 (0.0634)	0.0472 (0.0366)	-0.1727 (0.0533)	0.0323 (0.0342)
Welfare	-0.1079 (0.0567)	-0.2848 (0.0723)	-0.1593 (0.0950)	-0.2434 (0.0980)	-0.0547 (0.0634)	-0.2932 (0.0550)	-0.1001 (0.1589)	-0.2892 (0.1173)	-0.1973 (0.1000)	-0.1893** (0.0716)

(Continues)

TABLE 2. Continued.

	Full		No Latent		No Global		Latent Only		Activity Alone	
	Own	Context.	Own	Context.	Own	Context.	Own	Context.	Own	Context.
Job missing	-0.1335 (0.0369)	-0.0446 (0.0442)	-0.1261 (0.0662)	-0.0359 (0.0476)	-0.1808 (0.0551)	-0.0371 (0.0387)	-0.1625 (0.0503)	-0.0522 (0.0381)	-0.1500 (0.0525)	-0.0253 (0.0464)
Professional	-0.1104 (0.0309)	0.0010 (0.0267)	-0.1158 (0.0459)	0.0075 (0.0278)	-0.1162 (0.0376)	0.0119 (0.0209)	-0.1347 (0.0377)	0.0003 (0.0264)	-0.1097 (0.0387)	0.0076 (0.0246)
Other jobs	-0.0361 (0.0360)	0.0268 (0.0229)	-0.0482 (0.0412)	0.0283 (0.0222)	-0.0610 (0.0312)	0.0310 (0.0167)	-0.0679 (0.0377)	0.0214 (0.0227)	-0.0354 (0.0392)	0.0289 (0.0214)
Num. of other students at home	0.0257 (0.0230)	0.0123 (0.0104)	0.0195 (0.0257)	0.0131 (0.0133)	0.0220 (0.0212)	0.0154 (0.0119)	0.0216 (0.0232)	0.0096 (0.0140)	0.0184 (0.0232)	0.0068 (0.0131)
Latent ( $\rho_{11}$ )	0.0211 (0.0548)	-0.0198 (0.0283)	-	-	0.0067 (0.0275)	0.0014 (0.0230)	0.0272 (0.0369)	-0.0170 (0.0112)	-	-
Latent ( $\rho_{12}$ )	0.0905 (0.0537)	0.0070 (0.0254)	-	-	0.0260 (0.0610)	-0.0020 (0.0242)	0.0721 (0.0472)	-0.0235 (0.0197)	-	-
Latent ( $\rho_{13}$ )	0.0542 (0.0481)	0.0101 (0.0277)	-	-	0.0545 (0.0454)	0.0105 (0.0232)	0.0724 (0.0466)	-0.0045 (0.0188)	-	-
<i>Activity intensity—smoking</i>										
Endogenous ( $\lambda$ )	0.1052 (0.0196)		0.1103 (0.0225)		0.1068 (0.0193)		0.1056 (0.0197)		0.1125 (0.0191)	
Age	0.1462 (0.0201)	-0.0047 (0.0036)	0.1613 (0.0248)	-0.0054 (0.0045)	0.1837 (0.0386)	-0.0052 (0.0040)	0.2051 (0.0455)	-0.0120 (0.0053)	0.1655 (0.0212)	-0.0042 (0.0040)
Male	0.0126 (0.0650)	-0.0185 (0.0305)	-0.0598 (0.0691)	0.0390 (0.0424)	0.0307 (0.0594)	-0.0077 (0.0477)	-0.0494 (0.0915)	0.0414 (0.0436)	-0.0240 (0.0624)	0.0160 (0.0395)
Black	-0.1952 (0.0749)	0.0828 (0.0194)	-0.3633 (0.1395)	0.0726 (0.0357)	-0.2332 (0.1003)	0.0808 (0.0375)	-0.4153 (0.2075)	0.0810 (0.0377)	-0.0383 (0.0954)	0.0939 (0.0311)
Asian	-1.1052 (0.0615)	0.3028 (0.0444)	-0.5535 (0.1917)	0.1893 (0.0672)	-1.0746 (0.0857)	0.2965 (0.0639)	-1.2797 (0.3075)	0.3093 (0.1120)	-0.5277 (0.2115)	0.2472 (0.1586)
Hispanic	0.1317 (0.0613)	0.0734 (0.0472)	-0.0626 (0.1029)	0.0204 (0.0839)	0.1128 (0.0862)	0.0625 (0.0735)	-0.0591 (0.2321)	0.1509 (0.0849)	0.2026 (0.0990)	0.0232 (0.0734)
Other race	0.0610 (0.0704)	0.4019 (0.0578)	0.3284 (0.0894)	0.2831 (0.0896)	0.0755 (0.0659)	0.3913 (0.0535)	0.1891 (0.1459)	0.5043 (0.0714)	0.2874 (0.1449)	0.3868 (0.0814)
Both parents	-0.0808 (0.0492)	-0.0140 (0.0311)	0.0250 (0.0658)	-0.0431 (0.0434)	-0.0101 (0.0450)	-0.0564 (0.0415)	-0.0588 (0.0878)	-0.0351 (0.0579)	-0.0061 (0.0672)	-0.0437 (0.0382)
Less HS	0.3986 (0.0484)	0.0485 (0.0634)	0.1828 (0.0919)	0.0409 (0.0607)	0.2762 (0.0533)	0.0382 (0.0478)	0.2837 (0.0972)	0.0762 (0.0667)	0.1371 (0.1032)	0.0426 (0.0676)

(Continues)

TABLE 2. Continued.

