A Structural Model of Segregation in Social Networks*

Angelo Mele†

September 5, 2011

Abstract

In this paper, I develop and estimate a dynamic model of strategic network formation with heterogeneous agents. Individuals meet sequentially and update their links according to their preferences for socioeconomic attributes. While existing models have multiple equilibria, I prove the existence of a unique stationary equilibrium, which characterizes the likelihood of observing a specific network in the data. As a consequence, the structural parameters can be estimated using only one observation of the network at a single point in time. The estimation is challenging because the exact evaluation of the likelihood is computationally infeasible. To circumvent this problem, I propose an Bayesian Markov Chain Monte Carlo algorithm that avoids direct evaluation of the likelihood. I prove that the algorithm converges to the correct posterior distribution. This method drastically reduces the computational burden of estimating the posterior distribution and allows inference in high dimensional models.

I present an application to the study of segregation in school friendship networks, using data from Add Health containing the actual social networks of students in a representative sample of US schools. My results suggest that for white students, the value of a same-race friend decreases with the fraction of whites in the school. The opposite is true for African American students.

The model is used to study how different desegregation policies may affect the structure of the network in equilibrium. I find an inverted u-shaped relationship between the fraction of students belonging to a racial group and the expected equilibrium segregation levels. These results suggest that desegregation programs may decrease the degree of interracial interaction within schools.

JEL Codes: D85, C15, C73
Keywords: Social Networks, Bayesian Estimation, Markov Chain Monte Carlo

*I am grateful to Roger Koenker for continuous encouragement and advice, generous financial support and for the use of his computer cluster. I thank Ron Laschever for long and fruitful discussions about this research project. Dan Bernhardt and George Deltas have provided several suggestions at crucial stages of this work. I thank Alberto Bisin, Ethan Cole, Aureo de Paula, Shweta Gaonkar, Dan Karney, Darren Lubotsky, Antonio Mele, Luca Merlino, Tom Parker, Dennis O'Dea, Micah Pollak, Sergey Popov, Sudipta Sarangi, Giorgio Topa, Antonella Tutino and seminar participants at SED Meetings 2010, Add Health Users Conference 2010, Notre Dame, Johns Hopkins University - Carey Business School, UPENN, Southern Methodist University, FED Board, UQAM, SBIES 2011, PET 2011 for helpful comments and suggestions. I am especially grateful to Andriy Norets and Matt Jackson for suggestions on how to prove and bound convergence of the algorithms. Financial support from the Robert Ferber Award, the Robert Willis Harbeson Memorial Dissertation Fellowship, and the NET Institute Summer Research Grant 2010 is gratefully acknowledged. All remaining errors are mine

†*Address*: Dept. of Economics, University of Illinois at Urbana-Champaign, 419 David Kinley Hall, 1407 W. Gregory Dr., Urbana, IL 61801. *Email*: amele2@illinois.edu

1 Introduction

Social networks are important determinants of individuals' socioeconomic performance. An increasing amount of evidence shows that the number and composition of social ties affects employment prospects, school performance, risky behavior, adoption of new technologies, diffusion of information and health outcomes.[1] The structure of social ties is endogenously determined: individuals choose their peers and friends according to their socioeconomic characteristics and their relationships. As a consequence, in socially generated networks the agents tend to interact with similar individuals (homophily) and there exists segregation along socioeconomic attributes.[2] This paper develops a model and estimation methods to study the determinants of such segregated network structures.

The economics literature has relied on two complementary approaches to model the process of network formation. Strategic models interpret the network as equilibrium outcome of a strategic game, in which rational players choose their social ties by weighing the costs and benefits of each link.[3] Structural estimation of strategic models is challenging, because the process of link formation generates externalities: individuals benefit from indirect links created by other players. As a result, these models have multiple equilibria, a feature that greatly complicates the identification of structural parameters, requiring multiple network observation for estimation.[4] Furthermore, strategic models suffer a curse of dimensionality: the number of possible network structures increases exponentially with the number of players. Therefore the computation of equilibria may require an infeasible number of calculations for networks with a large number of players. On the other hand, in Random network formation models, the network is the realization of a stochastic process.[5] Each link occurs with some probability and the model characterizes the likelihood of observing a specific network configuration. While these models are able to match several important characteristics of social network data (high clustering, homophily, diameter) and their estimation is relatively standard, they lack the equilibrium microfoundation of the strategic models, limiting their use for policy analysis.

In this paper, I develop and estimate a strategic model of dynamic network formation

[1] For example, see the contributions of Topa (2001); Laschever (2009); Cooley (2010); De Giorgi et al. (2010); Nakajima (2007); Bandiera and Rasul (2006); Conley and Udry (forthcoming); Golub and Jackson (2011); Acemoglu et al. (2011).

[2] See Currarini et al. (2009, 2010), De Marti and Zenou (2009),Echenique et al. (2006).

[3] Jackson and Wolinsky (1996), Bala and Goyal (2000),Galeotti (2006),Jackson (2008).

[4] Comola (2008) and Mayer and Puller (2008) are examples of structural models where multiple equilibria may arise.

[5] Jackson (2008),Jackson and Rogers (2007), Snijders (2002), Wasserman and Pattison (1996)

with heterogeneous agents. I prove and characterize the existence of *unique stationary equilibrium*, which provides the likelihood of observing a specific network architecture in the long run. The absence of multiple equilibria is crucial in estimation: assuming that the network observed in the data is drawn from the stationary equilibrium of the model, the structural parameters can be estimated using *only one observation of the network at a single point in time*.

I estimate the posterior distribution of the structural parameters, using a Bayesian approach. The main challenge is that the model's likelihood is proportional to an intractable normalizing constant, that cannot be evaluated or approximated with precision. To overcome this problem, I propose a Markov Chain Monte Carlo algorithm that generate samples from the posterior distribution without evaluating the likelihood. I prove that the algorithm converges to the correct posterior and that the sample averages of the algorithm's output can be used to approximate posterior expectations. Using several properties of the model, I am able to decrease the computational burden even further, allowing inference in high dimensional models.

I present an application to the study of segregation in school friendship networks, using data from the *National Longitudinal Study of Adolescent Health* (Add Health).[6] This unique dataset contains detailed information on the actual friendship networks of students in a representative sample of US schools. The final sample contains 14 schools with a total of 1139 students.[7] I find that race, gender and grade are important determinants of network formation in schools. There is overwhelming evidence of homophily, i.e. students tend to interact and form social ties with similar people, other things being equal. My results suggest that for white students the value of a same-race friend decreases with the fraction of whites in the school. The opposite is true for African American students: the value of an African American friend increases with the proportion of blacks in the school. Hispanic preferences seem to mirror those of whites.

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

[7]I use only the schools from the *saturated sample*. The sampling scheme of Add Health involved in-school interviews for all the students. A subsample of 20745 students was also interviewed at home, to collect detailed individual information. The saturated sample contains schools for which both interviews were administered to *each* student enrolled. Therefore this sample does not contain any missing information about individual controls. This is not the case for most schools in Add Health.

The model can be employed in policy analysis, because it allows the researcher to simulate counterfactual policy experiments.[8] This model provides useful guidance to policymakers who care about promoting policies that affect the structure of the network. For example, I consider a busing program that affects two schools from the sample, one with 98% whites and the other with 96% blacks. I simulate alternative swaps of students across schools and then measure the average segregation in the new stationary equilibrium of the model. I find that there is an *inverted U-shape* relationship between the fraction of students belonging to a racial group and the expected levels of equilibrium segregation.[9] For example, a reduction in the white student share from 90% to 80% implies an average increase of expected segregation by .20, as measured by the Freeman (1972) segregation index.[10]

The paper provides both theoretical and empirical contributions to the literature on network formation. First, the model incorporates ingredients from both strategic and random network formation literature (Jackson, 2008). The link formation is sequential: each period only one agent is active and he updates only one link. At the beginning of the period, a random agent (John) is drawn from the population and he meets another agent (Liz) according to a random matching technology. At this point he can choose to update his social tie to Liz. The implicit assumption is that meetings occur often and the agents have frequent opportunities to revise their strategies.

