2019-04

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

Chih-Sheng Hsieh
Lung-Fei Lee
Vincent Boucher

Mai / May 2019

Centre de recherche sur les risques les enjeux économiques et les politiques publiques

www.crrep.ca

ABSTRACT

In this paper, we model network formation and network interactions under a unified framework. The key feature of our model is to allow individuals to respond to incentives that stem from interaction benefits of certain activities when they choose friends (network links), while capturing homophily in terms of unobserved characteristic variables in network formation and activities. There are two advantages of this modeling approach: first, one can evaluate whether incentives from certain interactions are important factors for friendship formation or not. Second, in addition to homophily effects in terms of unobserved characteristics, inclusion of incentive effects in the network formulation also corrects possible friendship selection bias on activity outcomes under network interactions. A theoretical foundation of this unified model is based on a sub-game perfect equilibrium of a two-stage game. A tractable Bayesian MCMC approach is proposed for the estimation of the model, and we demonstrate its finite sample performance in a simulation study. We apply the model to study empirically American high school students' friendship networks from the Add Health dataset. We consider two activity variables, GPA and smoking frequency, and find a significant incentive effect from GPA, but not from smoking, on friendship formation. These results suggest that the benefit of interactions in academic learning is an important factor for friendship formation, whereas the interaction benefit of smoking is not. On the other hand, from the perspective of network interactions, both GPA and smoking frequency are subject to significant positive interaction (peer) effects.

JEL Classification: C21, C25, I21, J13
Keywords: Social networks, social interaction, selectivity, spatial autoregression, Bayesian Estimation.

Chih-Sheng Hsieh: Department of Economics, The Chinese University of Hong Kong.
Lung-Fei Lee: Department of Economics, The Ohio State University.
Vincent Boucher: Department of Economics, Université Laval.

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 the editor, Christopher Taber, two anonymous referees, Matthew Jackson, 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 acknowledge 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.

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 transactions.[1] 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 achieve 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), and Johnsson and Moon (2016) study important 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

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

present a novel approach for examining whether peer-influenced economic outcomes play any role in the formation of friendship networks.[2] Specifically, we allow for economic choices that are subject to peer effect, e.g., smoking decisions, that impact an individuals' utility of forming network links.

Formally, we present a two-stage game. First, the network is formed, and then in the second stage, individuals choose the intensity of their involvement in some economic activities. We focus on a "sub-game perfect" equilibrium; we allow individuals to anticipate the second stage of the game when choosing the network.

The formation of the network, i.e., 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 et al., 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 the 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 et al., 2006; Calvó-Armengol et al., 2009; Bramoullé et al., 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 the individuals' incentives that stem from economic activity interactions related to friendship formation. That is, assessing how much individuals anticipate that they will be influenced once the network is formed. Second, use of a jointly coherent model permits controlling for possible friendship selection biases and allows for the study of peer effects of each activity.

A common empirical approach to model network formation is to assume pairwise in-

[2] 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, e.g., Add Health data (Udry, 2003) and Indian rural village data (Banerjee et al., 2013). Limited students' friendship network data having a panel structure can be found in the literature dealing with stochastic actor-based dynamic network modeling; for example, see Snijders (2001) and Snijders et al. (2010).

dependence. For example, Fafchamps and Gubert (2007a,b) 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 et al., 2002; Handcock et al., 2007; McCormick and Zheng, 2015) and models with unobserved degree heterogeneity (Graham, 2017; Jochmans, 2018). In these models, individuals are assumed to have unobserved positions in the network (or fixed effects) that reflect the heterogeneity of their social or economic status. These unobserved positions allow researchers to control the homophily effect in terms of unobserved individual characteristics.[3]

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

In this paper, we go beyond pairwise independence specification and consider the exponential probability distribution to model network data. The idea is to treat any observed network as one of the $2^{m(m-1)}$ possible configurations for links among a population of m individuals. This idea matches the Exponential Random Graph (ERG) model proposed by Frank and Strauss (1986) or, more generally, the p^* model of Wasserman and Pattison (1996) in the statistical literature.

In either an ERG 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 these network structures would appear in a network. However, the parameters of these network statistics in ERG and p^* models do not allow for casual interpretations.

Contrary to the standard literature on ERG models, 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 also control unobserved individual heterogeneity through latent variables, as in Hsieh and Lee (2016). As a result, our proposed network formation model handles three distinguished features: observed and

[3]Under pairwise independence, the likelihood of the entire network being conditional on the unobservables is the product of the likelihoods from all pairs.

unobserved individual heterogeneity, global network dependence, and 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 generic and flexible specification is that it complicates the estimation. Indeed, the likelihood function of an ERG model 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 et al., 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 algorithm proposed by Mele (2017b). 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.

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) and Jochmans (2018) introduce node-specific parameters to capture degree heterogeneity in a pairwise independent link formation model; however, they ignore any possible network externality effects.

Sheng (2014), Miyauchi (2016), and De Paula et al. (2018) specify strategic network formation models and characterize the equilibria of the model by pairwise stability condition (Jackson and Wolinsky, 1996). Instead of imposing equilibrium selection assumptions, they specify incomplete models and utilize a partial identification approach.

Christakis et al. (2010) and Mele (2017b) model 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 ERG models of network formation (e.g., Boucher and Mourifié, 2017; Chandrasekhar and Jackson, 2014; 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 endogeneous networks (e.g., Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee,2016).[4] In particular, this paper contributes to the emerging empirical literature studying the impact of economic actions on the formation of a network.[5] To our knowledge, the only three existing papers that deal with this aspect are Badev (2018), Boucher (2016) and Lewis-Faupel (2016).

Lewis-Faupel (2016) focuses on a setup where individuals take a single binary action under rational expectations. This binary action is anticipated in the network formation

[4]See also Boucher and Fortin (2016) for a recent discussion and additional references.

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

stage, where individuals form heuristic expectations.

Badev (2018) also focuses on a setup where individuals take a single binary action (affecting the preferences on the network structure). He presents a random utility model and an original equilibrium concept— based on stability constraints—that nests a pairwise-stable and pairwise-Nash network. The equilibrium also follows an ERG model in which all payoff relevant variables are assumed to be observed and where existence is only guaranteed for potential games.

Boucher (2016) presents 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.

This paper is organized 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 a simulation exercise to examine the finite sample performance of the proposed Bayesian MCMC estimation method. Section 5 provides an application of the model to high school students' friendship networks and activities with the Add Health data. Section 6 concludes the paper. Some additional technical details for the estimation and additional empirical results are relegated to the Supplementary Appendix.

2 An 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.[6] The $(i,j)^{\text{th}}$ 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

[6] It is also called spatial weights matrix, adjacency matrix, or sociomatrix in the literature.

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).[7] We normalize diagonal elements so that $w_{ii} = 0$ for all $i \in (1, ..., m)$.

Let X be the $m \times k$ matrix of individual characteristics and let x_i be the i^{th} row of X. Individuals choose the intensity of their involvement in economic activities. For each activity $d \in (1, ..., \bar{d})$, let $y_{i,d}$ denote the intensity of individual i in activity d. Let also $Y_d = (y_{1,d}, \cdots, y_{m,d})'$ 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 $y_{i,d} \in \mathbb{R}_+$.[8] The impact of left censoring on equilibrium activity intensity is formally described in Proposition 1 in Section 2.3.

2.2 Preferences

The preferences of any given individual i are represented by the utility function:

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

where $v_i(W)$ represents an explicit preference over the network structure 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 $\delta_d \geq 0$ therefore 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.[9]

It is worth noting that, conditional on W, the utility is separable across activities, so we assume no complementarity across activities. However, since the network structure is endogenously determined, the optimal choice for each activity will, at equilibrium, be a

[7] This non-reciprocity is motivated by our empirical application. In fact, 54.36% of friendship links in our dataset (Add Health) are non-reciprocal. This assumption is also present in Hsieh and Lee (2016), Mele (2017b), Jochmans (2018), and others.

[8] We abstract from discrete choices in this paper as it generally involves multiple equilibria (Krauth, 2006; Soetevent and Kooreman, 2007). This is left for future research.

[9] Since this is a two-stage game (see Section 2.3), δ_d can also be interpreted as an activity-specific discount rate.

function not only of the individual's preferences for this particular activity but also of their preferences for the other activities, and their explicit network preferences.[10]

Indeed, the equilibrium value for $(W, Y_1, ..., Y_{\bar{d}})$ is the result of a complex interplay between individuals' preferences over the network structure and their preferences regarding each type of activity. This has important consequences for estimation, as will be discussed in Section 5.

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

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

where $\mu_{i,d}$ captures individual exogenous heterogeneity. The first and second terms of Eq. (2) represent the private benefit and cost of increasing the *intensity* the activity (i.e., $y_{i,d}$). The third term reflects an additional social benefit (or cost) of increasing the intensity of the activity for the individual, i.e., 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:

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

where $\tau_{i,W}$ is an idiosyncratic shock on the value of the network W for individual i.

In Eq. (3), the local network effects captured by ψ_{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 individuals 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 $\varpi_i(w_i, W_{-i})\eta$ 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 summary statistics that are relevant to individual i's utility, and η is the corresponding vector of coefficients.

[10] We also assume that the unobserved (for the econometrician) part of the utility functions for each activity may be correlated. See Section 5.

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 ERG models 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 $(W, Y_1, ..., Y_{\bar{d}})$ and the timing of the game.

2.3 The game

The game occurs in two stages. In the first stage, the network links are determined. In the second stage, individuals play a non-cooperative game for the choice of the activity intensities $(Y_1, \cdots, Y_{\bar{d}})$, 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 et al., 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 ERG models (i.e., as in Eq. (6) below). Indeed, the usual microeconomic foundation is based on Christakis et al. (2010), Badev (2018), and Mele (2017b) where individuals are assumed to have the opportunity periodically to meet another and revise their friendship status (in a non-transferable 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.[11] 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

[11]The randomness is due to the unobserved preference shocks $\tau_{i,W}$.

they have, but also on the the global network effects (see Eq. (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.

Figure 1: Example of a simple global network effect.

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* $T(W)$. 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, i.e., $T(W) = \sum_i U_i(W)$.[12] This definition allows to see side payments as being made in "utility units". We focus on *strongly efficient* networks (Dutta and Jackson, 2000), i.e., 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, i.e., how side payments are made. This is formally described by the *allocation rule* $\Lambda_i(W, T) : T(W) = \sum_i \Lambda_i(W, T)$.[13] If no side payments are allowed, the allocation rule is simply given by the individuals' utility: $\Lambda_i(W, T) = U_i(W)$.[14] Alternatively, if the value of the network is shared equally among individuals, we get:

[12] The dependence of U_i on Y is omitted on purpose. The formal definition is presented in Definition 1 below.

[13] This definition assumes implicitly that the allocation rule is *balanced*, as in Dutta and Jackson (2000).

[14] 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: $\Lambda_i(W, T) - U_i(W)$.

$\Lambda_i(W,T) = T(W)/m$. Note that the balance requirement of $T(W) = \sum_i \Lambda_i(W,T)$ imposes that no outside resources are used in making side payments. Indeed, using our definition for the network value, we obtain $\sum_i \Lambda_i(W,T) = \sum_i U_i(W)$.

We say that a network W is *individually stable* if for all i, $\Lambda_i(W,T) \geq \Lambda_i(\widetilde{W},T)$ for all networks $\widetilde{W}$ such that $\widetilde{w}_{jk} = w_{jk}$ for all k and all $j \neq i$.[15] 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 (non-empty) set of allocation rules compatible with both strong efficiency and individual stability.

Once the network is formed, individuals are free to select the intensity of the activities in which they are involved, conditional on network structure. We follow the extensive literature for games on networks (e.g., Ballester et al., 2006; Calvó-Armengol et al., 2009; Bramoullé et al., 2014; Boucher, 2016) and assume that activity intensity choices are part of a Nash equilibrium.[16]

Formally, our equilibrium concept for the two-stage game is defined as follows:

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

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

2. *The network value*

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

is strongly efficient and individually stable 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 sub-game perfect since it is solved by backward induction. The next proposition follows.

[15] 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.

[16] That is, $(Y_1^*, ..., Y_{\bar{d}}^*)$ such that $y_{i,d}^* \in \arg\max_{y_{i,d}} \; U_i(W, Y_1^*, ..., y_{i,d}, Y_{-i,d}^*, \cdots, Y_{\bar{d}}^*)$ for all i and d.

Proposition 1. *Assume that* $|\lambda_d| < 1/\|W\|_\infty$, *where* $\|.\|_\infty$ *denotes the maximum row sum norm for all* $d = 1, ..., \bar{d}$ *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:*

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

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

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

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

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

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

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\widetilde{W}\in\Omega} \exp\{V(\widetilde{W})\}}, \tag{6}$$

where Ω *is the set of all* $m \times m$ *network matrices.*

The uniqueness of the Nash equilibrium uses standard arguments (e.g., Ballester et al., 2006). Whenever $|\lambda_d| < 1/\|W\|_\infty$ for all $d \in (1, ..., \bar{d})$, the best-response functions are contraction mappings, leading to a unique fixed point. 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, i.e., $T_{Y^*}(W) = T(W)$. 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 distribution.[17] This guarantees that the unique strongly efficient network is given by (ii) of Proposition 1.

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 non-decreasing 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.[18] Jackson (2005) also presents an extensive discussion of the type of allocation rules compatible with both strong efficiency and individual stability.