	Full			No Latent			No Global			Latent Only			Activity Alone		
		Yes	Yes		Yes	Yes		Yes	Yes		Yes	Yes		Yes	Yes
More HS	-0.1975 (0.0515)	-0.0357 (0.0369)	-0.2182 (0.0661)	-0.0329 (0.0479)	-0.0548 (0.0443)	-0.0381 (0.0383)	-0.1502 (0.0906)	0.0131 (0.0472)	-0.1895 (0.0679)	-0.0103 (0.0512)					
Edu missing	-0.0965 (0.0636)	-0.0551 (0.0351)	-0.1907 (0.0750)	0.0275 (0.0611)	-0.1202 (0.0744)	-0.0620 (0.0433)	-0.0775 (0.0874)	-0.0273 (0.0806)	-0.1388 (0.0877)	-0.0216 (0.0619)					
Welfare	1.5846 (0.0764)	-0.1062 (0.0550)	1.2705 (0.1107)	-0.1601 (0.2201)	1.5957 (0.0559)	-0.0860 (0.0653)	1.6348 (0.2791)	0.0728 (0.4136)	1.7299 (0.2190)	-0.5848 (0.4791)					
Job missing	0.4931 (0.0612)	0.0860 (0.0542)	0.3091 (0.1322)	0.1156 (0.0683)	0.5466 (0.0496)	0.0798 (0.0658)	0.5669 (0.1226)	0.1639 (0.0907)	0.2347 (0.1107)	0.0045 (0.0696)					
Professional	0.3740 (0.0464)	-0.0013 (0.0258)	0.3326 (0.0791)	0.0014 (0.0500)	0.3079 (0.0852)	0.0492 (0.0381)	0.4512 (0.1044)	0.0344 (0.0559)	0.2685 (0.0949)	-0.0150 (0.0436)					
Other jobs	0.0037 (0.0545)	0.0395 (0.0319)	0.1530 (0.0552)	0.0315 (0.0350)	0.1064 (0.0615)	0.0460 (0.0371)	0.1472 (0.0882)	0.0682 (0.0450)	-0.0551 (0.0758)	0.0047 (0.0311)					
Num. of other students at home	-0.1551 (0.0478)	-0.0105 (0.0234)	-0.0745 (0.0401)	-0.0028 (0.0246)	-0.1057 (0.0342)	0.0073 (0.0195)	-0.0636 (0.0472)	0.0173 (0.0291)	-0.0541 (0.0378)	0.0009 (0.0234)					
Latent ( $\rho_{21}$ )	-0.0921 (0.0288)	-0.0180 (0.0258)	-	-	0.0052 (0.0447)	0.0163 (0.0453)	-0.0119 (0.0510)	-0.0027 (0.0201)	-	-					
Latent ( $\rho_{22}$ )	0.0074 (0.0471)	-0.0057 (0.0331)	-	-	0.0147 (0.0712)	-0.0044 (0.0414)	-0.0332 (0.0574)	0.0168 (0.0290)	-	-					
Latent ( $\rho_{23}$ )	0.1055 (0.0520)	0.0264 (0.0462)	-	-	-0.0125 (0.0632)	0.0250 (0.0365)	0.0758 (0.0562)	-0.0579 (0.0216)	-	-					
Group fixed effect															
$\sigma_{\xi_{uc,g}}^{2(*)}$ (GPA)	0.4508 (0.1532)		0.4801 (0.1527)		0.4618 (0.1501)		0.4668 (0.1548)		0.4809 (0.1591)						
$\sigma_{\xi_{c,g}}^{2(*)}$ (smoking)	3.6320 (3.2199)		3.6435 (3.2489)		3.5293 (3.1070)		3.6085 (3.1239)		3.5702 (3.1364)						
$\sigma_{\xi_{uc,g}}^{(*)}$	-0.2522 (0.2708)		-0.2452 (0.2770)		-0.2448 (0.2738)		-0.2542 (0.2755)		-0.2474 (0.2725)						

Note: The full model contains the activity intensity equations for GPA and smoking and the network formation model, where the network formation model involves the latent characteristic variables, the global effect, and the incentive effect. In the second column, we remove the latent variables from the network formation model. In the third column, we remove the global effect from the network formation model. In the fourth column, we remove the global effect and the latent variables from the network formation model. In the fifth column, we estimate only the activity intensity equations. The MCMC runs for 100,000 iterations, and the first 50,000 runs are dropped due to burn-in. Values in parentheses are standard deviations of draws from MCMC. The parameters  $\sigma_{\xi_{uc,g}}^{2(*)}$ ,  $\sigma_{\xi_{c,g}}^{2(*)}$ , and  $\sigma_{\xi_{uc,g}}^{(*)}$  denote the average of the estimated variances for error terms in the activity intensity equations of GPA, smoking, and their covariances from different groups and the value in the parenthesis is the average of standard deviations. The trace plots of key parameters, i.e.,  $\lambda$  and  $\delta$  and the convergence diagnostics of Geweke (1992), are provided in the Online Supplementary Appendix Figure F.6.