The agents' preferences allow for a rich structure of indirect payoffs from link formation. Individuals care about the socioeconomic composition of their friends, friends of friends and feedback from those friends payoffs. Concretely, John's utility from linking to Liz depends on her socioeconomic attributes; additionally, he values the socioeconomic composition of her friends and how befriending her could affect his popularity among the other players. Finally, a link provides additional utility when it is reciprocated. When updating the link, John receives a random shock to his preferences, which is unobserved by the econometrician. This shock captures unobservable factors that could affect link formation: for example, John may be in a bad mood when he meets Liz, and this affects his linking strategy. The link is formed when the social relationship provides positive utility; otherwise the agent does not form (or severs, if already in place) the friendship. To preserve tractability, I assume that individuals do not take into account how their current linking strategy affects the shape

[8] Alternatively, it could be used as a guide for designing randomized experiments that modify students assignments.

[9] Currarini et al. (2010) use a different model and find the same relationship.

[10] The index measures the difference between the expected and actual number of links among individuals of different groups. An index of 0 means that the actual network closely resembles one in which links are formed at random. Higher values indicate more segregation. The maximum of 1 corresponds to a network in which there are no inter-group links.

of future networks: they follow a stochastic best-response dynamics *à la* Blume (1993).[11] This assumption reduces the computational complexity and makes analysis of the network dynamics feasible.[12]

The model has two desirable features. First, there are two sources of heterogeneity. Each individual is endowed with a set of *exogenous* attributes. Furthermore, the dynamics of network formation generates *endogenous* heterogeneity: each individual has a different set of friends and different compositions of friends' attributes. In equilibrium, two agents with the same exogenous attributes may exhibit very different linking strategies, due to their different endogenous positions in the network and the socioeconomic composition of their friends. Most models of strategic network formation incorporate the first level of heterogeneity but are unable to generate different equilibrium behavior, because the agents in these models only care about their direct links.[13]

Second, the network formation game can be characterized as a *potential game*.[14] All the players' incentives in any state of the network are completely summarized by an aggregate function, *the potential*, mapping networks and socioeconomic characteristics into potential levels. When an agent updates a link, the change in his utility is equal to the change in the aggregate potential. This simple characterization is key to making analysis of a network with many agents feasible because the potential summarizes the incentives of all players with a single number: there is no need to keep track of the choices and utility levels of all n players. The existence of a potential allows me to characterize the stationary equilibrium in closed form. Assuming that preference shocks follow an extreme value distribution (i.i.d. over time and across agents), and that any pair of agents can meet with positive probability, I prove that the unique stationary equilibrium characterizes the probability of observing a specific network structure as an exponential function of the potential. This result provides the likelihood function underlying the estimation.

The second main contribution is the development of a flexible algorithm that reduces

[11] It is possible to relax the assumption of myopic agents, but the computational burden becomes much more challenging. The simple characterization of equilibrium behavior, long run dynamics and the estimation strategy depend on the best-response dynamics and may not extend to networks with forward-looking agents.

[12] Alternatively, it is possible to interpret this model as an equilibrium selection device, that selects one of the possible networks as the result of an evolutionary game.

[13] An exception is the model of De Marti and Zenou (2009), where the cost of linking an individual also depends on the composition of friends of friends. While the structure of the preference is similar to mine, they present a static model and the link formation requires mutual agreement of the players. The consequence is that their model has multiple equilibria.

[14] See Monderer and Shapley (1996) for a description of games with a potential. Gilles and Sarangi (2004) investigates a model of network formation with a potential function. Their model only considers the utility from direct links, while mine includes indirect links, mutual links and popularity.

the computational burden for estimation.[15] The likelihood of this model is proportional to a normalizing constant that cannot be evaluated or approximated with precision. A state-of-the-art supercomputer would take several years to evaluate the likelihood at a single parameter value. This feature prevents the use of traditional MCMC schemes, e.g. Metropolis-Hastings, that evaluate the likelihood at each iteration. To circumvent this problem, I propose a Markov Chain Monte Carlo algorithm that samples from the posterior distribution of the parameters without evaluating the likelihood. The algorithm belongs to the class of *exchange algorithms*, first proposed by Murray et al. (2006) to generate samples from posterior distributions with similar intractable likelihoods. I develop an approximate version of the algorithm that reduces the computational burden even further, allowing inference in high dimensional models.[16]

The idea behind the sampler is as follows. At each iteration we perform a double Metropolis-Hastings step. First, we propose a new parameter vector as in the standard Metropolis-Hastings scheme; second, we draw a new network from the stationary equilibrium of the model at the proposed parameter, using a second Metropolis-Hastings algorithm. If the network generated by the latter simulation is similar to the network observed in the data, then the parameter is accepted with high probability; viceversa, if the simulated network is very different from the data, the probability of acceptance is lower. The similarity between networks is measured using likelihood ratios at the current and proposed parameters. The intuition is that if the observed and simulated networks are similar, then the proposed parameter is also likely to generate the data as a draw from the stationary equilibrium.

I prove that the samples generated by my approximate algorithm converge to the correct posterior distribution. Therefore, sample averages computed using the algorithm's output can be used to approximate posterior expectations and to make inference. In the empirical application I use an extension of the algorithm which makes use of parallel computing techniques and allows me to estimate the model using multiple network data. This is extremely important if the model is used to perform counterfactual policy experiments: using multiple networks allows to control for network fixed effects and additional network level characteristics that could affect the outcome of the implemented policy.

The paper proceeds as follows. In section 2, I present the model's details and show that

[15] Christakis et al. (2010) propose an alternative model and estimation procedure. In their model myopic agents meet over time and must mutually agree to form a link. The meeting is treated as a latent variable and integrated out in estimation, using a MCMC algorithm that samples parameters from the posterior while simulating meeting sequences.

[16] I use arguments similar to Liang (2010)

it belongs to the class of potential games. Then, I provide a full characterization of the unique stationary equilibrium. Section 3 identifies the computational problem in estimation and proposes an approximate exchange algorithm for the estimation of the posterior distribution. I prove that the proposed algorithm converges to the correct posterior distribution. The empirical application is presented in Section 4, where I estimate preferences for interracial interaction in a sample of US high schools. I show that the model can be used for policy analysis: I study the impact of school desegregation programs on the structure of the school friendship network. Section 5 concludes. The proofs for the theoretical results are in Appendix A, while the details of the estimation algorithm are contained in Appendix B.

2 A Model of Network Formation

2.1 Setup

Let $\mathcal{I} = \{1, 2, ..., n\}$ be the set of agents, each identified by a vector of A (exogenous) attributes $X_i = \{X_{i1}, ..., X_{iA}\}$, e.g. gender, wealth, age, location, etc. The attributes of the population are contained in the matrix $X = \{X_1, X_2, ..., X_n\}$ and $\mathcal{X}$ denotes the set of all possible matrices X. Time is discrete.

The social network is represented as a $n \times n$ binary matrix $G \in \mathcal{G}$, where $\mathcal{G}$ is the set of all $n \times n$ binary matrices. The generic element of the matrix G is

$$G_{ij} = \begin{cases} 1 & \text{if individual } i \text{ nominates individual } j \text{ as a friend} \\ 0 & \text{otherwise} \end{cases}$$

and I follow the convention in the literature, assuming $G_{ii} = 0$, for any i.