It is worth noting that the expression $\mu_d' Y_d^*(W) - \frac{1}{2}Y_d^{*\prime}(W)Y_d^*(W) + \lambda_d Y_d^{*\prime}(W)WY_d^*(W)$ in the equilibrium value for $V(W)$ reduces to $\frac{1}{2}Y_d^{*\prime}(W)Y_d^*(W)$ when activity d's intensity is uncensored since we can exploit the closed-form solution in Eq. (4). In that case, the incentive effect of activities is always non-negative (recall that $\delta_d \geq 0$). Indeed, as noted by Ballester et al. (2006), this property holds whenever the choice of activities features complementarities.[19] As such, without explicit network preferences, i.e., $v_i(W)$, individuals would want to have as many links as possible, leading to the complete network. 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 further specify individual exogenous heterogeneity via $\mu_{i,d} = x_i\beta_{1d} + \sum_{j=1}^m w_{ij}x_j\beta_{2d} + \alpha_d + \epsilon_{i,d}$, where α_d is a constant term, and $\epsilon_{i,d}$ is the unobserved heterogeneity of i's preferences regarding activity d. Following Proposition 1, the equilibrium for uncensored activities is

[17] Brock and Durlauf (2001) use the same assumption when specifying the social welfare function.

[18] See, in particular, their Theorem 4.

[19] Boucher (2016) studies a model of conformism having an endogenous network. He shows that conformism has non-monotonic effects on the value of network links.

given by:[20]

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

where l_m is the m-dimensional vector of ones, and $\epsilon_d = (\epsilon_{1,d}, \cdots, \epsilon_{m,d})'$. Eq. (7) matches the reduced form of the spatial autoregressive (SAR) model (Bramoullé et al., 2009; Lee et al., 2010; Lin, 2010) for studying social interactions. The coefficient λ_d in Eq. (7) represents the endogenous (peer) effect, which has been the focus of recent literature due to its policy implications (Glaeser et al., 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 from the literature is that the network structure W in Eq. (7) is explicitly endogenous. To understand the source of this endogeneity, recall that from Proposition 1, we have

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\widetilde{W}} \exp\{V(\widetilde{W})\}}.$$

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 ϵ_d. Then, this implies that any shock ϵ_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 the intensity of the activities (e.g., Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016). We also allow for such unobserved variables in the following way: $\epsilon_d = Z\rho_{1d} + WZ\rho_{2d} + \xi_d$, where $Z = (z_1', \cdots, z_m')'$ 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 Eq. (7) can be rewritten as:

$$Y_d^*(W) = (I_m - \lambda_d(W))^{-1} (X\beta_{1d} + WX\beta_{2d} + Z\rho_{1d} + WZ\rho_{2d} + l_m\alpha_d + \xi_d), \tag{8}$$

[20] The parametric form for left-censored $y_{i,d}$ follows a similar structure.

where we assume that $\xi_d \sim \mathscr{N}_m(0, \sigma^2_{\xi_d} I_m)$.

We assume that these unobserved variables may also affect individuals' preferences over links. Specifically, we follow Hsieh and Lee (2016) to introduce the multidimensional individual latent variables $z_i = (z_{i1}, \cdots, z_{i\bar{\ell}})'$ in the network formation process through the local network effect as follows:

$$\psi_{ij} = \gamma_0 + c_i\gamma_1 + c_j\gamma_2 + c_{ij}\gamma_3 + \sum_{\ell=1}^{\bar{\ell}} \gamma_{4\ell}|z_{i\ell} - z_{j\ell}|. \tag{9}$$

The variables c_i and c_j in Eq. (9) 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.[21] In particular, the individual and dyadic characteristics $C = \{(c_i, c_j, c_{ij}) : i = 1, \cdots, m, j = 1, \cdots, m, i \neq j\}$ control for observed homophily in the network formation process (see e.g., Fafchamps and Gubert (2007a,b) in the context of risk-sharing network formation). For exposition purposes, we denote $\gamma = (\gamma_0, \gamma_1', \gamma_2', \gamma_3', \gamma_4')'$.

The variables $|z_{i\ell} - z_{j\ell}|$ for $\ell = 1, \cdots, \bar{\ell}$ in Eq. (9) are meant to capture the homophily on *unobserved* characteristics. We therefore 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. Note also that for identification purposes, we follow Hsieh and Lee (2016) and Hsieh and Van Kippersluis (2018) to impose the assumptions on $z_{i,\ell}$ as follows: (1) the variance of $z_{i,\ell}$ is normalized to one; (2) $z_{i\ell}$ is independent across i and ℓ; (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 $|\gamma_{41}| \geq |\gamma_{42}| \geq \cdots \geq |\gamma_{4\bar{\ell}}|$.[22] Given the role played by the latent variables Z in Eq. (9) for network formation, Z and WZ that appear in Eq. (8) can also be interpreted as control functions for solving endogeneity due to individual and contextual unobserved correlated effects (Fruehwirth, 2014; Hsieh and Van Kippersluis, 2018).

[21]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.

[22]The justification of these identification restrictions can be found in the Supplementary Appendix of Hsieh and Van Kippersluis (2018).

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

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

We discuss the interpretation of each term in Eq. (10) in turn and provide a visualization of each effect in Figure 2.[23]

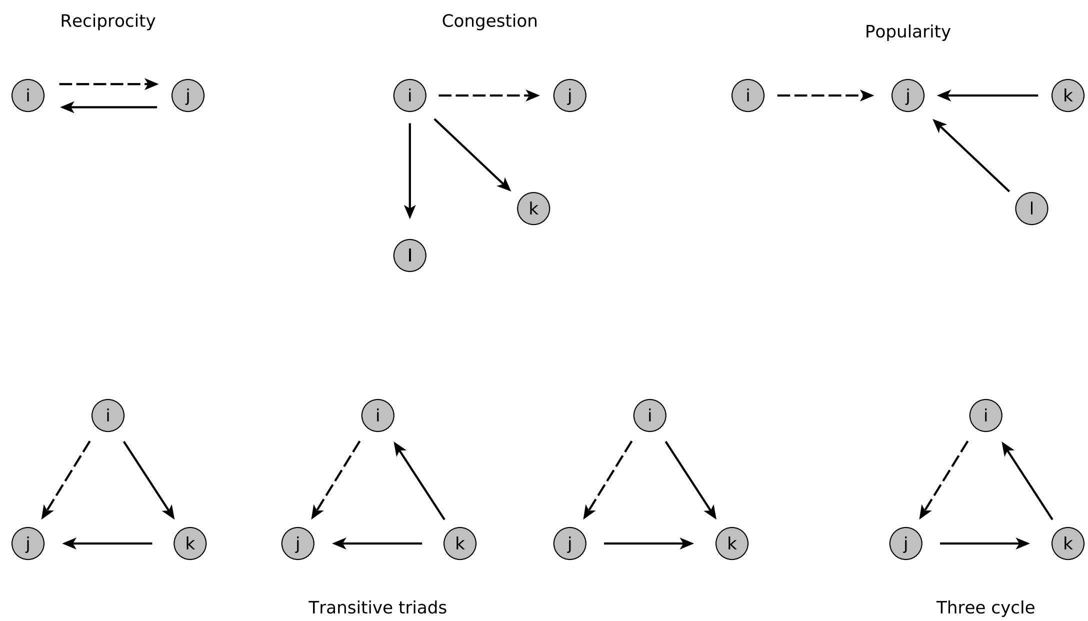
The reciprocity effect implies that (provided that $\eta_1 > 0$) i enjoys more utility from a link with j if j also has a link with i. The congestion effect (provided that either $\eta_2 < 0$ or $\eta_3 < 0$) implies that the value for i of a link with j decreases with the number of links that i has. This may represent the fact that i has limited resources, e.g., 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, i.e., receives many links.

The transitive triads effects include preferences for cliques, i.e., 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 in Figure 2. A similar intuition holds for the three-cycle effect, although as noted by Snijders et al. (2010), the emergence of more three cycles in a network (see the bottom right of Figure 2) implies fewer hierarchical relationships among individuals. As discussed by Davis (1970), social networks usually feature fewer three cycles. We therefore expect η_6 to be negative.[24]

[23]We also provide additional discussion in the Supplementary Appendix D.

[24]This is indeed what we find in our empirical study, see Table 4.

Figure 2: Global network effects



Given the parametric assumption in Eq. (10), we can write:

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

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 η_{51}, η_{52}, and η_{53} are not identified separately from Eq. (11). Hence, we will use $\eta_5 = \eta_{51} + \eta_{52} + \eta_{53}$ hereafter. We further denote $\eta = (\eta_1, \cdots, \eta_6)'$ 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 individuals will not form links to everyone (and thus results in a complete graph). Moreover, as discussed by Snijders et al. (2006), Bhamidi et al. (2011), Chatterjee et al. (2013), and Mele (2017b), we need negative externalities in ERGMs to produce sparser graphs so that they can be distinguished from Erdös-Rény random graphs when the number of individuals increases. In our case, η_2 (or η_3) and η_6 are expected to create such negative externalities, and we confirm that these parameter estimates are significantly negative in our empirical study in Section 5.

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 $g \in \{1, \cdots, G\}$ 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. In Appendix B, we present the analysis for the censored activity outcome case. 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. We refer the interested reader to Supplementary Appendix C for details.

Since both Y_g and W_g are endogenous variables, we focus on the joint likelihood, i.e.,[25]

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

where $\theta_g = (\gamma', \eta', \delta, \lambda, \beta', \rho', \sigma^2_{\xi_g})$. Note that to describe group heterogeneity on activity outcomes at both the mean and variance levels, we use α_g and $\sigma^2_{\xi_g}$ to capture the group fixed effect and group heteroskedasticity, respectively. To adhere to the principle of model parsimony, the other coefficients are assumed to be common across groups. Using the parametric

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

forms assumed in Section 2.4, we can write:

$$\begin{aligned}
P(W_g, Y_g|\theta_g, \alpha_g, Z_g) &= P(Y_g|W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|W_g, \theta_g, \alpha_g, Z_g) \cdot P(W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g, W_g|\theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|\theta_g, \alpha_g, Z_g) \cdot P(W_g|\xi_g, \theta_g, \alpha_g, Z_g) \\
&= |S_g(W_g)| \cdot f(\xi_g|\theta_g, \alpha_g, Z_g) \cdot \frac{\exp(\mathrm{V}(W_g, \xi_g, \theta_g, \alpha_g, Z_g))}{\sum_{\widetilde{W}_g \in \Omega_g} \exp\left(\mathrm{V}(\widetilde{W}_g, \xi_g, \theta_g, \alpha_g, Z_g)\right)},
\end{aligned} \tag{12}$$

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

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

The main challenge in estimating this model is to compute the denominator in Eq. (12). As it sums all possible network structures, its direct evaluation is impossible, even for small-sized networks. For example, in a network with only five individuals, 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, 2018; Chandrasekhar and Jackson, 2014; Mele, 2017b; Boucher and Mourifié, 2017) and can be traced back to the spatial analysis in Besag (1974).

To deal with this problem, several estimation methods have been proposed. The first is the maximum pseudo-likelihood approach (MPL). This approach was first mentioned in Besag (1974) and later applied to the network study in Strauss and Ikeda (1990). Recently, Boucher and Mourifié (2017) have shown that the special case of the estimator proposed in Strauss and Ikeda (1990) can be estimated consistently under an assumption of specific homophily.

Another approach is the Monte Carlo maximum likelihood (MCML) estimation approach that simulates auxiliary networks to approximate the denominator of the exponential distribution function with its simulated counterpart (Geyer and Thompson, 1992). One shortcoming of the MCML approach is that the choice of initial values during the optimization algorithm plays a critical role. The initial values must produce extremely precise estimates of the parameter values; otherwise, the convergence of the algorithm is not ensured (Bartz et al., 2009; Caimo and Friel, 2011). The Robbins-Monro approach, used in Snijders (2002) to simulate auxiliary networks for constructing simulated moments, usually accepts a wide range of initial values that can lead to a convergent algorithm.

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). 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 variable A_g across G groups, i.e., $\{A_g\} := (A_1, \cdots, A_G)$. Now let $y = (\{Y_g\}, \{W_g\})$ and $\theta = (\{\theta_g\}, \{\alpha_g\}, \{Z_g\})$. From Section 3.2, the likelihood function of y, given θ, has the following form: $P(y|\theta) = f(y;\theta)/D(\theta)$, where $D(\theta)$ is an intractable normalizing term.

The standard M-H algorithm to simulate random draws of θ operates as follows: given an old draw θ_{old} one proposes a new draw θ_{new} from a proposal distribution $q(\cdot|\theta_{old})$, and one then updates the old draw to the new draw with an acceptance ratio $\alpha_{MH}(\theta_{new}, \theta_{old})$. Denoting the prior distribution of θ as $\pi(\theta)$, the acceptance ratio is given by:

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

One can see that in Eq. (13), the normalizing terms $D(\theta_{old})$ and $D(\theta_{new})$ do not cancel out; thus the evaluation of the acceptance-rejection criterion in Eq. (13) is intractable.

To solve this problem, Murray et al. (2006) consider including auxiliary variables $\widetilde{y}$ into the acceptance probability, i.e., the acceptance probability can be written as:

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

where $\widetilde{y}$ are simulated from the likelihood function $P(\widetilde{y}|\theta_{new}) = f(\widetilde{y};\theta_{new})/D(\theta_{new})$ with the exact sampling (Propp and Wilson, 1996).