For the global network effects, we find a positive and strong reciprocity effect ( $\eta_1$ ), which is consistent with findings in the literature (Snijders, Van De Bunt, and Steglich (2010), Badev (2013), Mele (2017a)). This reflects the fact that mutual friendship nominations among students are still common (45.64% of friendship links) in our sample. The congestion effect is concave on individual's out-degree. Indeed, the linear effect ( $\eta_2$ ) is positive, while the quadratic effect ( $\eta_3$ ) is negative. This result confirms our conjecture that limited resources, for example, time, energy, and money, may constrain students from making too many friends. The popularity effect ( $\eta_4$ ) is small and insignificant.

The positive and strong transitive-triads effect ( $\eta_5$ ) shows that students value transitive relationships. As expected, the three-cycles effect ( $\eta_6$ ) is negative. As discussed in Snijders, Van De Bunt, and Steglich (2010), this reveals a certain degree of local hierarchy among students. Our estimated results of those global network effects confirm that there are nontrivial negative externalities on link formation that distinguish our model from the Erdős–Rényi random graph model.

The incentive effect from GPA ( $\delta_1$ ) is strong and significant. Therefore, for high school students in our sample, the utility of interactions in academic achievement influences their friendship decisions. In contrast, the incentive effect from smoking ( $\delta_2$ ) is small and insignificant. Hence, it implies that students in our sample barely consider the utility of interactions in smoking as a factor in their friendship decisions.

**4.2.2 Network Interactions on GPA** Our main finding for network interactions on academic performance is that, by controlling network endogeneity through latent variables and the incentive effect ( $\delta_1$ ), the estimated endogenous effect ( $\lambda$ ) on GPA drops from 0.0330 in the activity intensity equation alone (fifth column) to 0.0177 in the full model (first column). This highlights the effectiveness of our joint modeling approach for correcting the selection bias inherited in the activity intensity equation. Our estimate of endogenous effect coefficient on GPA is close but smaller than the estimate (0.019) found in Hsieh and Lee (2016), which is probably due to that they do not incorporate the utility of interactions in their network formation model. We interpret this estimated endogenous peer effect as follows. Through interactions, an individual could raise his/her GPA by 0.0177 units when any friend improves the GPA by one unit. Also note that the overall effect grows with the number of friends. The more friends that an individual has, the stronger the overall effect he/she receives. On average, one standard deviation increase in peers' GPA will increase an individual's GPA by 0.1530 points. This estimate also implies that the social multiplier effects, as measured by elements of  $(I_{m_g} - \lambda W_g)^{-1} l_{m_g}$  between individuals and groups, have an average of 1.0683 and a standard deviation of 0.0546.

Results from the second to the fourth columns in Table 2 show that correction of bias comes from both the incentive effect ( $\delta_1$ ) and the unobserved latent characteristic variables ( $\rho_{11}$  to  $\rho_{13}$ ). When the model only contains the latent variables (results in the fourth column), only 59.48% of the observed bias is corrected.<sup>34</sup> When the model only controls for the incentive effect (results in the second column), we correct for 92.16% of

<sup>34</sup>This percentage is obtained by dividing the difference of estimated  $\lambda$ 's between the fourth and the fifth columns with the difference of estimated  $\lambda$ 's between the first and the fifth columns.

the observed bias. These empirical results confirm the findings of the simulation study in the Online Supplementary Appendix E that omitting the global network effects from the network formation model can result in an upward bias on the estimated incentive effect (results in the third column), and thus indirectly cause pressure on biasing the estimated endogenous peer effect ( $\lambda$ ) downwards.

For the contribution of individual characteristics, we observe that students who are older, male, Hispanic, of other races, having a mother's education level as missing or lower than a high-school level, having a mother's occupation as missing or as a professional, or having a mother that participates in social welfare programs tend to have lower GPA scores. On the contrary, students who live with both parents or have a mother having an education level higher than high school, tend to have a higher GPA. We also see that one latent variable shows a significant positive effect on GPA. The estimates of the contextual effects for students that are either Black, Asian, other race, living with both parents, have a mother with less than a high-school level of education, or have received welfare are found to be positively significant. The estimates of the contextual effects for mothers having greater than a high-school level of education or missing education levels are found to be positively significant in the full model.

**4.2.3 Network Interactions on Smoking** For the smoking outcome, we observe that the estimated endogenous (peer) effect ( $\lambda$ ) drops from 0.1125 in the activity intensity equation alone (fifth column) to 0.1052 in the full model (first column). The smaller selection bias for smoking (as opposed to GPA) is likely due to the small incentive effect ( $\delta_2$ ) for smoking.

We nonetheless see that the correction of bias is largely due to the inclusion of latent variables (comparing the fourth column with the fifth column) rather than the inclusion of the incentive effect (comparing the second column with the fifth column). On average, one standard deviation increase in peers' smoking frequencies will increase an individual's smoking frequency by 0.4092 times. Our estimate also implies that the social multipliers have respectively an average and a standard deviation equal to 1.9692 and 1.2002.