The network represented by G is *directed*: the existence of a link from i to j does not imply the existence of the link from j to i, i.e. $g_{ij} \neq g_{ji}$. This modeling choice reflects the structure of the Add Health data, where friendship nominations are not necessarily mutual. Some authors refer to this data as *perceived* networks.[17]

Let the *realization* of the network at time t be denoted as g^t and the *realization* of the link between i and j at time t be g^t_{ij}. The network including all the current links but g^t_{ij}, i.e. $g^t \backslash g^t_{ij}$, is denoted as g^t_{-ij}; while g^t_{-i} denotes the network matrix excluding the i-th row (i.e. all the links of player i).

Preferences are defined over network realizations and population characteristics. I assume there is an utility function $U_i : \mathcal{G} \times \mathcal{X} \to \mathbb{R}$ for each i, mapping networks and individual characteristics into utility levels.

[17] See Wasserman and Faust (1994) for references.

2.1.1 Network Formation Process

The process of link formation follows a *stochastic best-response dynamic*, as in Blume (1993), and generates a Markov chain of networks. The main ingredients of this process are random meetings and utility maximization. The implicit assumption is that individuals meet very frequently and have the opportunity to revise their links.

Meeting Technology. At the beginning of each period an agent i is randomly selected from the population, and he meets another individual j according to a matching technology. Formally, the *meeting process* is a stochastic sequence $m = \{m^t\}_{t=1}^{\infty}$ with support $\mathcal{I} \times \mathcal{I}$. The realizations of the meeting process are ordered pairs $m^t = \{i, j\}$, indicating which agent i should play and which link g_{ij} can be updated at period t.[18]

The probability that player i is randomly chosen from the population and meets agent j is defined as

$$\Pr\left(m^t = ij | g^{t-1}, X\right) = \rho\left(g^{t-1}, X_i, X_j\right) \tag{1}$$

where $\sum_{i=1}^{n}\sum_{j=1}^{n}\rho\left(g, X_i, X_j\right) = 1$ for any $g \in \mathcal{G}$. The matching probability depends on the current network (e.g. the existence of a common friend between i and j) and the characteristics of the pair. This structure includes matching technologies with a bias for same-type individuals as in Currarini et al. (2009). The simplest example of matching technology is an i.i.d. discrete uniform process with $\rho\left(g^{t-1}, X_i, X_j\right) = \frac{1}{n(n-1)}$. An example with bias for same-type agents is $\rho\left(g^{t-1}, X_i, X_j\right) \propto \exp\left[-d\left(X_i, X_j\right)\right]$, where $d\left(\cdot,\cdot\right)$ is a distance function.

Utility Maximization Conditional on the meeting $m^t = ij$, player i updates the link ij to maximize his current utility, taking the existing network g^t_{-ij} as given. The agents have *complete information* since they can observe the entire network and the individual attributes of all agents. Before updating his link to j, individual i receives an idiosyncratic shock $\varepsilon \sim F\left(\varepsilon\right)$ to his preferences that the econometrician cannot observe. This shock is meant to model unobservable events that could influence the utility of a link, e.g. mood, gossips, fights, etc. Player i links player j at time t, i.e. $g^t_{ij} = 1$, if and only if it is a best response to the current network configuration, i.e. $g^t_{ij} = 1$ if and only if

$$U_i\left(g^t_{ij} = 1, g^{t-1}_{-ij}, X\right) + \varepsilon_{1t} \geq U_i\left(g^t_{ij} = 0, g^{t-1}_{-ij}, X\right) + \varepsilon_{0t}. \tag{2}$$

[18] Several models incorporate a matching technology in the network formation process. Jackson and Watts (2002) assume individuals meet randomly according to a discrete uniform distribution. Currarini et al. (2009) introduce a matching process that is biased towards individuals of the same type, similar to the one modeled here.

I assume that when the equality holds, the agent plays the status quo.[19] The match formation process generates a sequence $[g^0, g^1, ..., g^t]$ of networks. In each period only one element of the random matrix G is updated, conditioning on the existing network. Therefore the sequence is a Markov chain, with transition probabilities determined by the meeting process and agents' linking choices.[20]

2.1.2 Preferences

The preferences are defined over networks and individual characteristics. The utility of player i from a network g and population attributes $X = (X_1, ..., X_n)$ is given by

$$U_i(g, X) = \underbrace{\sum_{j=1}^{n} g_{ij} u_{ij}}_{\text{direct friends}} + \underbrace{\sum_{j=1}^{n} g_{ij} g_{ji} m_{ij}}_{\text{mutual friends}} + \underbrace{\sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{jk} v_{ik}}_{\text{friends of friends}} + \underbrace{\sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ki} w_{kj}}_{\text{popularity}} \quad (4)$$

where $u_{ij} \equiv u(X_i, X_j)$, $m_{ij} \equiv m(X_i, X_j)$, $v_{ij} \equiv v(X_i, X_j)$ and $w_{ij} \equiv w(X_i, X_j)$ are (bounded) real-valued functions of the attributes. The utility of the network is the sum of the net benefits received from each link. The total benefit from an *additional link* has four components.

- When the agent links another individual, she receives an additional *direct* net benefit u_{ij}. The direct utility includes both costs and benefits and it may possibly be negative: when only homophily enters payoffs of direct links, the net utility u_{ij} is positive if i and j belong to the same group, while it is negative when they are of different types. This is illustrated in Panel A of Figure 1 with a simple network of 8 agents. Each agent can belong to either the blue group or the yellow group. The link that agent 4 forms to individual 5 provides different direct utility in the two networks, since the identity of 5 is different: blue for the left network and yellow for the right one. In many models this component is parameterized as $u_{ij} = b_{ij} - c_{ij}$, where b_{ij} indicates the (gross) benefit

[19] This assumption does not affect the main result and is relevant only when the distribution of the preference shocks is discrete.

[20] The derivation of the transition matrix is straightforward. The set of all possible states corresponds to the set of all possible binary $n \times n$ matrices $\mathcal{G}$; the probability of transition from a network $g^t = g$ to next period network $g^{t+1} = (g'_{ij}, g_{-ij})$ is

$$\rho(g, X_i, X_j) \Pr\left[U_i\left(g'_{ij}, g_{-ij}, X\right) + \varepsilon\left(g'_{ij}\right) \geq U_i\left(g_{ij}, g_{-ij}, X\right) + \varepsilon\left(g_{ij}\right)\right] \quad (3)$$

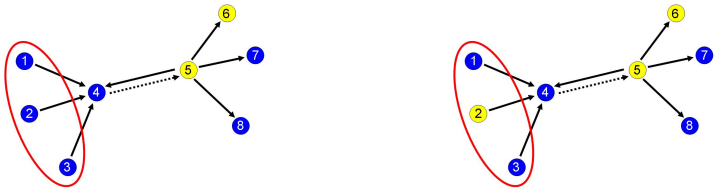
The transition probability is zero if the networks g and g' differ by more than one element.

Figure 1: Components of the utility function

A. Direct friends

B. Mutual friends

C. Friends of friends



D. Popularity

The network contains $n = 8$ agents, belonging to two groups: blue and yellow. All the panels show a situation in which player 4 decides whether to form a link to individual 5 (the dashed arrow from 4 to 5). Agent 4 receives different direct utility when he links a blue (Panel A, left) or a yellow (Panel A, right) individual. Agent 4's utility from an additional link is different if the link is unilateral (Panel B, left) or reciprocated (Panel B, right). Furthermore, agent 4's utility from friends of friends varies with their socioeconomic composition: 3 blue individuals (Panel C, left) provide different utility than 2 blue and 1 yellow agents (Panel C, right). Finally, agent 4 values how his new link affects his popularity, since he creates a new indirect friendship for those who already have a link to him (agents 1,2 and 3). The utility of a link to agent 5 (which is yellow) when agents 1,2 and 3 are all blue (Panel D, left) is different than when agent 2 is yellow and 1 and 2 are blue (Panel D, right).

and c_{ij} the cost of forming the additional link g_{ij}. I use the notation u_{ij}, since it does not require assumptions on the cost function.