In the conditional acceptance probability of Eq (14), all normalizing terms cancel out, and the remaining terms are computable. This algorithm is called the "exchange algorithm" because a swapping operation between (θ_{old}, y) and $(\theta_{new}, \widetilde{y})$ 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 $q(\theta_{new}|\theta_{old})$ into $q(\theta_{new}|\theta_{old})P(\widetilde{y}|\theta_{new})$.

The exchange algorithm defines a valid Markov chain for simulating from $P(\theta|y)$ (Murray et al., 2006; Liang, 2010; Liang et al., 2016). However, implementing the exchange algorithm is time-consuming because it requires the exact sampling of $\widetilde{y}$ from $P(\widetilde{y}|\theta_{new})$. To save computation time, Liang (2010) proposes a "double M-H algorithm" that utilizes the reversibility condition and shows that when $\widetilde{y}$ is simulated by the M-H algorithm—starting from y with R iterations—the conditional acceptance probability in Eq. (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.[26] With this mixed proposal, Mele (2017b) shows 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 θ from $P(\theta|y)$ requires simulating the auxiliary variable $\widetilde{y}$. In our context, however, the auxiliary activity variables $\{\widetilde{Y}_g\}$ in $\widetilde{y}$

[26] Mele (2017b) uses 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.

are redundant during simulation as they can be replaced by either a closed-form function (as in Eq. (4)) or a contraction mapping (as in Eq. (5)) of auxiliary networks and estimated individual heterogeneity. Thus, we can simplify $\widetilde{y}$ to $\widetilde{w} = \{\widetilde{W}_g\}$ and modify the conditional acceptance probability in Eq. (14) to:

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

To evaluate $\alpha(\widetilde{w}, \theta_{new}, \theta_{old})$ in Eq. (15), we need only to simulate the auxiliary networks $\widetilde{w}$ from their probability density function $P(\widetilde{w}|\theta_{new}) = f(\widetilde{w}; \theta_{new})/D(\theta_{new})$ that shares the same normalizing term, i.e., $D(\theta_{new})$ as $P(\widetilde{y}|\theta_{new})$.

3.4 Posterior distributions and the MCMC

We now present the MCMC procedure. Although we regard unobserved latent variables $\{Z_g\}$ as individual random effects, we follow Tanner and Wong (1987) and Albert and Chib (1993) to proceed with Bayesian data augmentation that treats $\{Z_g\}$ basically as parameters to be estimated.[27] By Bayes' theorem, the joint posterior distribution of the parameters and unobservables in the model can be written as:

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

where $\pi(\cdot)$ represents the density function of the prior distribution, and 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 Section 3.5.

Obtaining draws directly from the joint posterior distribution of Eq. (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

[27] The idea is that instead of trying to approach the posterior distribution of the parameters θ (i.e., $P(\theta|y) = \int_Z P(\theta|y, Z)P(Z)$, where $P(\theta|y, Z) \propto P(y|Z, \theta)\pi(\theta)$), we try to approach the posterior distribution of the parameters *and* the latent variables Z (i.e., $P(\theta, Z|y) \propto P(y|Z, \theta)\pi(\theta, Z)$). It turns out that this is much easier for simulating draws. Moreover, the posterior distribution of θ can simply be recovered by integrating over Z, i.e., $p(\theta|y) = \int_Z P(\theta, Z|y)$.

the Gibbs sampler in the Supplementary Appendix C. A subset of the parameters admit close-form conditional posterior distributions; therefore, they can be drawn directly to improve convergence. However, this is not true for most parameters, and we must therefore use the double M-H algorithm discussed in Section 3.3 to draw from these 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 activity—is presented in Supplementary Appendix C.

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

(a) *Propose $\widetilde{\Xi}_g$ from $q(\Xi_g|\Xi_g^{(t-1)})$.*

(b) *Compute the residuals $\widetilde{\xi}_g$ from the activity intensity equation conditional on the proposed $\widetilde{\Xi}_g$ and the value of the other unknown parameters and variables at iteration $t-1$, i.e., $(\theta \setminus \Xi_g)^{(t-1)}$.*

(c) *Simulate auxiliary network $\widetilde{W}_g$. Set the initial auxiliary network equal to the observed network $\widetilde{W}_g^{(0)} = W_g$. Conditional on $\widetilde{\xi}_g$, $\widetilde{\Xi}_g$, and $(\theta \setminus \Xi_g)^{(t-1)}$, use R repetitions of the following procedure:*[28]

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

[28] The 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 $P_{inv} = 0.01$ in the following simulation and empirical studies.

the probability

$$\alpha_{MH,local}(\widetilde{w}_{ij,g}^{(r)}, \widetilde{w}_{ij,g}^{(r-1)}) = \min\left\{1, \frac{P(\widetilde{w}_{ij,g}^{(r)}, \widetilde{W}_{-ij,g}^{(r-1)}|\widetilde{\xi}_g, \widetilde{\Xi}_g, (\theta \setminus \Xi_g)^{(t-1)})}{P(\widetilde{w}_{ij,g}^{(r-1)}, \widetilde{W}_{-ij,g}^{(r-1)}|\widetilde{\xi}_g, \widetilde{\Xi}_g, (\theta \setminus \Xi_g)^{(t-1)})}\right\}, \quad (17)$$

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

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

$$\alpha_{MH,global}(\widetilde{W}_g^{(r)}, \widetilde{W}_g^{(r-1)}) = \min\left\{1, \frac{P(\widetilde{W}_g^{(r)}|\widetilde{\xi}_g, \widetilde{\Xi}_g, (\theta \setminus \Xi_g)^{(t-1)})}{P(\widetilde{W}_g^{(r-1)}|\widetilde{\xi}_g, \widetilde{\Xi}_g, (\theta \setminus \Xi_g)^{(t-1)})}\right\}, \quad (18)$$

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

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

$$\begin{aligned} &\alpha_{MH,\Xi_g}(\widetilde{w}, \widetilde{\Xi}_g, \Xi_g^{(t-1)}) \\ &= \min\left\{1, \frac{\pi(\widetilde{\Xi}_g)P(y|\widetilde{\Xi}_g, (\theta \setminus \Xi_g)^{(t-1)})q(\Xi_g^{(t-1)}|\widetilde{\Xi}_g)}{\pi(\Xi_g^{(t-1)})P(y|\Xi_g^{(t-1)}, (\theta \setminus \Xi_g)^{(t-1)})q(\widetilde{\Xi}_g|\Xi_g^{(t-1)})} \cdot \frac{P(\widetilde{w}|\Xi_g^{(t-1)}, (\theta \setminus \Xi_g)^{(t-1)})}{P(\widetilde{w}|\widetilde{\Xi}_g, (\theta \setminus \Xi_g)^{(t-1)})}\right\}, \end{aligned} \quad (19)$$

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

Note that Step (c) of Algorithm 1 represents the simulation of the auxiliary variable $\widetilde{w} = \{\widetilde{W}_g\}$ discussed in Section 3.3. It is worth noting that some authors have used similar estimation strategies. For example, Mele (2017b) uses a special case of Algorithm 1 where $\theta = \{\theta_g\}$ and Step (b) is left out. Note also that Algorithm 1 can be adapted to the case where $y_{i,d}$ is left-censored. To do so, one only has to incorporate the additional latent variable $\ddot{Y}_g$ to the list in Ξ_g. The formal algorithm is presented in Supplementary Appendix C.

3.5 Prior distributions

We assume independence across prior distributions of common parameters, group effects, and latent variables, namely $\pi(\theta) = \pi_1(\{\theta_g\})\pi_2(\{\alpha_g\})\pi_3(\{Z_g\})$. Corresponding to the Gibbs sampling that divides parameters in $\{\theta_g\}$ properly into subgroups, we define prior distributions for parameters and other unknown variables in the model as follows:

(i) Latent variables in both network formation and activity intensity equations,

$$z_{i,g}|\mu_{z,g} \sim \mathcal{N}_{\bar{\ell}}(\mu_{z,g}, I_{\bar{\ell}}) \text{ and } \mu_{z,g} \sim \mathcal{N}_{\bar{\ell}}(0, \varsigma I_{\bar{\ell}}) \quad i = 1, \cdots, m_g; \quad g = 1, \cdots, G.$$

(ii) Coefficients of network formation utility,

$$\phi = (\gamma', \eta', \delta) \sim \mathcal{N}_{2\bar{s}+\bar{q}+\bar{\ell}+\bar{h}+1}(\phi_0, \Phi_0 I_{2\bar{s}+\bar{q}+\bar{\ell}+\bar{h}+1}),$$
$$\phi \in O = \{\phi \in \mathbb{R}^{2\bar{s}+\bar{q}+\bar{\ell}+\bar{h}+1} || \gamma_{41}| \geq |\gamma_{42}| \geq \cdots \geq |\gamma_{4\bar{\ell}}|, \delta \geq 0\}.$$

(iii) Coefficients of endogenous effect in the activity intensity equation,

$$\lambda \sim U[-1/\Delta_G, 1/\Delta_G].$$

(iv) Coefficients of own and contextual effects in the activity intensity equation,

$$\beta \sim \mathcal{N}_{2k}(\beta_0, B_0 I_{2k}).$$

(v) Coefficients of own and contextual correlated effects in the activity intensity equation,

$$\rho \sim \mathcal{N}_{2\bar{\ell}}(\rho_0, R_0 I_{2\bar{\ell}}).$$

(vi) Variance of disturbances in the economic activity equation,

$$\sigma^2_{\xi_g} \sim \mathcal{IG}(\tfrac{\kappa_0}{2}, \tfrac{\nu_0}{2}), \quad g = 1, \cdots, G.$$

(vii) Group fixed effects in the activity intensity equation,

$$\alpha_g \sim \mathcal{N}(\alpha_0, A_0), \quad g = 1, \cdots, G.$$

Most of the above prior distributions are conjugate priors used commonly in the Bayesian literature. However, following Hsieh and Lee (2016) and Hsieh and Van Kippersluis (2018), we set up a hierarchical prior for $z_{i,g}$. That is, we assume $z_{i,g}$ is normally distributed having a unit variance and a prior mean equal to $\mu_{z,g}$, reflecting the identification restrictions on $z_{i,g}$, as presented in Section 2.4. Then, we further assume that $\mu_{z,g}$ is normally distributed having a hyperprior mean of zero and a hyperprior variance of ς. As a result, latent variables $\{Z_g\}$ only add G "real" parameters ($\{\mu_{z,g}\}$) into the estimation procedure (excluding their coefficients γ_4s).

Note that the prior of ϕ is constrained to a parameter space O in which $|\gamma_{41}| \geq |\gamma_{42}| \geq \cdots \geq |\gamma_{4\bar{\ell}}|$ for the identification of latent variables Z. We also assume that δ is non-negative to maintain the coherence with our microeconomic model. The prior on λ is also constrained on $[-1/\Delta_G, 1/\Delta_G]$, where $\Delta_G = \max_g \|W_g\|_\infty$, to ensure that Proposition 1 holds and that the equilibrium is always unique and generates a well-defined data generating process (Kelejian and Prucha, 2010). The use of a uniform prior follows Smith and LeSage (2004).

The prior of $\sigma^2_{\xi_g}$ is specified as an inverse Gamma ($\mathscr{IG}$) distribution with the shape and scale parameters governed by $\frac{\kappa_0}{2}$ and $\frac{\nu_0}{2}$. In the specification of (vii), we treat the group effects α_g as fixed effects with the hyperparameters α_0 and A_0 fixed in their prior distributions. The distinction between fixed and random effects in a Bayesian approach lies on prior assignment at the second and third levels of hierarchy (Lancaster, 2004; Rendon, 2013). For a fixed-effect model, a Bayesian approach updates distributions of fixed-effect parameters, whereas a random-effect model updates distributions of hyperparameters in the prior distribution of random-effect parameters. If it is preferred to model the correlation between covariates and group effects explicitly, one may follow Mundlak (1978) to have a correlated random-effect specification that allows the mean of the random effect (i.e., the mean hyperparameter in the prior distribution of the random effect) to be a linear function of covariates (e.g., Boucher and Goussé,2019). To determine if there is any impact due to the specification of random group effects, we also examine the estimation results of our model based on the correlated random-effect specification for a robustness check.[29]

4 Simulation study

In this section, we conduct a simulation study to examine the finite sample performance of the Bayesian MCMC sampler proposed in Section 3. The simulation study is designed to accommodate four different purposes.

First, we carry out a Monte Carlo experiment (with 100 repetitions) to demonstrate that the MCMC sampler can successfully recover the true parameters from the artificially

[29]We specify α_g as follows:

$$\alpha_g = \overline{X}_g \beta_3 + \overline{Z}_g \rho_3 + \zeta_g, \ \ \zeta_g \sim \mathscr{N}(0, \sigma^2_{\alpha,c}),$$

where $\overline{X}_g$ and $\overline{Z}_g$ are, respectively, the group averages of X_g and Z_g. The β_3, ρ_3, and σ^2_α are unknown parameters, and we also specify prior distributions for them such that $\beta_3 \sim \mathscr{N}_k(\beta_0, B_0 I_k)$, $\rho_3 \sim \mathscr{N}_{\bar{\ell}}(\rho_0, R_0 I_{\bar{\ell}})$, and $\sigma^2_\alpha \sim \mathscr{IG}(\frac{\kappa_0}{2}, \frac{\nu_0}{2})$. As a result, we form a hierarchical prior for α_g where the prior mean and the prior variance of α_g follow other prior distributions.

generated network data. *Second*, the same Monte Carlo experiment is used to show the issue of model misspecifications and the consequent estimation biases. *Third*, using samples from one simulation repetition, we begin the MCMC sampler with different initial values to see if the Markov chains converge toward the true values within a reasonable amount of draws, i.e., examining the issues of non-convergence and the slow mixing of the Markov chain. *Fourth*, we also report the computation time required by network samples of different sizes to offer users additional information on the feasibility of our approach when applied to their own network data.