The effects of individual characteristics show that students who are either Black, Asian, have a mother's education level higher than high school, or have more school-age children at their home tend to smoke less than their school counterparts. On the contrary, students who are either older, Hispanic, having mothers that have less than a high-school level education, having mothers that participate in welfare programs, or having mothers that have professional jobs or missing job information tend to smoke more than others. For contextual effects, a student may smoke more if they are surrounded by more friends who are Black, Asian, or other races. A student may smoke less if they have friends whose mothers participate in welfare programs. For the estimated covariances of disturbances in the outcome equations of GPA and smoking, we find the values are generally negative with an average of  $-0.2522$  and standard deviation of  $0.2708$ .

Finally, as an additional robustness check, we estimate the model with correlated random group effects in activity intensity equations, as discussed in the Online Supplementary Appendix D.2. The estimation results are available in the Online Supplementary Appendix Table E.6. We find the estimates of the local, global network, and incentive



effects in network formation, and the endogenous peer effects in the activity intensity equations remain similar to those in Table 2. Since all of the group averages on  $X_g$  and  $Z_g$  used in capturing the mean of correlated random effects are insignificant, there are no significant changes of the estimated own and contextual effects. As a result, our estimation results are robust between the fixed and random group effect settings.

## 5. CONCLUSION

Researchers are interested in network structures to analyze possible impacts of those structures on activity outcomes. As mentioned in Jackson (2010, Section 5), if networks only serve as conduits for diffusion, for example, diseases or ideas, their impacts on outcomes are somewhat mechanical, and one need not worry about any feedback effects from outcomes. However, for studying the impact of a friendship network on outcomes, both the network structure and strategic interactions between the network and outcomes should be considered. This extra consideration should be reflected in a dynamic or static equilibrium model.

In this paper, we propose a static equilibrium model that accounts for those features. We present a complete information game in which students respond to incentives stemming from their interactions with friends that in turn affect their friendship decisions. We also allow for unobserved individual characteristics in network formation and activity outcome equations.

Our empirical results show that American high school students regard the utility of interactions in academic learning as a significant incentive for forming friendships, whereas the utility of interactions in smoking is not. Another novelty of our approach to the social interaction literature is to present a model that allows correcting for possible friendship selection biases in activity outcomes that can be attributed to the specification of incentive effects, latent characteristic variables, or both.

Some issues that are not emphasized in this paper remain important for future extensions. First, we focus on a complete information setup. If this assumption were appropriate for a school setting, it may be questionable in other economic contexts. Second, we abstract from network games with multiple outcome equilibria. In the paper, we circumvent this issue by focusing on continuous outcome variables. In a multiple equilibria setting with discrete choice outcomes, one could either provide an equilibrium selection rule or characterize the estimation problem with moment inequalities. Finally, an interesting way forward would be to apply our model to the study of other types of networks, for example, criminal networks, physician referral networks, or academic coauthor networks.

## APPENDIX A: PROOF OF PROPOSITION 1

The existence and uniqueness of the Nash equilibrium for a fixed network structure follows directly from the literature (e.g., Ballester, Calvó-Armengol, and Zenou (2006), Calvó-Armengol, Patacchini, and Zenou (2009)). We nonetheless include a short proof for completeness.

We start with the case where  $y_{i,d}$  is uncensored. Taking the first-order conditions of  $U_i(W, Y_1, \dots, Y_{\bar{d}})$  with respect to  $y_{i,d}$  leads to

$$\mu_{i,d} - y_{i,d} + \lambda_d \sum_{j=1}^m w_{ij} y_{j,d} = 0,$$

or, rearranging and writing in a matrix form:

$$B_u(Y_d) \equiv Y_d = \mu_d + \lambda_d W Y_d, \quad (\text{A.1})$$

where  $B_u(Y_d)$  denotes the best response function.

For any  $Y_d, \tilde{Y}_d$ , we have  $\|B_u(Y_d) - B_u(\tilde{Y}_d)\|_\infty = |\lambda_d| \|W(Y_d - \tilde{Y}_d)\|_\infty \leq |\lambda_d| \times \|W\|_\infty \|Y_d - \tilde{Y}_d\|_\infty$ . Then  $B_u(Y_d)$  is a contraction mapping whenever  $|\lambda_d| < 1/\|W\|_\infty = 1/\bar{n}$ . By the Banach fixed-point theorem, this implies that there exists a unique Nash equilibrium of  $Y_d$  such that  $B_u(Y_d) = Y_d$ . It also implies that the linear system (A.1) has a unique solution so that  $Y_d^* = [I_m - \lambda_d W]^{-1} \mu_d$ , where the inverse is well-defined.