- Agents receive additional utility m_{ij} if the link is mutual; friendship is valued differently if the other agent reciprocates. The idea is that an agent may perceive another individual as a friend, but that person may not perceive the relationship in the same way. Panel B of Figure 1 isolates this component: a link from agent 4 to agent 5 has a different value if agent 5 reciprocates (right network).

- The players value the composition of friends of friends. When i is deciding whether to link j, she observes j's friends and their socioeconomic characteristics. Each of j's friend provides additional utility $v(X_i, X_k)$ to i. In this model, an agent who has the opportunity to form an additional link, values a white student with three Hispanic friends as a different good than a white student with two white friends and one African American friend.[21] In other words, individuals value both *exogenous* heterogeneity and *endogenous* heterogeneity: the former is determined by the socioeconomic characteristics of the agents, while the latter arises endogenously with the process of network formation. I assume that only friends of friends are valuable and they are perfect substitutes: individuals do not receive utility from two-links-away friends. In Panel C of Figure 1, from the perspective of agent 4, agent 5 in the left network is a different good than agent 5 in the right network, since the composition of his friends is different.

- The fourth component corresponds to a *popularity effect*. Consider Panel D in Figure 1. When agent 4 forms a link to agent 5, he automatically creates an indirect link for agents 1, 2 and 3. Thus agent 4 generates an externality. For example, suppose there is homophily in indirect links. Then in the left network the externality is negative for all three agents (1, 2 and 3); and in the right network it is negative for 1 and 3, but positive for 2. Therefore, in the left network the popularity of 4 goes down, while in the right network the fall in popularity is less pronounced. [22]

[21] A similar assumption is used in De Marti and Zenou (2009) where the agents' cost of linking depend on the racial composition of friends of friends. Their model is an extension of the connection model of Jackson and Wolinsky (1996), and the links are formed with mutual consent. The corresponding network is undirected.

[22] One can contemplate an alternative interpretation of this last component. One can view it as a feature that captures *forward-looking* behavior in the model in a reduced form, since the "popularity" affects how more/less likely the other agents are to maintain or create a link to individual i in future meetings.

2.2 Equilibrium Analysis

I impose an additional assumption on the functional forms of the utility functions. The assumption is not too strong, but it provides an important identification restriction. I assume that the utility m_{ij} obtained from mutual links is symmetric and that the utility of an indirect link v_{ij} has the same functional form as the utility from the popularity effect w_{ij}.

ASSUMPTION 1 ***(Preferences)*** *The utility function satisfies the following restrictions*

$$\begin{aligned} m(X_i, X_j) &= m(X_j, X_i) \text{ for all } i, j \in \mathcal{I} \\ w(X_k, X_j) &= v(X_k, X_j) \text{ for all } k, j \in \mathcal{I} \end{aligned}$$

therefore the utility function is

$$U_i(g, X) = \sum_{j=1}^{n} g_{ij} u_{ij} + \sum_{j=1}^{n} g_{ij} g_{ji} m_{ij} + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{jk} v_{ik} + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ki} v_{kj} \tag{5}$$

The symmetry in m_{ij} does not imply that a mutual link between i and j gives both the same utility. Indeed if i and j have a mutual link, they receive the same common utility component (m_{ij}) but they may perceive that particular friendship in a different way, as long as the utility from direct or indirect links are different for i and j. As a result, two individuals with the same exogenous characteristics $X_i = X_j$ (say two males, whites, enrolled in eleventh grade) who form a mutual link receive the same u_{ij} and m_{ij}, but they may have different utilities from that additional link because of the composition of their friends of friends and their popularity. Therefore, this part of the assumption helps in identifying the utility from indirect links and popularity.

The second restriction is more technical. When i forms a link to j, i creates an externality for all k's who have linked her: any such k now has an additional indirect friend, i.e. j, who agent k values by an amount $v(X_k, X_j)$. When $w(X_k, X_j) = v(X_k, X_j)$, an individual i values his popularity effect as much as k values the indirect link to j, i.e., i internalizes the externality he creates.

Assumption 1 is the main ingredient that guarantees a closed form solution for the stationary equilibrium of the model. Without this assumption, the model would still have a unique stationary equilibrium, however it would be impossible to characterize the likelihood function in closed form.[23] The first part of the assumption is a normalization of the utility

[23] Estimation of such a model could be performed using Approximate Bayesian Computations (see Marjoram et al. (2003) for example), but the computational burden is even more challenging.

function that allows identification for the utility of indirect links and popularity. The second part of the assumption is an identification restriction, that guarantees the model's coherency in the sense of Tamer (2003). In simple words, this part of the assumption guarantees that the system of conditional linking probabilities implied by the model generates a proper joint distribution of the network matrix.[24]

The assumption delivers a very simple characterization of the stationary equilibrium. The following proposition highlights a crucial result of this paper.

PROPOSITION 1 ***(Potential Function)***
Under Assumption 1, the deterministic component of the incentives of any player in any state of the network are summarized by a ***potential function****,* $Q : \mathcal{G} \times \mathcal{X} \to \mathbb{R}$

$$Q(g, X) = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} u_{ij} + \sum_{i=1}^{n} \sum_{j>i}^{n} g_{ij} g_{ji} m_{ij} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ij} g_{jk} v_{ik}, \tag{6}$$

and the network formation game is a Potential Game.

Proof. See Appendix A ■

The intuition for the result is simple. Under the restrictions of Assumption 1, for any player i and any link g_{ij} we have

$$Q(g_{ij}, g_{-ij}, X) - Q(1 - g_{ij}, g_{-ij}, X) = U_i(g_{ij}, g_{-ij}, X) - U_i(1 - g_{ij}, g_{-ij}, X)$$

Consider two networks, $g = (g_{ij}, g_{-ij})$ and $g' = (1 - g_{ij}, g_{-ij})$, that differ only with respect to one link, g_{ij}, chosen by individual i: the difference in utility that agent i receives from the two networks, $U_i(g, X) - U_i(g', X)$, is exactly equal to the difference of the *potential* function evaluated at the two networks, $Q(g, X) - Q(g', X)$. That is, the potential is an aggregate function that summarizes both the state of the network and the deterministic incentives of the players in each state.

Characterizing the network formation as a potential game facilitates analysis. In order to compute the equilibria of the model, there is no need to keep track of each player's behavior: the potential function contains all the relevant information. This property is key for the analysis of networks with many players: the usual check for existence of profitable deviations from the Nash equilibrium can be performed using the potential, instead of checking each player's possible deviation in sequence.

[24] Similar restrictions are also encountered in spatial econometrics models (Besag, 1974) and in the literature on qualitative response models (Heckman, 1978; Amemiya, 1981)

It should be emphasized that the potential $Q(g, X)$ is not equivalent to the welfare function $W(g, X)$, that describes the total utility of all agents in the network,

$$\begin{aligned} W(g, X) &= \sum_{i=1}^{n} U_i(g, X) \\ &= Q(g, X) + \sum_{i=1}^{n}\sum_{j>i}^{n} g_{ij}g_{ji}m_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ k\neq i,j}}^{n} g_{ij}g_{ki}v_{kj} \end{aligned}$$

To analyze the long run behavior of the model, I impose more structure on the matching technology.[25]

ASSUMPTION 2 ***(Meeting Process)*** *Any meeting is possible, i.e., for any* $ij \in \mathcal{I} \times \mathcal{I}$

$$\rho(g^{t-1}, X_i, X_j) > 0 \tag{7}$$

The meeting process is such that any individual can be chosen and any pair of agents can meet. This assumption guarantees that any equilibrium network can be reached with positive probability. For example, a discrete uniform distribution satisfies this assumption.