We design the data generating process (DGP) throughout the simulation based on the exponential distribution of Eq. (6). Continuous (uncensored) activity variables are generated by the activity intensity equation of Eq. (7), and the censored activity variables are generated by Eq. (5). We set the network size for the uncensored case at 30 and the censored case at 40—to compensate the loss of information due to censoring—and fix the number of networks at 30, i.e., there are 900 and 1,200 individual observations for the uncensored and censored cases, respectively, at each simulation repetition.

In the activity intensity equations, we generate the exogenous variable X from a normal distribution $\mathcal{N}(0,4)$. The group fixed effects are generated from $\mathcal{N}(3,1)$ for the uncensored case and $\mathcal{N}(-1.5,1)$ for the censored case. The disturbance term ξ is generated from $\mathcal{N}(0,0.5)$. The latent variable Z is specified as one dimensional and generated from $\mathcal{N}(0,1)$. For the other effects on network formation, the local network effect is specified based on Eq. (9). We include a constant term and a dyad-specific exogenous variable C_{ij} that is generated as follows: first, we draw two uniform random variables from $U(0,1)$, denoted as U_1 and U_2. If U_1 and U_2 are both larger than 0.7 or less than 0.3, then we set C_{ij} to one. Otherwise, we set C_{ij} to zero. We also include the distance of latent variables $|z_i - z_j|$ as part of the local network effect. The global network effects are specified according to Eq. (10). All true parameter values of the DGP are reported in the second column of Tables 1 and 2.

Each artificial network W is simulated by the M-H algorithm from an empty network based on the exponential distribution of Eq. (6). The following steps are implemented iteratively corresponding to the local and global updates in Step (c) of the pseudo-MCMC algorithm in Section 3.4. Activity intensity variables are simulated along with networks. The

M-H algorithm runs through the entire network for a total of 10,000 iterations, and realizations of the network and the activity intensity variables from the last iteration are used as the data. The networks generated out of the design have an average out-degree of 3.6978 for the uncensored case and 2.6458 for the censored case; the average density is 0.1275 (uncensored) and 0.0678 (censored), and the average clustering coefficient is 0.0493 (uncensored) and 0.0263 (censored).[30] These network statistics are comparable to those of the empirical samples in Section 5. The generated uncensored activity variable has a mean of 4.0767, and the censored variable has a mean of 1.1923. A total of 21.67% of the observations are censored.

To estimate the model, a total of 50,000 draws were simulated using the double M-H algorithm discussed in Section 3.3. The values of hyperparameters in the prior distributions are set as follows: $\phi_0 = 0$; $\Phi_0 = 10$; $\beta_0 = 0$; $B_0 = 10$; $\rho_0 = 0$; $R_0 = 10$; $\sigma_0 = 0$; $\Sigma_0 = 10$; $\alpha_0 = 0$; $A_0 = 400$. These specified values of hyperparameters are chosen to form very flat prior densities over the range of parameter spaces so that estimation results are less influenced by our choice of priors. We discard the first 10,000 draws and use the remaining 40,000 draws to compute the posterior mean as a point estimate. We now summarize our findings from the simulation study.

First, we report the Monte Carlo simulation results for uncensored and censored activities in Tables 1 and 2, respectively. For both activity cases, we find that the proposed Bayesian estimation can recover successfully the true parameter values when the correct model—"full" model—is used. We also estimate four misspecified models: the "no latent" model that ignores the latent variable from network formation and activity intensity; the "no global" model that ignores the global network effects from network formation; the 'latent only" model that only includes the latent variable; and the "activity only" model that regards networks as exogenously given and only estimates the activity intensity equation. The results reveal

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

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

different levels of estimation biases in these misspecified models.

When ignoring the latent variable, the "no latent" results show significant upward biases of the endogenous peer effect (λ) and some of the global network effects (e.g., η_1, η_2, and η_5), and there is a significant downward bias on the incentive effect (δ). This result demonstrates that omitting latent variables that affect both activity intensity and network formation not only causes an upload bias on the estimate of endogenous peer effect (Hsieh and Lee, 2016), but also biases the estimates of other network effects.

When ignoring the global network effects, the "no global" results show that the estimated incentive effect confounds with the uncontrolled global network effects and becomes upward biased. Meanwhile, since the incentive and endogenous peer effects are highly interdependent through the activity intensity, the upward bias on the estimated incentive effect therefore leads to the downward bias on the estimated endogenous peer effect. These results reveal the necessity of controlling the global network effects in our network formation model.

When ignoring the global network effects and the incentive effect, the estimate of endogenous peer effect is also upward biased in the "latent only" results. This highlights the importance of the new avenue explored with our network formation model, i.e., the incentive effect, in which the issue of network endogeneity on social interactions can be formulated as well as adding to the existing literature related to joint-modeling network formation and the interactions with latent variables (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2016). Finally, the results of the "activity only" model display the most severe upward bias on the endogenous peer effect among the four misspecified models due to uncontrolled network endogeneity.

Next, we inspect the convergence of the Markov chain in our Bayesian MCMC estimation under different initial values to determine if there are issues of non-convergence and slow mixing. To implement this inspection, we take the network and activity samples from the first simulation repetition of the above Monte Carlo experiment and estimate the true model with the proposed MCMC procedure in Section 3.3. We focus on the two important parameters of the model—the endogenous peer effect (λ) and the incentive effect (δ)—for the inspection and assign different initial values to begin the MCMC sampling. To keep the exercise tractable, we let the sampling of other parameters begin with the true values.

For illustration purposes, we consider the case of uncensored activity and present the results of inspection by the trace plots of MCMC draws in Figure 3 and Figure 4. The plots for the case of censored activity are relegated to the Supplementary Appendix Figures E.1 and E.2.

To interpret these trace plots, we take the endogenous peer effect λ in Figure 3 as an example. The true value of λ in the simulation is set at 0.05 (see Table 1). Accordingly, we assign five evenly spaced values between 0 and 0.1, namely 0, 0.025, 0.05, 0.075, and 0.1, to start the MCMC sampling. With these five different initial values, we run five independent MCMC samplings (for all parameters) and plot the first 5,000 draws of λ in each chain. The plots in Figure 3 show that regardless of the different initial values, the draws of λ converge and stabilize swiftly near the true values. We also find a similar pattern in the other figures for λ in the censored outcome and for δ in both the censored and uncensored outcomes; this occurs despite the convergence of δ requiring slightly more draws than λ.[31]

Finally, we use this simulation environment to illustrate the computational cost of our estimation algorithm. We focus on the case of uncensored activity and generate artificial data with different network sizes of 20, 40, 60, 80, and 100. We fix the number of network groups at 30 given that the number of groups is less of a concern for computation because we can easily digest the cost of many groups by applying the parallel computation at the group level.

In Figure 5, we show the computation cost, measured by the average CPU time (in seconds) of one MCMC iteration for the five studied models (in Table 1); these models include the full model, three nested network formation models, and the activity intensity equation alone. This timing task is done using a desktop PC having an Intel i7-6700 CPU (4.00 GHz). We see that the computational cost increases exponentially with network size whenever estimation of the corresponding model requires the double M-H algorithm. Taking the median network size (60) in this simulation—which is also close to the average network size in the empirical study of Section 5—as a reference, completing the estimation of the full model with 100,000 MCMC iterations will require roughly 155 hours, a manageable amount

[31]The draws of λ and δ have posterior means close to but not exactly equal to the true parameter values. This occurs as we use only the sample from one simulation repetition, and therefore sampling error exists.

of time.

5 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 usually 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.[32] 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 do observe a significant fraction of non-smokers.

In the context of social interactions, students' academic performance and smoking behavior are studied extensively as they have important long-term consequences on students'

[32] Discussions about how academic performance and smoking affect friendship selections can be found in, e.g., Kiuru et al. (2010), Lomi et al. (2011), Flashman (2012), Schaefer et al. (2013). Other activities may affect friendship choices. We focus on academic performance and smoking because they are the key subjects of interest discussed in social interaction studies.

future lives and health. Studies of peer effects on students' academic performance, e.g., Hoxby (2000), Sacerdote (2001), Hanushek et al. (2003), and Zimmerman (2003) use the linear-in-means model; whereas Calvó-Armengol et al. (2009), Lin (2010), Boucher et al. (2014), and Liu et al. (2014) use the network interactions model. For studies of peer effects on students' smoking behaviors, evidence of peer effects can be found in Gaviria and Raphael (2001), Powell et al. (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) consider further the problem of endogenous friendship selection of peer effects on economic activities by modeling unobservables in both the network interaction 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 when students form friendships.

5.1 Data summary

As mentioned previously, 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 those 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 1,036 respondents from 15 schools (groups).[33] 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 Eq. (9), 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. For the variables used in the activity intensity equation of Eq. (8), 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 et al. (2010), and Hsieh and Lee (2016) to choose the independent variables. A complete list of variables and their summary statistics are provided in Table 3.

5.2 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

[33] To clarify, we do not use the Add Health saturation sample (Udry, 2003) having 16 schools. In the 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 any variables from the in-home survey, we do not use the saturation sample.

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 E.3 and E.4.[34] In addition, modeling multiple activity outcomes allows us to explicitly control the correlation of the error terms between the activity intensity equations.

Similar to the simulation study in Section 4, we compare the estimation results of the full model with the results of several possibly misspecified models to see how each model misspecification affects estimates of the full model, particularly for the estimate of endogenous peer effect (λ) on activity outcomes. We present the estimation results in Table 4. From the first to the fifth columns respectively, they are 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 the activity intensity equation alone assuming exogenous network links.[35]

Network formation

The results from local network effects in the full model are as follows. Staying in a same school for a longer time has a significant positive effect on receiving (but a negative effect on sending out) friendship nominations. The exogenous dyad-specific effects are all positive and significant, where the effect of the same age (0.3675) is strongest, followed by the effect of the same sex (0.3471), and then the effect of the same race (0.3116). We find that the distances of latent variables have significant negative effects on network formation, confirming the existence of homphily with respect to unobservables (Hoff et al., 2002; Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016).[36]

[34]Comparing the results of the multiple-activity model in Table 4 with the results of the single-activity models in 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 the values in Table 4. These differences illustrate the potential concern of having omitted-variable biases when activity outcomes are modeled separately.

[35]In Table 4, the mean and the standard deviation (in parentheses) of the MCMC posterior draws are reported as the point estimate of each parameter. We set the hyperparameters in the prior distributions to be identical to those used in the simulation study presented in Section 4.

[36] Because the exact likelihood value for the full model in Eq. (12) is unavailable due to the intractable

For the global network effects, we find a positive and strong reciprocity effect (1.4309), which is consistent with findings in the literature (Snijders et al., 2010; Badev, 2018; 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 is positive (0.2699), while the quadratic effect is negative (-0.0247). This result confirms our conjecture that limited resources, e.g., time, energy, and money, may constrain students from making too many friends. The popularity effect is small (0.0049) and insignificant.

The positive and strong transitive-triads effect (0.4715) shows that students value transitive relationships. As expected, the three-cycles effect is negative (-0.2071). As discussed in Snijders et al. (2010), this reveals a certain degree of local hierarchy among students. Our estimated results of the global network effects confirm that there are non-trivial negative externalities on link formation that distinguish our model from the Erdös-Rény random graph model.

The incentive effect from GPA (0.2145) is strong and significant. Therefore for high school students in our sample, the potential interaction benefit from their friends' academic achievement influences their friendship decisions. In contrast, the incentive effect from smoking (0.0197) is small and insignificant. Hence, it implies that students in our sample barely consider the interaction benefit from smoking as a factor in their friendship decisions.

Network interactions on GPA

Our main finding for the network interactions on academic performance is that, by controlling network endogeneity through the latent variables and the incentive effect, the es-

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 et al. (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. Dimension three is chosen due it having the smallest AICM value.

timated endogenous effect 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.

We interpret this estimated endogenous peer effect as follows. Through interactions, an individual could raise their GPA by 0.0177 units when any of his or her friends improves their own GPA by one unit. Also note that this effect grows with the number of friends. The more friends that an individual has, the stronger the effect they receive. 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 4 show that correction of bias comes from both the incentive effect and the unobserved latent characteristic variables. When the model only contains the latent variables (see the results in the fourth column), only 59.48% of the observed bias is corrected.[37] When the model only controls for the incentive effect (see results in the second column), we correct for 92.16% of the observed bias. Our empirical results confirm the findings of the simulation study in Section 4 that omitting the global network effects from the network formation model can result in an upward bias on the estimated incentive effect (see results in the third column) and thus indirectly cause pressure on biasing the estimated endogenous peer effect 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 that 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 contextual effects for students that are either Black, Asian, other race, living with both parents, have a mother with less

[37] This percentage is obtained by dividing the difference of estimated λ's between the fourth and the fifth columns with the difference of estimated λ's between the first and the fifth columns.

than a high-school level of education, or that 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.