The case where  $y_{i,d}$  is left censored, that is,  $y_{i,d} \geq 0$  is similar. Indeed, since  $U_i(W, Y_1, \dots, Y_{\bar{d}})$  is concave in  $y_{i,d}$ , the optimal solution  $y_{i,d}^* = \arg \max_{y_{i,d} \geq 0} U_i(W, Y_1, \dots, Y_{\bar{d}})$  is given by 0 or by the first order conditions. Formally,

$$y_{i,d}^* = \max \left\{ 0, \mu_{i,d} + \lambda_d \sum_{j=1}^m w_{ij} y_{j,d} \right\} \quad (\text{A.2})$$

Then, similar to the case where  $y_{i,d}$  is uncensored, we can write the vector-valued best response function  $B_c(Y_d) = [y_{1,d}^*, \dots, y_{m,d}^*]'$ . Now, for a fixed value of  $Y_d$ , note that we necessarily have:  $\|B_c(Y_d) - B_c(\tilde{Y}_d)\|_\infty \leq \|B_u(Y_d) - B_u(\tilde{Y}_d)\|_\infty$ . This implies that if  $B_u(Y_d)$  is a contraction mapping, then so is  $B_c(Y_d)$ . Using the same argument as before, there exists a unique Nash equilibrium of the game for left-censored activities.

We now turn to the first stage of the game. Since there exists a unique Nash equilibrium  $(Y_1^*(W), \dots, Y_{\bar{d}}^*(W))$ , the value of the network  $T(W)$  is uniquely defined. Also, since  $\tau_W$  is drawn from a Type I extreme value distribution, the probability of having more than one network structure maximizing  $T(W)$  is zero. There is, therefore, a generically unique strongly efficient network, and the probability that  $W$  maximizes  $T$  is given by

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\tilde{W} \in \Omega_{\bar{n}}} \exp\{V(\tilde{W})\}}.$$

Existence is guaranteed by letting the allocation rule  $\Lambda_i(W, T) = T(W)/m$  for all  $i$ , which implies that strongly efficient networks are individually stable. This completes the proof.

### A.1 Comment on the contraction mapping assumption

The argument in Appendix A depends on the best response functions being contraction mappings. This is important for uniqueness and for the solution of equation (A.2) to be found iteratively.

First, note that a contraction mapping is only a sufficient condition and in no way necessary. Also, the use of the infinity norm  $\|\cdot\|_\infty$  is stronger than what is needed. Indeed, the argument works as long as there exists a (submultiplicative) matrix norm for which the condition holds.

In particular, one could use the norm given by the spectral radius of  $W$ . We focus on the infinity norm since it is more intuitive. Indeed, otherwise, we would have to require individuals to choose their friends such that  $W$  has a bounded spectral radius, which carries much less economic intuition than requiring them to choose at most  $\bar{n}$  friends.

### REFERENCES

- Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012), “Estimating spillovers using panel data, with an application to the classroom.” *Quantitative Economics*, 3 (3), 421–470. [1362]
- Auerbach, E. (2019), “Identification and estimation of a partially linear regression model using network data.” [arXiv:1903.09679](https://arxiv.org/abs/1903.09679). [1350]
- Badev, A. (2013), “Discrete games in endogenous networks: Theory and policy.” Working paper. [1353, 1354, 1357, 1367, 1379]
- Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006), “‘Who’s who in networks. Wanted: The key player.’” *Econometrica*, 74 (5), 1403–1417. [1351, 1356, 1358, 1359, 1360, 1381]
- Ballester, C., Y. Zenou, and A. Calvó-Armengol (2010), “Delinquent networks.” *Journal of the European Economic Association*, 8 (1), 34–61. [1350]
- Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013), “The diffusion of microfinance.” *Science*, 341 (6144). <http://science.sciencemag.org/content/341/6144/1236498>. [1350]
- Battaglini, M., E. Patacchini, and E. Rainone (2019), “Endogenous social connections in legislatures.” NBER Working Paper No. 25988. [1353, 1354]
- Bayer, P., R. Hjalmarrsson, and D. Pozen (2009), “Building criminal capital behind bars: Peer effects in juvenile corrections.” *Quarterly Journal of Economics*, 124 (1), 105–147. [1350]
- 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 (6), 1150–1196. [1350]
- Besag, J. (1974), “Spatial interaction and the statistical analysis of lattice systems.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 36, 192–236. [1367]