It is helpful to consider a *special case* of the model, in which there are no preference shocks: the characterization of equilibria and long run behavior for such model provides intuition about the dynamic properties of the full structural model.
Let $\mathcal{N}(g)$ be the set of networks that differ from g by only one element of the matrix, i.e.

$$\mathcal{N}(g) \equiv \{g' : g' = (g'_{ij}, g_{-ij}), \text{ for all } g'_{ij} \neq g_{ij}, \text{ for all } i, j \in \mathcal{I}\}. \tag{8}$$

A Nash network is defined as a network in which any player has no profitable deviations from his current linking strategy, when randomly selected from the population. The following results characterize the set of the pure-strategy Nash equilibria and the long run behavior of the model with no shocks.

PROPOSITION 2 ***(Model without Shocks: Equilibria and Long Run)***
Consider the model without idiosyncratic preference shocks. Under Assumptions 1 and 2:

1. *There exists at least one pure-strategy Nash equilibrium network*

[25] Christakis et al. (2010) assume that individuals can meet only once and their link remains in place forever. This assumption is convenient when estimating a large network, but it does not allow the characterization of the stationary equilibrium.

2. *The set $\mathcal{NE}(\mathcal{G}, X, U)$ of all pure-strategy Nash equilibria of the network formation game is completely characterized by the local maxima of the potential function.*

$$\mathcal{NE}(\mathcal{G}, X, U) = \left\{ g^* : g^* = \arg \max_{g \in \mathcal{N}(g^*)} Q(g, X) \right\} \tag{9}$$

3. *Any pure-strategy Nash equilibrium is an* ***absorbing*** *state.*

4. *As $t \to \infty$, the network converges to one of the Nash networks with probability 1.*

Proof. In Appendix A ■

Suppose that the current network is a Nash network. As a consequence, if an agent deviates from the current linking strategy, he receives less utility.[26] Since the change in utility for any agent is equivalent to the change in potential, any deviation from the Nash network must decrease the potential. It follows that the Nash network must be a local maximizer of the potential function over the set of networks that differ from the current network for at most one link.

Furthermore, the network must converge to one of the Nash Equilibria in the long run, independently of the initial network. Suppose an agent is drawn from the meeting process. Such agent will play a best response to the current network configuration. Therefore, his utility cannot decrease. This holds for any player and any period. It follows that the potential is nondecreasing over time. Since there is a finite number of possible networks, in the long run, the sequence of networks must reach a local maximum of the potential, i.e., a Nash equilibrium.

With the intuition from the simpler model in mind, we can now analyze the full structural model with preference shocks. In the full model there is a high probability of hitting a Nash network. However, the shocks allow the network to escape from such networks: this makes the model ergodic and eliminates absorbing states.

I make the following parametric assumption on the shocks, that allows me to characterize the stationary distribution and transition probabilities.

ASSUMPTION 3 ***(Idiosyncratic Shocks)*** *The shock follows a Type I extreme value distribution, i.i.d. among links and across time.*

[26] When the utility from the equilibrium and the deviation is the the same, the agent plays the status quo, i.e., the Nash strategy.

The probability of a link between i and j, given a meeting $m^t = ij$ and previous period network configuration g^{t-1}

$$\Pr\left(g_{ij}^t = 1 \middle| g_{-ij}^{t-1}, X\right) = \Pr\left[\varepsilon_{0t} - \varepsilon_{1t} \leq U_i\left(1, g_{-ij}^{t-1}, X\right) - U_i\left(0, g_{-ij}^{t-1}, X\right)\right] \tag{10}$$

$$= \frac{\exp\left[u_{ij} + g_{ji}^{t-1} m_{ij} + \sum_{k \neq i,j} g_{jk}^{t-1} v_{ik} + \sum_{k \neq i,j} g_{ki}^{t-1} v_{kj}\right]}{1 + \exp\left[u_{ij} + g_{ji}^{t-1} m_{ij} + \sum_{k \neq i,j} g_{jk}^{t-1} v_{ik} + \sum_{k \neq i,j} g_{ki}^{t-1} v_{kj}\right]} \tag{11}$$

Under Assumptions 1-3, the network evolves as a Markov chain with transition probabilities given by the conditional choice probabilities (11) and the probability law of the meeting process m^t.

One can show that the sequence $[g^0, g^1, ..., g^t]$ is:

1. *irreducible*, i.e. every state of the network can be reached with positive probability in a finite number of steps

2. *aperiodic*, i.e. the chain does not get trapped in cycles, because the probability of moving from a state to another is always positive under the extreme value assumption

Intuitively, since the meeting probability $\Pr(m^t = ij) > 0$ for all ij, there is always a positive probability of reaching a new network in which the link g_{ij} can be updated. The logistic shock assumption implies that there is always a positive probability of switching to another state of the network, thus eliminating absorbing states.

THEOREM 1 ***(Uniqueness and Characterization of Stationary Equilibrium)*** *Consider the network formation game with idiosyncratic shocks, under Assumptions 1-3.*

1. *There exists a unique stationary distribution $\pi(g, X)$, i.e.,*

$$\lim_{t \to \infty} P\left(G^t = g \middle| G^0 = g^0, X\right) = \pi\left(g, X\right). \tag{12}$$

2. *Suppose that the meeting probability of i and j does not depend on the existence of a link between them, i.e.,*

$$\rho\left(g^{t-1}, X_i, X_j\right) = \rho\left(g_{-ij}^{t-1}, X_i, X_j\right). \tag{13}$$

for any $i, j \in \mathcal{I}$. Then the stationary distribution $\pi(g, X)$ is

$$\pi\left(g, X\right) = \frac{\exp\left[Q\left(g, X\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X\right)\right]}, \tag{14}$$

where $Q(g, X)$ is the potential function (6).

Proof. In Appendix A ■

The first part of the proposition follows directly from the irreducibility and aperiodicity of the Markov process generated by the network formation game. The uniqueness of the stationary distribution is crucial in estimation, since one does not need to worry about multiple equilibria. Furthermore, the stationary equilibrium characterizes the likelihood of observing a specific network configuration in the data. As a consequence, I can estimate the structural parameters from observations of only *one network at a specific point in time*, under the assumption that the observed network is drawn from the stationary equilibrium.

The second part of the proposition provides a closed-form solution for the stationary distribution. The latter can be interpreted as the probability of observing a specific network structure when the network is observed in the long run. In the long run, the system of interacting agents will visit more often those states/networks that have high potential, which correspond to the Nash equilibria described in Proposition 2. Therefore a high proportion of the possible networks generated by the network formation game, will correspond to Nash networks.

The stationary distribution $\pi(g, X)$ includes a normalizing constant

$$c(\mathcal{G}, X) \equiv \sum_{\omega \in \mathcal{G}} \exp\left[Q(\omega, X)\right] \tag{15}$$

that reflects the fact that it is a proper probability distribution. Unfortunately, this normalizing constant greatly complicates estimation, since it cannot be evaluated exactly or approximated with precision. How this is circumvented is explained in the next section.