Network interactions on smoking For the smoking outcome, we observe that the estimated endogenous (peer) effect 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 for smoking.

We nonetheless see that the correction of the 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). Our estimate implies that the social multipliers have a respective average and standard deviation of 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 between 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 Footnote 29 of Section 3.5. The estimation results are available in Supplementary Appendix Table E.1. We find the coefficient estimates of local, global network effects and the incentive effects, and the endogenous peer effects in the activity intensity equations remain similar to those in Table 4. Since all of the group averages on X_g and Z_g used in capturing the mean of correlated random effects have insignificant effects, there are also 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.

6 Conclusion

Researchers are interested in network structures to analyze the impact of these structures on outcomes. As mentioned in Jackson (2010, Section 5), if networks only serve as conduits for diffusion, e.g., diseases or ideas, their impact on outcomes is 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 these 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 outcome equations.

Our empirical results show that American high school students regard the interaction benefit from academic learning as a significant incentive for forming friendships, whereas the incentive effect of smoking is not significant. 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 is appropriate for a school setting, it is likely questionable in other economic contexts.

Second, we abstract from outcome games using multiple game equilibria. In the paper, we circumvent this issue by focusing on continuous outcome variables. In a multiple-equilibria setting, 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, e.g., criminal networks, physician referral networks, or academic coauthor networks.

References

Albert, James H and Siddhartha Chib (1993) "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, Vol. 88, pp. 669–679.

Badev, Anton (2018) "Discrete games in endogenous networks: Theory and policy," *working paper*.

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006) "Who's who in networks. Wanted: The key player," *Econometrica*, Vol. 74, pp. 1403–1417.

Ballester, Coralio, Yves Zenou, and Antoni Calvó-Armengol (2010) "Delinquent networks," *Journal of the European Economic Association*, Vol. 8, pp. 34–61.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson (2013) "The Diffusion of Microfinance," *Science*, Vol. 341, URL: http://science.sciencemag.org/content/341/6144/1236498, DOI: http://dx.doi.org/10.1126/science.1236498.

Bartz, Kevin, Jun Liu, and Joseph Blitzstein (2009) "Monte Carlo maximum likelihood for exponential random graph models: From snowballs to umbrella densities," *working paper*.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen (2009) "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections," *Quarterly Journal of Economics*, Vol. 124, pp. 105–147.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa (2008) "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, Vol. 116, pp. 1150–1196.

Besag, Julian (1974) "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 36, pp. 192–236.

Bhamidi, Shankar, Guy Bresler, Allan Sly et al. (2011) "Mixing time of exponential random graphs," *Annals of Applied Probability*, Vol. 21, pp. 2146–2170.

Boucher, Vincent (2015) "Structural homophily," *International Economic Review*, Vol. 56, pp. 235–264.

——— (2016) "Conformism and self-selection in social networks," *Journal of Public Economics*, Vol. 136, pp. 30–44.

Boucher, Vincent, Yann Bramoullé, Habiba Djebbari, and Bernard Fortin (2014) "Do peers affect student achievement? Evidence from Canada using group size variation," *Journal of Applied Econometrics*, Vol. 29, pp. 91–109.

Boucher, Vincent and Bernard Fortin (2016) "Some Challenges in the Empirics of the Effects of Networks," in Y. Bramoullé, A. Galeotti, and B. W. Rogers eds. *The Oxford Handbook on the Economics of Networks*, Oxford: Oxford University Press, Chap. 12, pp. 277–302.

Boucher, Vincent and Marion Goussé (2019) "Wage Dynamics and Peer Referrals," *forthcoming, Review of Economic Dynamics*.

Boucher, Vincent and Ismael Mourifié (2017) "My friend far, far away: a random field approach to exponential random graph models," *Econometrics Journal*, Vol. 20, pp. S14–S46.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009) "Identification of peer effects through social networks," *Journal of Econometrics*, Vol. 150, pp. 41–55.

Bramoullé, Yann and Bernard Fortin (2009) "The econometrics of social networks," *Cahier de recherche/Working Paper*, Vol. 9, p. 13.

Bramoullé, Yann, Rachel Kranton, and Martin D'amours (2014) "Strategic interaction and networks," *American Economic Review*, Vol. 104, pp. 898–930.

Brock, William A and Steven N Durlauf (2001) "Discrete choice with social interactions," *The Review of Economic Studies*, Vol. 68, pp. 235–260.

Caimo, Alberto and Nial Friel (2011) "Bayesian inference for exponential random graph models," *Social Networks*, Vol. 33, pp. 41–55.

Calvo-Armengol, Antoni and Matthew O. Jackson (2004) “The effects of social networks on employment and inequality,” *American Economic Review*, Vol. 94, pp. 426–454.

Calvó-Armengol, Antoni and Matthew O. Jackson (2007) “Networks in labor markets: Wage and employment dynamics and inequality,” *Journal of Economic Theory*, Vol. 132, pp. 27–46.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009) “Peer effects and social networks in education,” *Review of Economic Studies*, Vol. 76, pp. 1239–1267.

Caulier, Jean-François, Ana Mauleon, and Vincent Vannetelbosch (2015) “Allocation rules for coalitional network games,” *Mathematical Social Sciences*, Vol. 78, pp. 80–88.

Chandrasekhar, Arun G and Matthew O Jackson (2014) “Tractable and consistent random graph models,” *NBER working paper*.

Chandrasekhar, Arun and Randall Lewis (2011) “Econometrics of sampled networks,” *Unpublished manuscript, MIT.[422]*.

Chatterjee, Sourav, Persi Diaconis et al. (2013) “Estimating and understanding exponential random graph models,” *Annals of Statistics*, Vol. 41, pp. 2428–2461.

Chib, Siddhartha and Edward Greenberg (1996) “Markov chain Monte Carlo simulation methods in econometrics,” *Econometric Theory*, Vol. 12, pp. 409–431.

Chib, Siddhartha and Srikanth Ramamurthy (2010) “Tailored randomized block MCMC methods with application to DSGE models,” *Journal of Econometrics*, Vol. 155, pp. 19–38.

Christakis, Nicholas A. and James H. Fowler (2007) “The spread of obesity in a large social network over 32 years,” *New England Journal of Medicine*, Vol. 357, pp. 370–379.

Christakis, Nicholas A., James H. Fowler, Guido W. Imbens, and Karthik Kalyanaraman (2010) “An empirical model for strategic network formation,” *NBER working paper*.

Clark, Andrew E and Youenn Lohéac (2007) “It wasnt me, it was them! Social influence in risky behavior by adolescents,” *Journal of Health Economics*, Vol. 26, pp. 763–784.

Comola, Margherita (2007) "The network structure of informal arrangements: evidence from rural Tanzania," *Barcelona, Universitat Pompeu Fabra, mimeo.*

Conley, Timothy G. and Christopher R. Udry (2010) "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, Vol. 100, pp. 35–69.

Conley, Timothy and Christopher Udry (2001) "Social learning through networks: The adoption of new agricultural technologies in Ghana," *American Journal of Agricultural Economics*, Vol. 83, pp. 668–673.

Davis, James A (1970) "Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices," *American Sociological Review*, pp. 843–851.

De Paula, Áureo, Seth Richards-Shubik, and Elie Tamer (2018) "Identifying preferences in networks with bounded degree," *Econometrica*, Vol. 86, pp. 263–288.

Dutta, Bhaskar and Matthew O Jackson (2000) "The stability and efficiency of directed communication networks," *Review of Economic Design*, Vol. 5, pp. 251–272.

Fafchamps, Marcel and Flore Gubert (2007a) "The formation of risk sharing networks," *Journal of development Economics*, Vol. 83, pp. 326–350.

——— (2007b) "Risk sharing and network formation," *American economic review*, Vol. 97, pp. 75–79.

Flashman, Jennifer (2012) "Academic achievement and its impact on friend dynamics," *Sociology of Education*, Vol. 85, pp. 61–80.

Fletcher, Jason M (2010) "Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects," *Health Economics*, Vol. 19, pp. 466–484.

Fowler, James H. and Nicholas 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*, Vol. 27, pp. 1400–1405.

Frank, Ove and David Strauss (1986) "Markov graphs," *Journal of the American Statistical Association*, Vol. 81, pp. 832–842.

Fruehwirth, Jane Cooley (2014) "Can achievement peer effect estimates inform policy? A view from inside the black box," *Review of Economics and Statistics*, Vol. 96, pp. 514–523.

Gaviria, Alejandro and Steven Raphael (2001) "School-based peer effects and juvenile behavior," *Review of Economics and Statistics*, Vol. 83, pp. 257–268.

Geweke, John (1992) "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics*, Citeseer.

Geyer, Charles J (1991) "Markov chain Monte Carlo maximum likelihood," in *in Computer Science and Statistics: Proceedings of the 23rd symposium on the Interface*: Interface Foundation of North America.

Geyer, Charles J and Elizabeth A Thompson (1992) "Constrained Monte Carlo maximum likelihood for dependent data," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 657–699.

Glaeser, Edward L, Bruce I Sacerdote, and Jose A Scheinkman (2003) "The social multiplier," *Journal of the European Economic Association*, Vol. 1, pp. 345–353.

Goldsmith-Pinkham, Paul and Guido W. Imbens (2013) "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, Vol. 31, pp. 253–264.

Graham, Bryan S (2017) "An econometric model of network formation with degree heterogeneity," *Econometrica*, Vol. 85, pp. 1033–1063.

Handcock, Mark S., Adrian E. Raftery, and Jeremy M. Tantrum (2007) "Model-based clustering for social networks," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 170, pp. 301–354.

Hanushek, Eric A, John F Kain, Jacob M Markman, and Steven G Rivkin (2003) "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, Vol. 18, pp. 527–544.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock (2002) "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, Vol. 97, pp. 1090–1098.

Hoxby, Caroline (2000) "Peer effects in the classroom: Learning from gender and race variation."

Hsieh, Chih-Sheng and Lung Fei Lee (2016) "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, Vol. 31, pp. 301–319.

Hsieh, Chih-Sheng and Hans Van Kippersluis (2018) "Smoking Initiation: Peers and Personality," *Quantitative Economics*, Vol. 9, pp. 825–863.

Jackson, Matthew O (2005) "Allocation rules for network games," *Games and economic behavior*, Vol. 51, pp. 128–154.

Jackson, Matthew O. (2010) "An overview of social networks and economic applications," *The Handbook of Social Economics*, Vol. 1, pp. 511–85.

Jackson, Matthew O and Alison Watts (2002) "The evolution of social and economic networks," *Journal of Economic Theory*, Vol. 106, pp. 265–295.

Jackson, Matthew O and Asher Wolinsky (1996) "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, Vol. 71, pp. 44–74.

Jochmans, Koen (2018) "Semiparametric analysis of network formation," *Journal of Business & Economic Statistics*, Vol. 36, pp. 705–713.

Johnsson, Ida and Hyungsik Roger Moon (2016) "Estimation of peer effects in endogenous social networks: control function approach," *working paper.*

Kelejian, Harry H and Ingmar R Prucha (2010) "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of Econometrics*, Vol. 157, pp. 53–67.

Kiuru, Noona, William J Burk, Brett Laursen, Katariina Salmela-Aro, and Jari-Erik Nurmi (2010) "Pressure to drink but not to smoke: Disentangling selection and socialization in adolescent peer networks and peer groups," *Journal of Adolescence*, Vol. 33, pp. 801–812.

Kossinets, Gueorgi (2006) "Effects of missing data in social networks," *Social networks*, Vol. 28, pp. 247–268.

Krauth, Brian V (2006) "Simulation-based estimation of peer effects," *Journal of Econometrics*, Vol. 133, pp. 243–271.

Lancaster, Tony (2004) *An introduction to modern Bayesian econometrics*: Blackwell Oxford.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010) "Specification and estimation of social interaction models with network structures," *Econometrics Journal*, Vol. 13, pp. 145–176.

Lewis-Faupel, Sean (2016) "The Formation and Human Capital Consequences of High School Social Networks."

Liang, Faming (2010) "A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants," *Journal of Statistical Computation and Simulation*, Vol. 80, pp. 1007–1022.

Liang, Faming, Ick Hoon Jin, Qifan Song, and Jun S Liu (2016) "An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants," *Journal of the American Statistical Association*, Vol. 111, pp. 377–393.

Lin, Xu (2010) "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables," *Journal of Labor Economics*, Vol. 28, pp. 825–860.

Liu, Xiaodong (2013) "Estimation of a local-aggregate network model with sampled networks," *Economics Letters*, Vol. 118, pp. 243–246.

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou (2014) "Endogenous peer effects: local aggregate or local average?" *Journal of Economic Behavior & Organization*, Vol. 103, pp. 39–59.

Lomi, Alessandro, Tom AB Snijders, Christian EG Steglich, and Vanina Jasmine Torló (2011) "Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance," *Social Science Research*, Vol. 40, pp. 1506–1520.

Lundborg, Petter (2006) "Having the wrong friends? Peer effects in adolescent substance use," *Journal of Health Economics*, Vol. 25, pp. 214–233.

Manski, Charles F (1993) "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, Vol. 60, pp. 531–542.