- Bhamidi, S., G. Bresler, A. Sly et al. (2011), "Mixing time of exponential random graphs." *Annals of Applied Probability*, 21 (6), 2146–2170. [1363, 1368]
- Blume, L. E. (1993), "The statistical mechanics of strategic interaction." *Games and Economic Behavior*, 5 (3), 387–424. [1354]
- Boucher, V. (2015), "Structural homophily." *International Economic Review*, 56 (1), 235–264. [1355, 1362]
- Boucher, V. (2016), "Conformism and self-selection in social networks." *Journal of Public Economics*, 136, 30–44. [1351, 1353, 1354, 1356, 1358, 1360]
- Boucher, V., Y. Bramoullé, H. Djebbari, and B. Fortin (2014), "Do peers affect student achievement? Evidence from Canada using group size variation." *Journal of Applied Econometrics*, 29 (1), 91–109. [1371]
- Boucher, V. and I. Mourifié (2017), "My friend far, far away: A random field approach to exponential random graph models." *Econometrics Journal*, 20 (3), S14–S46. [1353, 1367]
- Bramoullé, Y., H. Djebbari, and B. Fortin (2009), "Identification of peer effects through social networks." *Journal of Econometrics*, 150 (1), 41–55. [1350, 1360]
- Bramoullé, Y. and B. Fortin (2009), "The econometrics of social networks." *Cahier de recherche/Working Paper*, 9, 13. [1351, 1356]
- Bramoullé, Y., R. Kranton, and M. D'amours (2014), "Strategic interaction and networks." *American Economic Review*, 104 (3), 898–930. [1351, 1356, 1358]
- Brock, W. A. and S. N. Durlauf (2001), "Discrete choice with social interactions." *The Review of Economic Studies*, 68 (2), 235–260. [1359]
- Calvo-Armengol, A. and M. O. Jackson (2004), "The effects of social networks on employment and inequality." *American Economic Review*, 94 (3), 426–454. [1350]
- Calvó-Armengol, A. and M. O. Jackson (2007), "Networks in labor markets: Wage and employment dynamics and inequality." *Journal of Economic Theory*, 132 (1), 27–46. [1350]
- Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009), "Peer effects and social networks in education." *Review of Economic Studies*, 76 (4), 1239–1267. [1350, 1351, 1356, 1358, 1371, 1381]
- Caulier, J.-F., A. Mauleon, and V. Vannetelbosch (2015), "Allocation rules for coalitional network games." *Mathematical Social Sciences*, 78, 80–88. [1351, 1357]
- Chandrasekhar, A. and R. Lewis (2011), "Econometrics of sampled networks." Unpublished manuscript, MIT. [422]. [1371]
- Chandrasekhar, A. G. and M. O. Jackson (2014), "Tractable and consistent random graph models." NBER working paper. [1353, 1367]
- Chatterjee, S., P. Diaconis et al. (2013), "Estimating and understanding exponential random graph models." *Annals of Statistics*, 41 (5), 2428–2461. [1363]

- Chib, S. and E. Greenberg (1996), “Markov chain Monte Carlo simulation methods in econometrics.” *Econometric Theory*, 12 (03), 409–431. [1369]
- Chib, S. and S. Ramamurthy (2010), “Tailored randomized block mcmc methods with application to dsge models.” *Journal of Econometrics*, 155 (1), 19–38. [1368]
- 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 (4), 370–379. [1350]
- Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010), “An empirical model for strategic network formation.” NBER working paper. [1353, 1357]
- Clark, A. E. and Y. Lohéac (2007), ““it wasn’t me, it was them!” social influence in risky behavior by adolescents.” *Journal of Health Economics*, 26 (4), 763–784. [1372]
- Cohen-Cole, E., X. Liu, and Y. Zenou (2018), “Multivariate choices and identification of social interactions.” *Journal of Applied Econometrics*, 33 (2), 165–178. [1355]
- Comola, M. (2007), “The network structure of informal arrangements: Evidence from rural Tanzania.” Report, Barcelona, Universitat Pompeu Fabra. [1351]
- Conley, T. and C. Udry (2001), “Social learning through networks: The adoption of new agricultural technologies in Ghana.” *American Journal of Agricultural Economics*, 83 (3), 668–673. [1350]
- Conley, T. G. and C. R. Udry (2010), “Learning about a new technology: Pineapple in Ghana.” *American Economic Review*, 100 (1), 35–69. [1350]
- Davis, J. A. (1970), “Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices.” *American Sociological Review*, 843–851. [1363]
- De Paula, Á., S. Richards-Shubik, and E. Tamer (2018), “Identifying preferences in networks with bounded degree.” *Econometrica*, 86 (1), 263–288. [1353, 1355]
- Dutta, B. and M. O. Jackson (2000), “The stability and efficiency of directed communication networks.” *Review of Economic Design*, 5 (3), 251–272. [1351, 1356, 1357, 1358, 1360]
- Dzanski, A. (2019), “An empirical model of dyadic link formation in a network with unobserved heterogeneity.” *Review of Economics and Statistics*, 101 (5). [1353]
- Fafchamps, M. and F. Gubert (2007a), “The formation of risk sharing networks.” *Journal of Development Economics*, 83 (2), 326–350. [1350, 1351, 1361]
- Fafchamps, M. and F. Gubert (2007b), “Risk sharing and network formation.” *American Economic Review*, 97 (2), 75–79. [1350, 1351, 1361]
- Flashman, J. (2012), “Academic achievement and its impact on friend dynamics.” *Sociology of Education*, 85 (1), 61–80. [1371]
- Fletcher, J. M. (2010), “Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects.” *Health Economics*, 19 (4), 466–484. [1372]