3 Estimation Strategy

3.1 Computational Problem

To estimate the model, I assume that the utility functions depend on a vector of parameters $\theta = (\theta_u, \theta_m, \theta_v)$:

$$\begin{aligned} u_{ij} &= u(X_i, X_j, \theta_u) \\ m_{ij} &= m(X_i, X_j, \theta_m) \\ v_{ij} &= v(X_i, X_j, \theta_v) \end{aligned}$$

The goal is to recover the parameters' posterior distribution, given the data, the model and the prior. Let $p(\theta)$ be the prior distribution. Given the likelihood function $\pi(g, X, \theta)$ of the

observed data (g, X), the posterior distribution of θ can be written as

$$p(\theta|g, X) = \frac{\pi(g, X, \theta)\, p(\theta)}{\int_\Theta \pi(g, X, \theta)\, p(\theta)\, d\theta}. \tag{16}$$

Intuitively, the posterior is the distribution of the parameters that are most likely to generate the data g, given the model $\pi(g, X, \theta)$ and the prior $p(\theta)$.

Estimation of the posterior faces two computational challenges. First, the posterior depends on the normalizing integral $\int_\Theta \pi(g, X, \theta)\, p(\theta)\, d\theta$. This problem is common to any Bayesian analysis, and is often solved using a Metropolis-Hastings algorithm that avoids direct computation of the integral. This algorithm generates a Markov chain of parameters $\{\theta^0, \theta^1, \theta^2, ..\}$ whose unique invariant distribution is the posterior (16). The empirical distribution of the chain is used as estimate of the posterior. A Weak Law of Large Numbers guarantees that sample averages computed using the sequence generated by the chain, converge to the expectation under the posterior distribution.

At each iteration s, with current parameter $\theta_s = \theta$, a new parameter vector θ' is proposed from a distribution $q_\theta(\cdot|\theta)$. At iteration $t+1$ the new parameter θ_{s+1} is updated according to

$$\theta_{s+1} = \begin{cases} \theta' & \text{with prob. } \alpha(\theta, \theta') \\ \theta & \text{with prob. } 1 - \alpha(\theta, \theta'), \end{cases} \tag{17}$$

where $\alpha(\theta, \theta')$ is computed as

$$\alpha(\theta, \theta') = \min\left\{1, \frac{p(\theta'|g, X)\, q_\theta(\theta|\theta')}{p(\theta|g, X)\, q_\theta(\theta'|\theta)}\right\} \tag{18}$$

The appealing feature of this scheme is that one does not need to evaluate the integral to compute $\alpha(\theta, \theta')$, because the ratio of the posteriors is $p(\theta'|g, X)/p(\theta|g, X) = \frac{\pi(g,X,\theta')p(\theta')}{\pi(g,X,\theta)p(\theta)}$.

However, the naive version of the Metropolis-Hastings algorithm cannot be used for the model formulated above. The likelihood function $\pi(g, X, \theta)$ is known up to a normalizing constant that cannot be computed in practice. The acceptance probability in (18) can be rewritten to make the likelihood contribution explicit

$$\begin{aligned} \alpha(\theta, \theta') &= \min\left\{1, \frac{\frac{\exp[Q(g,X,\theta')]}{c(\mathcal{G},X,\theta')} p(\theta')\, q_\theta(\theta|\theta')}{\frac{\exp[Q(g,X,\theta)]}{c(\mathcal{G},X,\theta)} p(\theta)\, q_\theta(\theta'|\theta)}\right\} \\ &= \min\left\{1, \frac{\exp[Q(g, X, \theta')]}{\exp[Q(g, X, \theta)]} \frac{c(\mathcal{G}, X, \theta)}{c(\mathcal{G}, X, \theta')} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)}\right\}. \end{aligned}$$

The Metropolis-Hastings acceptance $\alpha(\theta, \theta')$ depends on the ratio $c(\mathcal{G}, X, \theta)/c(\mathcal{G}, X, \theta')$, whose exact evaluation is computationally infeasible even for very small networks. To

be concrete, consider a small network with $n = 10$ agents. From (15) we know that $c(\mathcal{G}, X, \theta) = \sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X, \theta)]$. To compute the constant at the current parameter θ we would need to evaluate the potential function for all $2^{90} \backsimeq 10^{27}$ possible networks with 10 agents and compute their sum. This task would take a very long time even for a state-of-the art supercomputer. In general with a network containing n players, we have to sum over $2^{n(n-1)}$ possible network configurations.[27]

3.2 Estimation Algorithm

To solve the estimation problem, I develop a variation of the *exchange algorithm*, first developed by Murray et al. (2006). This algorithm uses a double Metropolis-Hastings step to avoid the computation of the normalizing constant $c(\mathcal{G}, X, \theta)$ in the likelihood. This improvement comes with a cost: the algorithm may produce MCMC chains that have very poor mixing properties (Caimo and Friel, 2010) and high autocorrelation. I partially correct for this problem by carefully calibrating the proposal distribution.[28]

While several authors have proposed similar algorithms in the related literature on Exponential Random Graphs Models (ERGM),[29] the models estimated with this methodology typically have very few parameters and use data from very small networks. To the best of my knowledge, this is the first attempt to estimate a high dimensional model using data from multiple networks.

In this section I describe the algorithm for a single network, while in the appendix I provide the extension for multiple independent networks.[30] This is especially important for policy: schools may have unobserved differences that impact the network formation process and using multiple networks may partially correct for that.

[27] A supercomputer that can compute 10^{12} potential functions in 1 second would take almost 40 million years to compute the constant once for a network with $n = 10$ players. The schools used in the empirical section have between 20 and 159 enrolled students. This translates into a minimum of 2^{380} and a maximum of 2^{25122} possible network configurations.

[28] In this paper I use a random walk proposal. Alternatively one could update the parameters in blocks or use recent random block techniques as in Chib and Ramamurthy (2009) to improve convergence and mixing.

[29] Caimo and Friel (2010) use the exchange algorithm to estimate ERGM. They improve the mixing of the sampler using the snooker algorithm. Koskinen (2008) proposes the Linked Importance Sampler Auxiliary variable (LISA) algorithm, which uses importance sampling to provide an estimate of the acceptance probability. Another variation of the algorithm is used in Liang (2010).

[30] When the data consist of several independent school networks, I use a parallel version of the algorithm that stores each network in a different processor. Each processor runs the simulations independently and the final results are summarized in the master processor, that updates the parameters for next iteration. Details in Appendix.

The idea of the algorithm is to sample from an augmented distribution using an auxiliary variable. At each iteration, the algorithm proposes a new parameter vector θ', drawn from a suitable proposal distribution $q_\theta(\theta'|\theta)$; in the second step, it samples a network g' (the auxiliary variable) from the likelihood $\pi(g', X, \theta')$; finally, the proposed parameter is accepted with a probability $\alpha_{ex}(\theta, \theta')$, such that the Markov chain of parameters generated by these update rules, has the posterior (16) as unique invariant distribution.

I first describe the algorithm used to sample a network from the stationary distribution of the model; then I provide the full algorithm for estimation of the posterior.

3.2.1 Network Simulations

To use the exchange algorithm, I need to draw random samples from the stationary distribution of the network formation model. Direct simulation is not possible because the normalizing constant $c(\mathcal{G}, X, \theta)$ is computationally infeasible, for the reasons explained above. Therefore I rely on Markov Chain Monte Carlo simulation methods.

The algorithm used in this paper is similar to the Metropolis-Hastings algorithm proposed in Snijders (2002).[31] For a fixed parameter value θ, the algorithm simulates a Markov chain of networks whose unique invariant distribution is (14). As the number of iterations R becomes large, the simulated networks are (approximate) samples from the stationary distribution of the model evaluated at parameter θ.