McCormick, Tyler H and Tian Zheng (2015) "Latent surface models for networks using Aggregated Relational Data," *Journal of the American Statistical Association*, Vol. 110, pp. 1684–1695.

Mele, Angelo (2017a) "Segregation in social networks: a structural approach," *Available at SSRN 3058985*.

——— (2017b) "A structural model of Dense Network Formation," *Econometrica*, Vol. 85, pp. 825–850.

——— (2017c) "A structural model of homophily and clustering in social networks," *Available at SSRN: https://ssrn.com/abstract=3031489*.

Mele, Angelo and Lingjiong Zhu (2017) "Approximate variational estimation for a model of network formation," *arXiv preprint arXiv:1702.00308*.

Miyauchi, Yuhei (2016) "Structural estimation of pairwise stable networks with nonnegative externality," *Journal of Econometrics*, Vol. 195, pp. 224–235.

Mundlak, Yair (1978) "On the pooling of time series and cross section data," *Econometrica*, Vol. 46, pp. 69–85.

Murray, Iain, Zoubin Ghahramani, and David MacKay (2006) "MCMC for doubly-intractable distributions," in *Proceedings of the 22nd Annual Conference on Uncertainity in Artificial Intelligence*: UAI.

Patacchini, Eleonora and Yves Zenou (2008) "The strength of weak ties in crime," *European Economic Review*, Vol. 52, pp. 209–236.

——— (2012) "Juvenile delinquency and conformism," *Journal of Law, Economics, and Organization*, Vol. 28, pp. 1–31.

Powell, Lisa M, John A Tauras, and Hana Ross (2005) "The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior," *Journal of Health Economics*, Vol. 24, pp. 950–968.

Propp, James Gary and David Bruce Wilson (1996) "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random structures and Algorithms*, Vol. 9, pp. 223–252.

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*, Vol. 8, pp. 1–45.

Rendon, Silvio R (2013) "Fixed and random effects in classical and Bayesian regression," *Oxford Bulletin of Economics and Statistics*, Vol. 75, pp. 460–476.

Sacerdote, Bruce (2001) "Peer effects with random assignment: results for Dartmouth roommates," *Quarterly Journal of Economics*, Vol. 116, pp. 681–704.

Schaefer, David R, Steven A Haas et al. (2013) "Social networks and smoking exploring the effects of peer influence and smoker popularity through simulations," *Health Education & Behavior*, Vol. 40, pp. 24S–32S.

Sheng, Shuyang (2014) "A structural econometric analysis of network formation games," *working paper*.

Smith, Tony E and James P LeSage (2004) "A Bayesian probit model with spatial dependencies," *Advances in Econometrics*, Vol. 18, pp. 127–160.

Snijders, Tom A., Gerhard G. Van de Bunt, and Christian E. Steglich (2010) "Introduction to stochastic actor-based models for network dynamics," *Social Networks*, Vol. 32, pp. 44–60.

Snijders, Tom AB (2001) "The statistical evaluation of social network dynamics," *Sociological Methodology*, Vol. 31, pp. 361–395.

——— (2002) "Markov chain Monte Carlo estimation of exponential random graph models," *Journal of Social Structure*, Vol. 3, pp. 1–40.

Snijders, Tom AB, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock (2006) "New specifications for exponential random graph models," *Sociological Methodology*, Vol. 36, pp. 99–153.

Soetevent, Adriaan R and Peter Kooreman (2007) "A discrete-choice model with social interactions: with an application to high school teen behavior," *Journal of Applied Econometrics*, Vol. 22, pp. 599–624.

Strauss, David and Michael Ikeda (1990) "Pseudolikelihood estimation for social networks," *Journal of the American Statistical Association*, Vol. 85, pp. 204–212.

Tanner, Martin A and Wing Hung Wong (1987) "The calculation of posterior distributions by data augmentation," *Journal of the American statistical Association*, Vol. 82, pp. 528–540.

Tierney, Luke (1994) "Markov chains for exploring posterior distributions," *Annals of Statistics*, pp. 1701–1728.

Udry, J Richard (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]," *Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill*.

Wasserman, Stanley and Philippa Pattison (1996) "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*," *Psychometrika*, Vol. 61, pp. 401–425.

Weinberg, Bruce A. (2007) "Social interactions with endogenous associations," *NBER working paper*.

Zimmerman, David J (2003) "Peer effects in academic outcomes: Evidence from a natural experiment," *Review of Economics and Statistics*, Vol. 85, pp. 9–23.

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 et al., 2006; Calvó-Armengol et al., 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 $U_i(W, Y_1, ..., Y_{\bar{d}})$ 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:

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

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

For any $Y_d, \widetilde{Y}_d$, we have: $\|B_u(Y_d) - B_u(\widetilde{Y}_d)\|_\infty = |\lambda_d| \|W(Y_d - \widetilde{Y}_d)\|_\infty \leq |\lambda_d| \|W\|_\infty \|Y_d - \widetilde{Y}_d\|_\infty$. Then, $B_u(Y_d)$ is a contraction mapping whenever $|\lambda_d| < 1/\|W\|_\infty$. By 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, i.e., $y_{i,d} \geq 0$ is similar. Indeed, since $U_i(W, Y_1, ..., Y_{\bar{d}})$ is concave in $y_{i,d}$, the optimal solution $y_{i,d}^* = \arg\max_{y_{i,d} \geq 0} U_i(W, Y_1, ..., Y_{\bar{d}})$ is given by 0 or by the first-order conditions. Formally:

$$y_{i,d}^* = \max\{0, \mu_{i,d} + \lambda_d \sum_{j=1}^{m} w_{ij} y_{j,d}\} \tag{A.2}$$

Then, similar to the case where $y_{i,d}$ is uncensored, we can write the vector-valued best response function $B_c(Y_d) = [y_{1,d}^*, ..., y_{m,d}^*]'$. Now, for a fixed value of Y_d, note that we necessarily have: $\|B_c(Y_d) - B_c(\widetilde{Y}_d)\|_\infty \leq \|B_u(Y_d) - B_u(\widetilde{Y}_d)\|_\infty$. 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 $(Y_1^*(W), ..., Y_{\bar{d}}^*(W))$, the value of the network $T(W)$ is uniquely defined. Also, since τ_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:

$$P(W) = \frac{\exp\{V(W)\}}{\sum_{\widetilde{W}} \exp\{V(\widetilde{W})\}}.$$

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.

B Likelihood function for the full model with a censored activity intensity

For the censored activity intensity $y_{ig} \in \mathbb{R}_+$, the equilibrium outcome vector based on Eq. (5) can be expressed as:

$$\begin{aligned} Y_g &= \max\left(0, \ddot{Y}_g\right), \\ \ddot{Y}_g &= \lambda W_g Y_g + X_g\beta_1 + W_g X_g \beta_2 + Z_g\rho_1 + W_g Z_g \rho_2 + l_g\alpha_g + \xi_g. \end{aligned} \tag{B.1}$$

We can divide the m_g individuals in group g into two blocks, such that the first m_{g1} individuals have activity variables equal to zero, and the remaining individuals who are arranged from $m_{g1}+1$ to m_g have positive values. $\ddot{Y}_g$ of Eq. (B.1) and the network W_g can be conformably decomposed into:

$$\begin{aligned} \begin{pmatrix} \ddot{Y}_{g1} \\ Y_{g2} \end{pmatrix} =& \lambda \begin{pmatrix} W_{11,g} & W_{12,g} \\ W_{21,g} & W_{22,g} \end{pmatrix} \begin{pmatrix} Y_{g1} \\ Y_{g2} \end{pmatrix} + \begin{pmatrix} X_{g1} \\ X_{g2} \end{pmatrix} \beta_1 + \begin{pmatrix} W_{11,g} & W_{12,g} \\ W_{21,g} & W_{22,g} \end{pmatrix} \begin{pmatrix} X_{g1} \\ X_{g2} \end{pmatrix} \beta_2 \\ &+ \begin{pmatrix} Z_{g1} \\ Z_{g2} \end{pmatrix} \rho_1 + \begin{pmatrix} W_{11,g} & W_{12,g} \\ W_{21,g} & W_{22,g} \end{pmatrix} \begin{pmatrix} Z_{g1} \\ Z_{g2} \end{pmatrix} \rho_2 + \begin{pmatrix} l_{g1} \\ l_{g2} \end{pmatrix} \alpha_g + \begin{pmatrix} \xi_{g1} \\ \xi_{g2} \end{pmatrix}, \end{aligned} \tag{B.2}$$

where $Y_{g2} > 0$ and $Y_{g1} = 0$, with the corresponding latent $\ddot{Y}_{g1} \leq 0$. Based on Eq. (B.2), the joint probability function of Y_g and W_g can be written as:

$$\begin{aligned}
&P(Y_g, W_g|\theta_g, \alpha_g, Z_g) \\
&= P(Y_{g1} = 0, Y_{g2}, W_g|\theta_g, \alpha_g, Z_g) \\
&= \int I(Y_{g1} = 0, \ddot{Y}_{g1}) \cdot P(\ddot{Y}_{g1}, Y_{g2}, W_g|\theta_g, \alpha_g, Z_g) \cdot d\ddot{Y}_{tg1} \\
&= \int_{-\infty}^{-(\lambda W_{12,g} Y_{g2} + X_{g1}\beta_1 + (W_{11,g}X_{g1} + W_{12,g}X_{g2})\beta_2 + Z_{g1}\rho_1 + (W_{11,g}Z_{g1} + W_{12,g}Z_{g2})\rho_2 + l_{g1}\alpha_g)} \\
&\quad \left|I_{m_g - m_{g1}} - \lambda W_{22,g}\right| \cdot f(\xi_{g1}, \xi_{g2}|\theta_g, \alpha_g, Z_g) \cdot P(W_g|\xi_{g1}, \xi_{g2}, \theta_g, \alpha_g, Z_g) \cdot d\xi_{g1},
\end{aligned} \tag{B.3}$$

where $I(Y_{g1} = 0, \ddot{Y}_{g1})$ is a dichotomous indicator that is equal to 1 when $\ddot{Y}_{g1}$ is negative, and equal to 0, otherwise; $\xi_{g2} = \left(I_{m_g - m_{g1}} - \lambda W_{22,g}\right) Y_{g2} - X_{g2}\beta_1 - (W_{21,g}X_{g1} + W_{22,g}X_{g2})\beta_2 - Z_{g2}\rho_1 - (W_{21,g}Z_{g1} + W_{22,g}Z_{g2})\rho_2 - l_{g2}\alpha_g$ and $\theta_g = (\gamma', \eta', \delta, \lambda, \beta', \rho', \sigma^2_{\xi_g})$.

Table 1: Results of Monte Carlo experiments for the uncensored activity variable.

		Full		No latent		No global		Latent only		Activity alone	
Parameter	True	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
λ	0.0500	0.0500	0.0073	0.0679	0.0091	0.0167	0.0052	0.0595	0.0091	0.0846	0.0073
β_1	0.3000	0.2817	0.0280	0.2986	0.0283	0.2992	0.0299	0.3038	0.0271	0.2965	0.0281
β_2	0.1000	0.0927	0.0099	0.1036	0.0196	0.1084	0.0116	0.1317	0.0185	0.1243	0.0214
ρ_1	0.3000	0.2838	0.0605	-	-	0.2359	0.0689	0.2713	0.0923	-	-
ρ_2	0.1000	0.0946	0.0248	-	-	0.0696	0.0316	0.0864	0.0685	-	-
γ_0	-0.5000	-0.5262	0.0655	-1.9885	0.1387	-1.4266	0.0617	-1.0418	0.0362	-	-
γ_3	0.3000	0.3106	0.0505	0.2733	0.0538	0.3519	0.0723	0.3620	0.0611	-	-
γ_4	-1.0000	-1.1367	0.0715	-	-	-1.1137	0.0784	-1.0580	0.0514	-	-
η_1	0.3000	0.2959	0.0405	0.4401	0.0499	-	-	-	-	-	-
η_2	0.2000	0.2172	0.0475	0.3101	0.0602	-	-	-	-	-	-
η_3	-0.1000	-0.1003	0.0084	-0.1098	0.0100	-	-	-	-	-	-
η_4	0.0400	0.0418	0.0036	0.0325	0.0042	-	-	-	-	-	-
η_5	0.3000	0.3023	0.0241	0.4015	0.0285	-	-	-	-	-	-
η_6	-0.2000	-0.1976	0.0216	-0.1777	0.0273	-	-	-	-	-	-
δ	0.3000	0.3000	0.0259	0.2094	0.0569	1.0049	0.1001	-	-	-	-
σ_ξ^2	0.5000	0.5381	0.1759	0.6852	0.1843	0.6647	0.2403	0.5259	0.1764	0.6740	0.1784

Note: This Monte Carlo study consists of 100 repetitions. The values reported in this table are the mean and the standard deviation of parameter estimates across repetitions. In each repetition, we estimate each of the corresponding models with 50,000 MCMC draws. We drop the first 10,000 draws due to burn-in and calculate the (posterior) mean of the remaining draws as parameter estimates.