- Fowler, J. H. and N. A. Christakis (2008), “Estimating peer effects on health in social networks: A response to Cohen-cole and Fletcher; trogdon, nonnemaker, pais.” *Journal of Health Economics*, 27 (5), 1400–1405. [1350]
- Frank, O. and D. Strauss (1986), “Markov graphs.” *Journal of the American Statistical Association*, 81 (395), 832–842. [1352]
- Fruehwirth, J. C. (2014), “Can achievement peer effect estimates inform policy? A view from inside the black box.” *Review of Economics and Statistics*, 96 (3), 514–523. [1362]
- Gaviria, A. and S. Raphael (2001), “School-based peer effects and juvenile behavior.” *Review of Economics and Statistics*, 83 (2), 257–268. [1372]
- Geweke, J. (1992), “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.” In *Bayesian Statistics*, Citeseer. [1378]
- Geyer, C. J. (1991), “Markov chain Monte Carlo maximum likelihood.” In *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, Interface Foundation of North America. [1368]
- Geyer, C. J. and E. A. Thompson (1992), “Constrained Monte Carlo maximum likelihood for dependent data.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 657–699. [1352, 1367]
- Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (2003), “The social multiplier.” *Journal of the European Economic Association*, 1 (2–3), 345–353. [1360]
- Goldsmith-Pinkham, P. and G. W. Imbens (2013), “Social networks and the identification of peer effects.” *Journal of Business & Economic Statistics*, 31 (3), 253–264. [1350, 1353, 1361, 1374]
- Graham, B. S. (2017), “An econometric model of network formation with degree heterogeneity.” *Econometrica*, 85 (4), 1033–1063. [1351, 1353]
- Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007), “Model-based clustering for social networks.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170 (2), 301–354. [1351, 1369]
- 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 (5), 527–544. [1371]
- 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 (460), 1090–1098. [1351, 1369, 1374]
- Hoxby, C. (2000), “Peer effects in the classroom: Learning from gender and race variation.” NBER Working Paper No. 7867. [1371]
- Hsieh, C.-S., M. D. König, and X. Liu (2019), “A structural model for the coevolution of networks and behavior.” CEPR Discussion Paper No. DP13911. [1353, 1354]

- Hsieh, C.-S., L.-F. Lee, and V. Boucher (2020), “Supplement to ‘Specification and estimation of network formation and network interaction models with the exponential probability distribution.’” *Quantitative Economics Supplemental Material*, 11, <https://doi.org/10.3982/QE944>. [1354]
- Hsieh, C.-S. and L. F. Lee (2016), “A social interactions model with endogenous friendship formation and selectivity.” *Journal of Applied Econometrics*, 31 (2), 301–319. [1350, 1352, 1353, 1354, 1361, 1366, 1372, 1373, 1374, 1379]
- Hsieh, C.-S. and H. Van Kippersluis (2018), “Smoking initiation: Peers and personality.” *Quantitative Economics*, 9 (2), 825–863. [1353, 1362, 1366, 1372]
- Jackson, M. O. (2005), “Allocation rules for network games.” *Games and Economic Behavior*, 51 (1), 128–154. [1351, 1356, 1360]
- Jackson, M. O. (2010), “An overview of social networks and economic applications.” *The Handbook of Social Economics*, 1, 511–585. [1381]
- Jackson, M. O. and A. Watts (2002), “The evolution of social and economic networks.” *Journal of Economic Theory*, 106 (2), 265–295. [1353]
- Jackson, M. O. and A. Wolinsky (1996), “A strategic model of social and economic networks.” *Journal of Economic Theory*, 71, 44–74. [1351, 1353, 1356]
- Jochmans, K. (2018), “Semiparametric analysis of network formation.” *Journal of Business & Economic Statistics*, 36 (4), 705–713. [1351, 1353, 1354]
- Johnsson, I. and H. R. Moon (forthcoming), “Estimation of peer effects in endogenous social networks: Control function approach.” *Review of Economics and Statistics*. [1350]
- Kiuru, N., W. J. Burk, B. Laursen, K. Salmela-Aro, and J.-E. Nurmi (2010), “Pressure to drink but not to smoke: Disentangling selection and socialization in adolescent peer networks and peer groups.” *Journal of Adolescence*, 33 (6), 801–812. [1371]
- Kossinets, G. (2006), “Effects of missing data in social networks.” *Social Networks*, 28 (3), 247–268. [1372]
- Krauth, B. V. (2006), “Simulation-based estimation of peer effects.” *Journal of Econometrics*, 133 (1), 243–271. [1355]
- Lee, L.-F., X. Liu, and X. Lin (2010), “Specification and estimation of social interaction models with network structures.” *Econometrics Journal*, 13 (2), 145–176. [1360, 1373]
- Lehmann, E. L. and G. Casella (2006), *Theory of Point Estimation*. Springer Science & Business Media. [1366]
- Lewis-Faupel, S. (2016), “The formation and human capital consequences of high school social networks.” Working paper. [1353, 1354]
- Liang, F. (2010), “A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants.” *Journal of Statistical Computation and Simulation*, 80 (9), 1007–1022. [1352, 1368]