ALGORITHM 1 *Fix a parameter value θ. At iteration r, with current network g_r*

1. *Propose a network g' from a proposal distribution*

$$g' \sim q_g(g'|g_r) \tag{19}$$

2. *Update the network according to*

$$g_{r+1} = \begin{cases} g' & \text{with prob. } \alpha_{mh}(g_r, g') \\ g_r & \text{with prob. } 1 - \alpha_{mh}(g_r, g') \end{cases} \tag{20}$$

where

$$\alpha_{mh}(g_r, g') = \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g_r, X, \theta)]} \frac{q_g(g_r|g')}{q_g(g'|g_r)}\right\} \tag{21}$$

[31] I also experimented with the Simulated Tempering algorithm proposed in Mele (2010). The latter is extremely useful when the stationary distribution of the network formation model has more than one mode. It also improves the mixing of the chain. However, it does so by increasing the time needed to collect a sample. In this context, a set of experiments with artificial data revealed virtually no difference between the Simulated Tempering results and the simpler Metropolis-Hastings updates, so I use the latter in this paper.

At each iteration a random network g' is proposed, and the update is accepted with probability $\alpha_{mh}(g_r, g')$. The main advantage of this simulation strategy is that the acceptance ratio (21) does not contain the normalizing constant $c(\mathcal{G}, X, \theta)$ of the stationary distribution. Each quantity in the acceptance ratio can be computed exactly.

The Metropolis-Hastings structure of the algorithm guarantees that the sampled networks are drawn from the stationary equilibrium of the model.

PROPOSITION 3 *The updates in ALGORITHM 1 produce a Markov Chain of networks.*

1. *The unique invariant distribution of ALGORITHM 1 is the stationary equilibrium of the model at parameter θ*

$$\lim_{r\to\infty} \Pr(G_r = g | G_0 = g_0, X, \theta) = \pi(g, X, \theta) \tag{22}$$

2. *The Markov Chain generated by ALGORITHM 1 is uniformly ergodic, i.e. there exist constants $B > 0$ and $\kappa \in (0,1)$ such that*

$$\max_{g_0 \in \mathcal{G}} \left\| \mathcal{P}_\theta^{(r)}(\cdot | g_0) - \pi(\cdot, X, \theta) \right\|_{TV} \leq B\kappa^r \tag{23}$$

   where $\mathcal{P}_\theta^{(r)}(g|g_0)$ is the r-th step transition probability of the algorithm, starting at network g_0 and using parameter θ; and $\|\cdot\|_{TV}$ is the total variation distance.

3. *For any initial network g_0, the rate of convergence can be bounded as*

$$2 \sup_{g\in\mathcal{G}} \left| \pi(g,\theta) - \mathcal{P}_\theta^{(r)}(g|g_0) \right| \leq \sqrt{\frac{1-\pi(g_0,\theta)}{\pi(g_0,\theta)}} \lambda_*(\theta)^r \tag{24}$$

   where $\lambda_*(\theta) = \max\{\lambda_2(\theta), |\lambda_{2^{n(n-1)}}(\theta)|\}$ is the second highest eigenvalue of ALGORITHM 1's transition matrix.

Proof. See Appendix B ■

The first part of the proposition confirms that the algorithm converges to the stationary equilibrium of the model. The second part establishes uniform convergence to the stationary distribution. The last part of the proposition is a classical results on Markov chain convergence (see Diaconis and Stroock (1991) or Meyn and Tweedie (2009)). For certain regions of the parameter space, the convergence would be fast, while for other parameter regions the rate of convergence could be extremely slow. Recent work on similar algorithms for the simulation of exponential random graphs, have shown that in several cases the convergence rates are so slow that the simulation is infeasible. Diaconis and Chatterjee (2011) and Bhamidi

et al. (forthcoming) show that for the simplest Erdos-Renyi random graph, the bound to the convergence rate scales up with $m\log(m)$, where m is the total number of states. In our setting, $m = 2^{n(n-1)}$. However, their bounds are computed to be valid for any parameter θ and for any initial network g_0, therefore a good choice of the initial network would improve convergence as indicated in (24).

To increase the speed of convergence and decrease such problems, in the empirical implementation of the algorithm, I use several alternative proposals $q_g(\cdot|\cdot)$. First, a move that updates only one link per iteration, proposing to swap the link value. At each iteration a random pair of agents (i, j) is selected from a discrete uniform distribution, and it is proposed to swap the value of the link g_{ij} to $1 - g_{ij}$. Second, to improve convergence, I allow the sampler to propose bigger moves: with a probability p_{inv}, the sampler proposes to swap the entire network matrix.[32] With a small probability p_{inv}, the sampler proposes a new network $g' = \mathbf{1} - g$, which is accepted with probability $\alpha_{mh}(g, g')$.

3.2.2 Posterior Simulation

I propose a modified version of the exchange algorithm developed by Murray et al. (2006) to sample from distributions with intractable constants. In the original algorithm, one needs to draw *exact* samples from the stationary equilibrium of the model. However, this would require an enormous number of steps using the network simulation algorithm. My strategy is instead to exploit the result in Lemma 1 to decrease the number of simulations needed to collect an approximate sample from the stationary equilibrium. The samples from the posterior distribution are generated using the following steps

ALGORITHM 2 ***(APPROXIMATE EXCHANGE ALGORITHM)***
Fix the number of simulations R. At each iteration t, with current parameter $\theta_t = \theta$ and network data g:

1. *Propose a new parameter θ' from a distribution $q_\theta(\cdot|\theta)$,*

$$\theta' \sim q_\theta(\cdot|\theta). \tag{25}$$

2. *Start **ALGORITHM 1** at the observed network g, iterating for R steps using parameter θ' and collect the last simulated network g'*

$$g' \sim \mathcal{P}_{\theta'}^{(R)}(g'|g). \tag{26}$$

[32] This move is suggested in Geyer (1992) and Snijders (2002). Snijders (2002) argues that this is particularly useful in case of a bimodal distribution.

3. Update the parameter according to

$$\theta_{t+1} = \begin{cases} \theta' & \text{with prob. } \alpha_{ex}(\theta, \theta', g', g) \\ \theta & \text{with prob. } 1 - \alpha_{ex}(\theta, \theta', g', g) \end{cases}$$

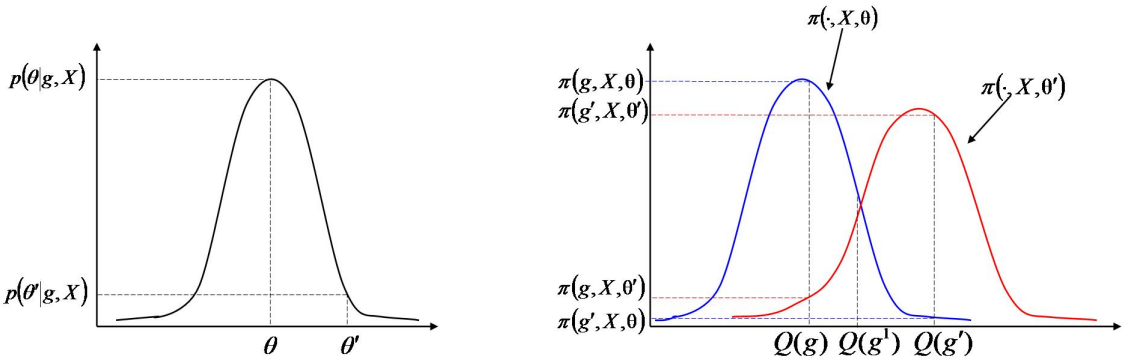
where

$$\alpha_{ex}(\theta, \theta', g', g) = \min \left\{ 1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]} \right\}. \quad (27)$$

The appeal of this algorithm is that all quantities in the acceptance ratio (27) can be evaluated: there are no integrals or normalizing constants to compute. I provide the algorithm details, and the relative proofs of convergence to the posterior and some evidence on mixing in Appendix B. The algorithm used to estimate the model using multiple school networks on parallel processors is an extension of ALGORITHM 2. I also present it in Appendix B. Here I explain intuitively why the sampler works, with the help of Figure 2.

For ease of exposition, suppose that the prior is relatively flat, so that $p(\theta)/p(\theta') \simeq 1$.