Table 2: Results of Monte Carlo experiments for the censored activity variable

		Full		No latent		No global		Latent only		Activity alone	
Parameter	True	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
λ	0.0500	0.0492	0.0131	0.0890	0.0157	0.0242	0.0121	0.0568	0.0131	0.1187	0.0147
β_1	0.3000	0.2695	0.0169	0.2798	0.0130	0.2726	0.0153	0.2779	0.0122	0.2761	0.0123
β_2	0.1000	0.0820	0.0091	0.0880	0.0060	0.0685	0.0085	0.1094	0.0058	0.0935	0.0065
ρ_1	0.3000	0.2464	0.0320	-	-	0.2567	0.0615	0.2496	0.0478	-	-
ρ_2	0.1000	0.0972	0.0228	-	-	0.1145	0.0249	0.1073	0.0401	-	-
γ_0	-1.7000	-1.6505	0.1155	-2.7295	0.1327	-1.9584	0.0895	-1.7348	0.0502	-	-
γ_3	0.3000	0.3022	0.0426	0.2673	0.0474	0.3211	0.0556	0.3265	0.0530	-	-
γ_4	-1.0000	-1.0715	0.0848	-	-	-1.0638	0.1144	-1.1122	0.0692	-	-
η_1	0.3000	0.2873	0.0392	0.4504	0.0520	-	-	-	-	-	-
η_2	0.2000	0.2085	0.0487	0.2107	0.0684	-	-	-	-	-	-
η_3	-0.1000	-0.0998	0.0099	-0.0975	0.0129	-	-	-	-	-	-
η_4	0.0400	0.0391	0.0041	0.0352	0.0044	-	-	-	-	-	-
η_5	0.3000	0.2802	0.0325	0.4073	0.0370	-	-	-	-	-	-
η_6	-0.2000	-0.2062	0.0252	-0.1609	0.0312	-	-	-	-	-	-
δ	0.3000	0.3166	0.0117	0.1885	0.0521	0.4211	0.1242	-	-	-	-
σ_ξ^2	0.5000	0.5163	0.1406	0.6132	0.1477	0.5348	0.1611	0.4739	0.1289	0.6013	0.1425

Note: This Monte Carlo study consists of 100 repetitions. The values reported in this table are the mean and the standard deviation of parameter estimates across repetitions. In each repetition, we estimate each of the corresponding models with 50,000 MCMC draws. We drop the first 10,000 draws due to burn-in and calculate the (posterior) mean of the remaining draws as parameter estimates.

Table 3: 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
Female	0	1	0.543	0.498
White	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
HS	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
Staying home	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		1,036		
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.

Table 4: Estimation results based on both GPA and smoking.

	Full		No latent		No global		Latent only		Activity alone	
Local & global & incentive effects										
Constant (γ_0)	-4.8531***		-5.7071***		-3.9329***		1.1101***		-	
	(0.0366)		(0.1194)		(0.1134)		(0.0958)		-	
Experience in school (sender) (γ_1)	-0.0405***		-0.0159		0.0892***		0.0528***		-	
	(0.0114)		(0.0212)		(0.0211)		(0.0148)		-	
Experience in school (receiver) (γ_2)	0.0374**		0.0515***		0.1586***		0.1301***		-	
	(0.0143)		(0.0198)		(0.0247)		(0.0147)		-	
Same age (γ_{31})	0.3675***		0.5340***		1.0300***		0.4450***		-	
	(0.0431)		(0.0496)		(0.0773)		(0.0586)		-	
Same sex (γ_{32})	0.3471***		0.4889***		0.3454***		0.5477***		-	
	(0.0441)		(0.0608)		(0.0686)		(0.0517)		-	
Same race (γ_{33})	0.3116***		0.2152***		0.3716***		0.3961***		-	
	(0.0430)		(0.0789)		(0.0839)		(0.0595)		-	
Latent distance (γ_{41})	-0.1469***		-		-0.3945***		-3.9387***		-	
	(0.0173)		-		(0.0671)		(0.1214)		-	
Latent distance (γ_{42})	-0.0966**		-		-0.2231***		-2.5592***		-	
	(0.0414)		-		(0.0662)		(0.1000)		-	
Latent distance (γ_{43})	-0.0125		-		-0.0868		-2.3793***		-	
	(0.0528)		-		(0.0594)		(0.0887)		-	
Reciprocity (η_1)	1.4309***		1.4080***		-		-		-	
	(0.0476)		(0.0552)		-		-		-	
Congestion (η_2)	0.2699***		0.3521***		-		-		-	
	(0.0151)		(0.0281)		-		-		-	
Congestion (η_3)	-0.0247***		-0.0304***		-		-		-	
	(0.0017)		(0.0022)		-		-		-	
Popularity (η_4)	0.0049		0.0015		-		-		-	
	(0.0068)		(0.0061)		-		-		-	
Trans. triads (η_5)	0.4715***		0.4767***		-		-		-	
	(0.0220)		(0.0189)		-		-		-	
Three cycles (η_6)	-0.2071***		-0.2083***		-		-		-	
	(0.0171)		(0.0166)		-		-		-	
Incentive from GPA (δ_1)	0.2145**		0.1825***		0.2921***		-		-	
	(0.0956)		(0.0344)		(0.1122)		-		-	
Incentive from smoking (δ_2)	0.0197		0.0083		0.0214		-		-	
	(0.0134)		(0.0053)		(0.0131)		-		-	
Activity intensity – GPA										
Endogenous (λ)	0.0177***		0.0189***		0.0162**		0.0239**		0.0330***	
	(0.0063)		(0.0070)		(0.0071)		(0.0091)		(0.0105)	
	Own	Contex.	Own	Contex.	Own	Contex.	Own	Contex.	Own	Contex.
Age	−0.0327**	0.0018	−0.0368***	0.0013	−0.0382**	0.0009	−0.0388*	0.0009	−0.0366**	−0.0013
	(0.0164)	(0.0024)	(0.0136)	(0.0024)	(0.0148)	(0.0025)	(0.0202)	(0.0025)	(0.0167)	(0.0026)
Male	−0.1984***	−0.0152	−0.1926***	−0.0166	−0.1854***	−0.0143	−0.1863***	−0.0166	−0.1906***	−0.0161
	(0.0307)	(0.0215)	(0.0381)	(0.0239)	(0.0335)	(0.0212)	(0.0381)	(0.0232)	(0.0359)	(0.0244)
Black	−0.0014	−0.0636***	−0.0564	−0.0564**	−0.0265	−0.0608***	−0.0109	−0.0489**	−0.0607	−0.0577**

Table – Continued

	(0.0473)	(0.0247)	(0.0613)	(0.0234)	(0.0433)	(0.0226)	(0.0716)	(0.0219)	(0.0626)	(0.0212)
Asian	0.0028	−0.1957***	−0.0791	−0.2240***	−0.0574	−0.1647***	−0.0384	−0.1824***	−0.0309	−0.2251***
	(0.0723)	(0.0566)	(0.1004)	(0.0659)	(0.0501)	(0.0393)	(0.0902)	(0.0567)	(0.0826)	(0.0643)
Hispanic	−0.2332***	−0.0362	−0.2241***	−0.0381	−0.2406***	−0.0644**	−0.2433***	−0.0502	−0.1990***	−0.0278
	(0.0360)	(0.0485)	(0.0776)	(0.0576)	(0.0442)	(0.0337)	(0.0737)	(0.0500)	(0.0680)	(0.0526)
Other race	−0.0689	−0.0967**	−0.1168*	−0.0873*	−0.0671	−0.1137***	−0.0555	−0.0838*	−0.0914*	−0.1020**
	(0.0418)	(0.0449)	(0.0673)	(0.0483)	(0.0524)	(0.0385)	(0.0575)	(0.0473)	(0.0556)	(0.0491)
Both parents	0.1149***	−0.0701***	0.1140**	−0.0598**	0.1147***	−0.0525**	0.1145***	−0.0602**	0.1126***	−0.0637**
	(0.0368)	(0.0213)	(0.0510)	(0.0271)	(0.0382)	(0.0268)	(0.0427)	(0.0258)	(0.0449)	(0.0293)
Less HS	−0.3424***	−0.0889**	−0.3466***	−0.0895**	−0.3074***	−0.1027***	−0.2972***	−0.0924**	−0.3070***	−0.0839**
	(0.0586)	(0.0392)	(0.0702)	(0.0450)	(0.0507)	(0.0333)	(0.0623)	(0.0435)	(0.0608)	(0.0378)
More HS	0.1593***	0.0485**	0.1561***	0.0419*	0.1686***	0.0397	0.1536***	0.0261	0.1553***	0.0272
	(0.0334)	(0.0211)	(0.0397)	(0.0236)	(0.0324)	(0.0223)	(0.0435)	(0.0248)	(0.0367)	(0.0256)
Edu missing	−0.1739***	0.0634**	−0.1751***	0.0457	−0.1696***	0.0355	−0.1624**	0.0472	−0.1727***	0.0323
	(0.0532)	(0.0334)	(0.0588)	(0.0391)	(0.0362)	(0.0341)	(0.0634)	(0.0366)	(0.0533)	(0.0342)
Welfare	−0.1079**	−0.2848***	−0.1593*	−0.2434**	−0.0547	−0.2932***	−0.1001	−0.2892**	−0.1973**	−0.1893**
	(0.0567)	(0.0723)	(0.0950)	(0.0980)	(0.0634)	(0.0550)	(0.1589)	(0.1173)	(0.1000)	(0.0716)
Job missing	−0.1335***	−0.0446	−0.1261*	−0.0359	−0.1808***	−0.0371	−0.1625***	−0.0522	−0.1500***	−0.0253
	(0.0369)	(0.0442)	(0.0662)	(0.0476)	(0.0551)	(0.0387)	(0.0503)	(0.0381)	(0.0525)	(0.0464)
Professional	−0.1104***	0.0010	−0.1158**	0.0075	−0.1162***	0.0119	−0.1347***	0.0003	−0.1097**	0.0076
	(0.0309)	(0.0267)	(0.0459)	(0.0278)	(0.0376)	(0.0209)	(0.0377)	(0.0264)	(0.0387)	(0.0246)
Other jobs	−0.0361	0.0268	−0.0482	0.0283	−0.0610**	0.0310**	−0.0679*	0.0214	−0.0354	0.0289
	(0.0360)	(0.0229)	(0.0412)	(0.0222)	(0.0312)	(0.0167)	(0.0377)	(0.0227)	(0.0392)	(0.0214)
Num. of other students at home	0.0257	0.0123	0.0195	0.0131	0.0222	0.0154	0.0216	0.0096	0.0184	0.0068
	(0.0230)	(0.0104)	(0.0257)	(0.0133)	(0.0212)	(0.0119)	(0.0232)	(0.0140)	(0.0232)	(0.0131)
Latent (ρ_{11})	0.0211	−0.0198	-	-	0.0067	0.0014	0.0272	−0.0170	-	-
	(0.0548)	(0.0283)	-	-	(0.0275)	(0.0230)	(0.0369)	(0.0112)	-	-
Latent (ρ_{12})	0.0905*	0.0070	-	-	0.0260	−0.0020	0.0721*	−0.0235	-	-
	(0.0537)	(0.0254)	-	-	(0.0610)	(0.0242)	(0.0472)	(0.0197)	-	-
Latent (ρ_{13})	0.0542	0.0101	-	-	0.0545	0.0105	0.0724*	−0.0045	-	-
	(0.0481)	(0.0277)	-	-	(0.0454)	(0.0232)	(0.0466)	(0.0188)	-	-
Activity intensity − smoking										
Endogenous (λ)	0.1052***		0.1103***		0.1068***		0.1056***		0.1125***	
	(0.0196)		(0.0225)		(0.0193)		(0.0197)		(0.0191)	
	Own	Contex.	Own	Contex.	Own	Contex.	Own	Contex.	Own	Contex.
Age	0.1462***	−0.0047	0.1613***	−0.0054	0.1837***	−0.0052	0.2051***	−0.0120**	0.1655***	−0.0042
	(0.0201)	(0.0036)	(0.0248)	(0.0045)	(0.0386)	(0.0040)	(0.0455)	(0.0053)	(0.0212)	(0.0040)
Male	0.0126	−0.0185	−0.0598	0.0390	0.0307	−0.0077	−0.0494	0.0414	−0.0240	0.0160
	(0.0650)	(0.0305)	(0.0691)	(0.0424)	(0.0594)	(0.0427)	(0.0915)	(0.0436)	(0.0624)	(0.0395)
Black	−0.1952***	0.0828***	−0.3633***	0.0726**	−0.2332**	0.0808**	−0.4153**	0.0810**	−0.0383	0.0939***
	(0.0749)	(0.0194)	(0.1395)	(0.0357)	(0.1003)	(0.0375)	(0.2075)	(0.0377)	(0.0954)	(0.0311)
Asian	−1.1052***	0.3028***	−0.5535***	0.1893***	−1.0746***	0.2965***	−1.2797***	0.3093***	−0.5277**	0.2472
	(0.0615)	(0.0444)	(0.1917)	(0.0672)	(0.0857)	(0.0639)	(0.3075)	(0.1120)	(0.2115)	(0.1586)
Hispanic	0.1317**	0.0734	−0.0626	0.0204	0.1128	0.0625	−0.0591	0.1509*	0.2026**	0.0232
	(0.0613)	(0.0472)	(0.1029)	(0.0839)	(0.0862)	(0.0735)	(0.2321)	(0.0849)	(0.0990)	(0.0734)
Other race	0.0610	0.4019***	0.3284***	0.2831**	0.0755	0.3913***	0.1891	0.5043***	0.2874**	0.3868***