- Liang, F, I. H. Jin, Q. Song, and J. S. Liu (2016), "An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants." *Journal of the American Statistical Association*, 111 (513), 377–393. [1368]
- Lin, X. (2010), "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables." *Journal of Labor Economics*, 28 (4), 825–860. [1360, 1371, 1373]
- Liu, X. (2013), "Estimation of a local-aggregate network model with sampled networks." *Economics Letters*, 118 (1), 243–246. [1371]
- Liu, X., E. Patacchini, and Y. Zenou (2014), "Endogenous peer effects: Local aggregate or local average?" *Journal of Economic Behavior & Organization*, 103, 39–59. [1350, 1371]
- Lomi, A., T. A. Snijders, C. E. Steglich, and V. J. Torló (2011), "Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance." *Social Science Research*, 40 (6), 1506–1520. [1371]
- Lundborg, P. (2006), "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics*, 25 (2), 214–233. [1372]
- Manski, C. F. (1993), "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60 (3), 531–542. [1372]
- McCormick, T. H. and T. Zheng (2015), "Latent surface models for networks using aggregated relational data." *Journal of the American Statistical Association*, 110 (512), 1684–1695. [1351]
- Mele, A. (2017a), "Segregation in social networks: A structural approach." Available at SSRN 3058985. [1379]
- Mele, A. (2017b), "A structural model of dense network formation." *Econometrica*, 85, 825–850. [1352, 1353, 1354, 1357, 1363, 1366, 1367, 1368, 1370]
- Mele, A. (2017c), "A structural model of homophily and clustering in social networks." Available at SSRN, <https://ssrn.com/abstract=3031489>. [1353]
- Mele, A. and L. Zhu (2017), "Approximate variational estimation for a model of network formation." arXiv:1702.00308. [1353]
- Miyauchi, Y. (2016), "Structural estimation of pairwise stable networks with nonnegative externality." *Journal of Econometrics*, 195 (2), 224–235. [1353]
- Murray, I., Z. Ghahramani, and D. MacKay (2006), "Mcmc for doubly-intractable distributions." In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, UAI. [1352, 1367, 1368]
- Patacchini, E. and Y. Zenou (2008), "The strength of weak ties in crime." *European Economic Review*, 52 (2), 209–236. [1350]
- Patacchini, E. and Y. Zenou (2012), "Juvenile delinquency and conformism." *Journal of Law, Economics, and Organization*, 28 (1), 1–31. [1350]

Powell, L. M., J. A. Tauras, and H. Ross (2005), “The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior.” *Journal of Health Economics*, 24 (5), 950–968. [1372]

Propp, J. G. and D. B. Wilson (1996), “Exact sampling with coupled Markov chains and applications to statistical mechanics.” *Random Structures and Algorithms*, 9 (1–2), 223–252. [1367]

Raftery, A. E., M. A. Newton, J. M. Satagopan, and P. N. Krivitsky (2007), “Estimating the integrated likelihood via posterior simulation using the harmonic mean identity.” *Bayesian Statistics*, 8, 1–45. [1374]

Sacerdote, B. (2001), “Peer effects with random assignment: Results for dartmouth roommates.” *Quarterly Journal of Economics*, 116 (2), 681–704. [1371]

Schaefer, D. R., S. A. Haas et al. (2013), “Social networks and smoking exploring the effects of peer influence and smoker popularity through simulations.” *Health Education & Behavior*, 40 (1 suppl), 24S–32S. [1371]

Sheng, S. (2014), “A structural econometric analysis of network formation games.” Working paper. [1353]

Snijders, T. A. (2001), “The statistical evaluation of social network dynamics.” *Sociological Methodology*, 31 (1), 361–395. [1350]

Snijders, T. A. (2002), “Markov chain Monte Carlo estimation of exponential random graph models.” *Journal of Social Structure*, 3 (2), 1–40. [1352, 1370]

Snijders, T. A., P. E. Pattison, G. L. Robins, and M. S. Handcock (2006), “New specifications for exponential random graph models.” *Sociological Methodology*, 36 (1), 99–153. [1352, 1363]

Snijders, T. A., G. G. Van De Bunt, and C. E. Steglich (2010), “Introduction to stochastic actor-based models for network dynamics.” *Social Networks*, 32 (1), 44–60. [1350, 1362, 1379]

Soetevent, A. R. and P. Kooreman (2007), “A discrete-choice model with social interactions: With an application to high school teen behavior.” *Journal of Applied Econometrics*, 22 (3), 599–624. [1355]

Strauss, D. and M. Ikeda (1990), “Pseudolikelihood estimation for social networks.” *Journal of the American Statistical Association*, 85 (409), 204–212. [1367]

Tierney, L. (1994), “Markov chains for exploring posterior distributions.” *Annals of Statistics*, 1701–1728. [1369]

Udry, J. R. (2003), *The National Longitudinal Study of Adolescent Health (Add Health), Waves i and ii, 1994–1996; Wave iii, 2001–2002 [Machine-Readable Data File and Documentation]*. Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC. [1350, 1371, 1372]

Wasserman, S. and P. Pattison (1996), “Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and  $p^*$ .” *Psychometrika*, 61 (3), 401–425. [1352]

Weinberg, B. A. (2007), “Social interactions with endogenous associations.” NBER working paper. [1350]

Zeger, S. L. and M. R. Karim (1991), “Generalized linear models with random effects; a Gibbs sampling approach.” *Journal of the American Statistical Association*, 86 (413), 79–86. [1369]

Zimmerman, D. J. (2003), “Peer effects in academic outcomes: Evidence from a natural experiment.” *Review of Economics and Statistics*, 85 (1), 9–23. [1371]

---

Co-editor Christopher Taber handled this manuscript.

Manuscript received 15 August, 2017; final version accepted 1 June, 2020; available online 12 June, 2020.