Table – Continued

	(0.0704)	(0.0578)	(0.0894)	(0.0896)	(0.0659)	(0.0535)	(0.1459)	(0.0714)	(0.1449)	(0.0814)
Both parents	−0.0808	−0.0140	0.0250	−0.0431	−0.0101	−0.0564	−0.0588	−0.0351	−0.0061	−0.0437
	(0.0492)	(0.0311)	(0.0658)	(0.0434)	(0.0450)	(0.0415)	(0.0878)	(0.0579)	(0.0672)	(0.0382)
Less HS	0.3986***	0.0485	0.1828**	0.0409	0.2762***	0.0382	0.2837***	0.0762	0.1371	0.0426
	(0.0484)	(0.0634)	(0.0919)	(0.0607)	(0.0533)	(0.0478)	(0.0972)	(0.0667)	(0.1032)	(0.0676)
More HS	−0.1975***	−0.0357	−0.2182***	−0.0329	−0.0548	−0.0381	−0.1502*	0.0131	−0.1895***	−0.0103
	(0.0515)	(0.0369)	(0.0661)	(0.0479)	(0.0443)	(0.0383)	(0.0906)	(0.0472)	(0.0679)	(0.0512)
Edu missing	−0.0965	−0.0551	−0.1907**	0.0275	−0.1202*	−0.0620	−0.0775	−0.0273	−0.1388	−0.0216
	(0.0636)	(0.0351)	(0.0750)	(0.0611)	(0.0744)	(0.0433)	(0.0874)	(0.0806)	(0.0877)	(0.0619)
Welfare	1.5846***	−0.1062**	1.2705***	−0.1601	1.5957***	−0.0860	1.6348***	0.0728	1.7299***	−0.5848
	(0.0764)	(0.0550)	(0.1107)	(0.2201)	(0.0559)	(0.0653)	(0.2791)	(0.4136)	(0.2190)	(0.4791)
Job missing	0.4931***	0.0860	0.3091**	0.1156*	0.5466***	0.0798	0.5669***	0.1639*	0.2347**	0.0045
	(0.0612)	(0.0542)	(0.1322)	(0.0683)	(0.0496)	(0.0658)	(0.1226)	(0.0907)	(0.1107)	(0.0696)
Professional	0.3740***	−0.0013	0.3326***	0.0014	0.3079***	0.0492	0.4512***	0.0344	0.2685***	−0.0150
	(0.0464)	(0.0258)	(0.0791)	(0.0500)	(0.0852)	(0.0381)	(0.1044)	(0.0559)	(0.0949)	(0.0436)
Other jobs	0.0037	0.0395	0.1530***	0.0315	0.1064*	0.0460	0.1472*	0.0682	−0.0551	0.0047
	(0.0545)	(0.0319)	(0.0552)	(0.0350)	(0.0615)	(0.0371)	(0.0882)	(0.0450)	(0.0758)	(0.0311)
Num. of other students at home	−0.1551***	−0.0105	−0.0745*	−0.0028	−0.1057***	0.0073	−0.0636	0.0173	−0.0541	0.0009
	(0.0478)	(0.0234)	(0.0401)	(0.0246)	(0.0342)	(0.0195)	(0.0472)	(0.0291)	(0.0378)	(0.0234)
Latent (ρ_{21})	−0.0921***	−0.0180	-	-	0.0052	0.0163	−0.0119	−0.0027	-	-
	(0.0288)	(0.0258)	-	-	(0.0447)	(0.0453)	(0.0510)	(0.0201)	-	-
Latent (ρ_{22})	0.0074	−0.0057	-	-	0.0147	−0.0044	−0.0332	0.0168	-	-
	(0.0471)	(0.0331)	-	-	(0.0712)	(0.0414)	(0.0574)	(0.0290)	-	-
Latent (ρ_{23})	0.1055**	0.0264	-	-	−0.0125	0.0250	0.0758	−0.0579**	-	-
	(0.0520)	(0.0462)	-	-	(0.0632)	(0.0365)	(0.0562)	(0.0216)	-	-
Group fixed effect	Yes		Yes		Yes		Yes		Yes	
$\sigma^{2(*)}_{\xi_{uc,g}}$ (GPA)	0.4508		0.4801		0.4618		0.4668		0.4809	
	(0.1532)		(0.1527)		(0.1501)		(0.1548)		(0.1591)	
$\sigma^{2(*)}_{\xi_{c,g}}$ (smoking)	3.6320		3.6435		3.5293		3.6085		3.5702	
	(3.2199)		(3.2489)		(3.1070)		(3.1239)		(3.1364)	
$\sigma^{(*)}_{\xi_{ucc,g}}$	-0.2522		-0.2452		-0.2448		-0.2542		-0.2474	
	(0.2708)		(0.2770)		(0.2738)		(0.2755)		(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 asterisks $^{***}(^{**},^{*})$ indicate that its 99% (95%, 90%) highest posterior density range does not cover zero. $\sigma^{2(*)}_{\xi_{uc,g}}$, $\sigma^{2(*)}_{\xi_{c,g}}$, and $\sigma^{(*)}_{\xi_{ucc,g}}$ 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., λ and δ and the convergence diagnostics of Geweke (1992), are provided in Supplementary Appendix Figure E.3.

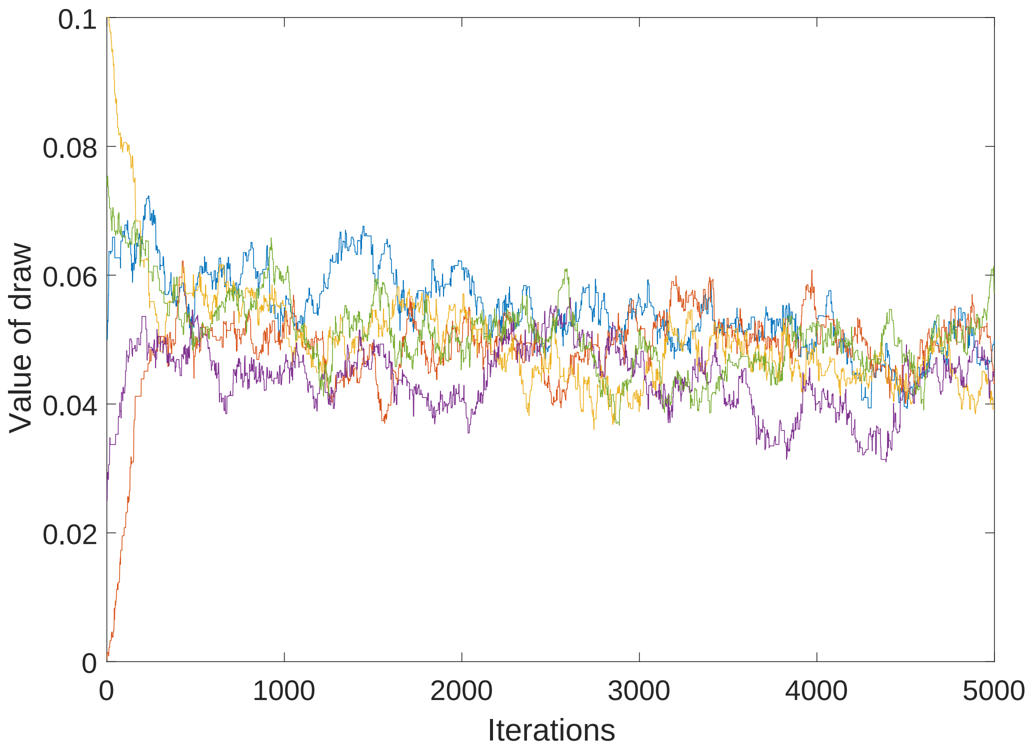


Figure 3: Trace plot of MCMC draws for the parameter λ at different initial values (for the uncensored outcome). The true value of λ in the simulation is 0.05.

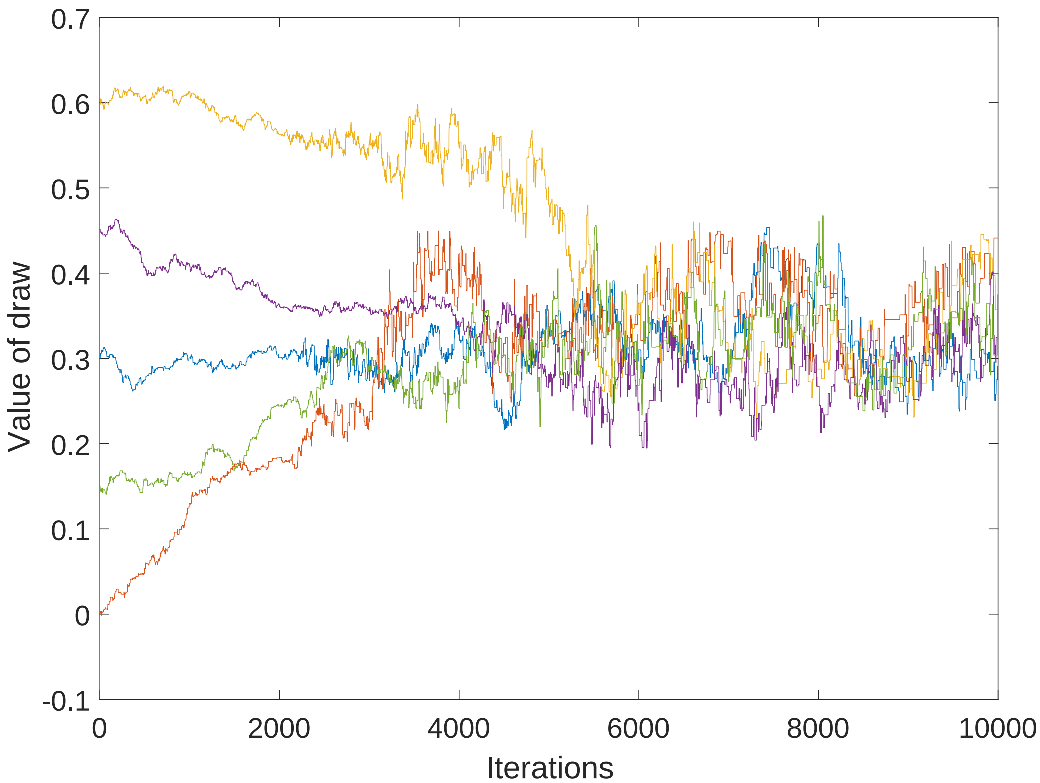


Figure 4: Trace plot of MCMC draws for the parameter δ at different initial values (for the uncensored outcome). The true value of δ in the simulation is 0.3.

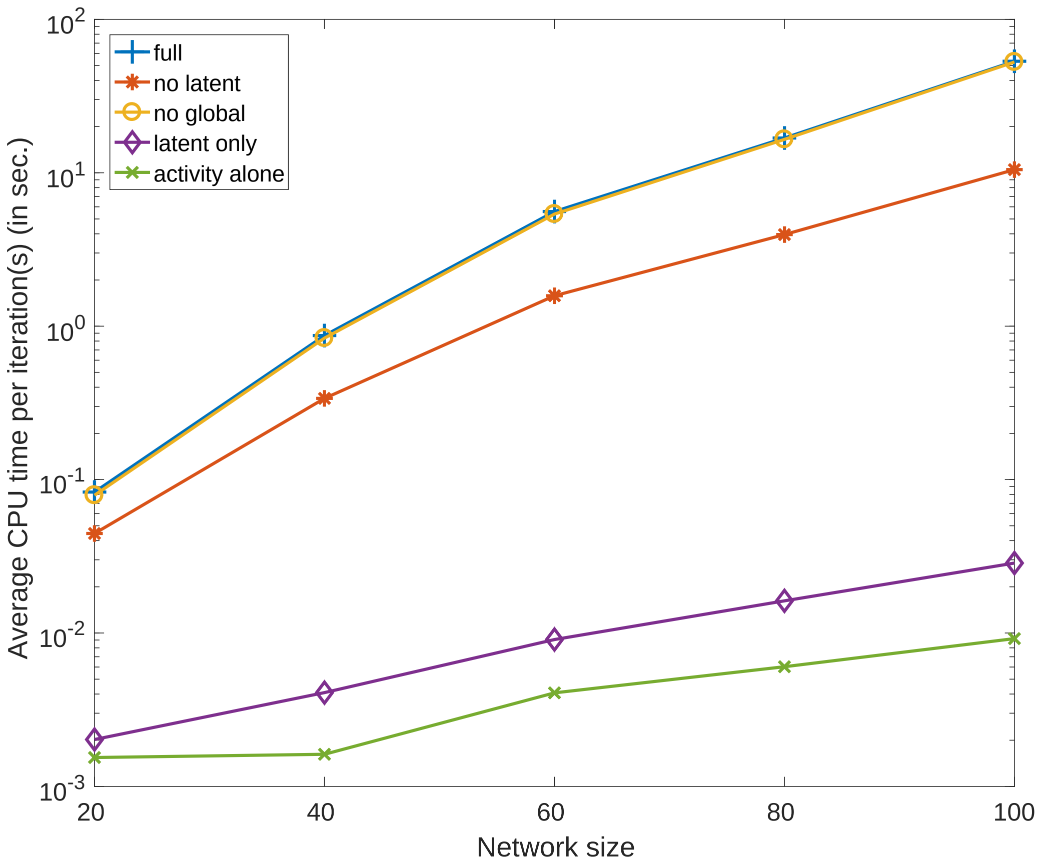


Figure 5: The average computation time (in seconds) for a single MCMC iteration. The scale of the vertical axis is logarithmic.