Figure 2: The Exchange Algorithm



A. Posterior Distribution B. Two Stationary Equilibria

The graph on the left is the posterior distribution, given the data. The graph on the right represents two stationary equilibria of the model, one at parameter θ (blue) and one at parameter θ' (red). The iteration t starts with parameter θ. It is proposed to update the parameter using proposal θ'. The algorithm start sampling networks from the stationary distribution at parameter θ' (red) and quickly moves from g to g'. The probability of accepting the proposed parameter θ' is proportional to the ratio $\frac{\pi(g',X,\theta)}{\pi(g',X,\theta')} \frac{\pi(g,X,\theta')}{\pi(g,X,\theta)}$, which is small as indicated in the graph. In summary, a move from the high density region of the posterior (θ) to a low density region (θ') is likely to be rejected. For the same reasoning a move from θ' to θ is very likely to be accepted. Therefore the algorithm produces samples from the correct posterior distribution.

Suppose we start the sampler from a parameter θ that has high posterior probability, given the data g. That is, there is good agreement between the data and the parameter, so it is likely that the data are generated from a model with parameter θ. This is displayed on the left panel of Figure 2. Now, suppose we propose a parameter θ' that belongs to a low probability region of the posterior. This means that there is a low probability that the observed network g is generated by parameter θ'. As a consequence the ratio

$$\frac{p(\theta'|g,X)}{p(\theta|g,X)} \simeq \frac{\pi(g,X,\theta')}{\pi(g,X,\theta)}$$

would be very small, as indicated in the right panel of Figure 2. Let's start the network simulations using parameter θ'. The sequence of simulated networks will start approaching the new stationary distribution $\pi(\cdot,X,\theta')$, moving away from the stationary distribution $\pi(\cdot,X,\theta)$. This is indicated in Figure 2 with a simulation of 2 steps: starting from g we obtain two networks, g^1 and g'. Network g' is closer to a high probability region of $\pi(\cdot,X,\theta')$ than to a high probability region of $\pi(\cdot,X,\theta)$, as long as the algorithm was run for a sufficiently large number of steps R. It also follows that the ratio

$$\frac{\pi(g',X,\theta)}{\pi(g',X,\theta')}$$

is small. Notice that

$$\begin{aligned}\frac{\pi(g',X,\theta)}{\pi(g',X,\theta')}\frac{\pi(g,X,\theta')}{\pi(g,X,\theta)} &= \frac{\exp\left[Q(g',X,\theta)\right]}{\exp\left[Q(g,X,\theta)\right]}\frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]}\frac{c(\mathcal{G},X,\theta')}{c(\mathcal{G},X,\theta)}\frac{c(\mathcal{G},X,\theta)}{c(\mathcal{G},X,\theta')} \\ &= \frac{\exp\left[Q(g',X,\theta)\right]}{\exp\left[Q(g,X,\theta)\right]}\frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]}.\end{aligned}$$

This ratio is contained in (27). As a consequence the acceptance ratio of the exchange algorithm is low and the proposed parameter θ' is very likely to be rejected. Let's repeat the reasoning while starting the sampler at θ' and proposing an update θ: this proposal is very likely to be accepted by the same intuitive argument.

In summary, the sampler is likely to accept proposals that move towards high density regions of the posterior, but it is likely to reject proposals that move towards low density regions of the posterior. Therefore, it produces samples of parameters that closely resemble draws the posterior distribution. The formal statement about ergodicity is contained in the following theorem.

THEOREM 2 *(Ergodicity of the Approximate Exchange Algorithm). The approximate exchange algorithm is ergodic, and it converges to the correct posterior distribution.*

1. (Convergence)) *Let* $\widetilde{P}_R^{(s)}(\theta_0, \cdot)$ *be the s-th step transition of the approximate exchange algorithm, when the auxiliary network is sampled using* R *steps of the network simulation algorithm and the initial parameter of the simulation is* θ_0*. Let* $\|\cdot\|_{TV}$ *be the total variation distance and* $p(\cdot|g, X)$ *the posterior distribution.*
Then, for any $\epsilon > 0$ *there exist* $R_0 \in \mathbb{N}$ *and* $S_0 \in \mathbb{N}$ *such that for any* $R > R_0$ *and* $s > S_0$ *and any initial parameter vector* $\theta_0 \in \Theta$

$$\left\| \widetilde{P}_R^{(s)}(\theta_0, \cdot) - p(\cdot|g, X) \right\|_{TV} \leq \epsilon \tag{28}$$

2. (WLLN) *A Weak Law of Large Numbers holds: for any initial parameter vector* $\theta_0 \in \Theta$ *and any bounded integrable function* $h(\cdot)$

$$\frac{1}{S} \sum_{s=1}^{S} h(\theta_s) \xrightarrow{P} \int_{\Theta} h(\theta) p(\theta|g, X) \, d\theta \tag{29}$$

Proof. In Appendix B. ■

The theorem states that the algorithm produces good samples as long as the number of steps of the network simulation algorithm is big enough and the algorithm is run for a sufficient number of iterations.

An important tuning parameter of the algorithm is R, the number of network simulations to be performed in the second step. As $R \to \infty$ the algorithm converges to the exact exchange algorithm of Murray et al. (2006), producing exact samples from the posterior distribution. At the same time an higher value of R would increase the computational cost and result in a higher rejection rate for the proposed parameters. I do not propose an *optimal* way to choose R, but I provide some evidence with simulated data in Appendix B, showing that there is not much difference in the estimates or convergence using different R's. The value of R has a stronger effect on the standard deviation than on the mean of the posterior, as one would expect.

3.3 Connections to Exponential Random Graphs

Considerable attention has been paid to exponential random graph models (ERGM).[33] These models are statistical models of random network formation, with complex dependence structures among links. These models have been successfully used to fit social networks, providing

[33] Frank and Strauss (1986) developed the theory of Markov random graphs. These are models of random network formation in which there is dependence among links: the probability that a links occur depends on the existence of other links. Wasserman and Pattison (1996) generalized the Markov random graphs to general dependence structures, developing the Exponential Random graph models. Snijders (2002) reviews these models and the related estimation techniques.

a useful benchmark for alternative models.

A remarkable feature of my model is that it contains ERGMs as a special case. Assume that the utility functions u, m and v depend *linearly* on a vector of parameters. Define $\theta_u = (\theta_{u1}, \theta_{u2}, ..., \theta_{uP})'$, $\theta_m = (\theta_{m1}, \theta_{m2}, ..., \theta_{mL})'$ and $\theta_v = (\theta_{v1}, \theta_{v2}, ..., \theta_{vS})'$. Define the function $H : \mathbb{R}^A \times \mathbb{R}^A \to \mathbb{R}$.

ASSUMPTION 4 ***(Linearity of Utility)*** *The utility functions are linear in parameters*

$$\begin{aligned} u_{ij} &= u(X_i, X_j, \theta_u) = \sum_{p=1}^{P} \theta_{up} H_{up}(X_i, X_j) = \theta_u' \mathbf{H}_u(X_i, X_j) \\ m_{ij} &= m(X_i, X_j, \theta_m) = \sum_{l=1}^{L} \theta_{ml} H_{ml}(X_i, X_j) = \theta_m' \mathbf{H}_m(X_i, X_j) \\ v_{ij} &= v(X_i, X_j, \theta_v) = \sum_{s=1}^{S} \theta_{vs} H_{vs}(X_i, X_j) = \theta_v' \mathbf{H}_v(X_i, X_j) \end{aligned}$$

This assumption leaves room for many interesting specifications. In particular, the functions H do not exclude interactions among different characteristics, for example interactions of race and gender of both individuals. We can consider different specifications and include different sets of variables for direct, mutual and indirect links.

PROPOSITION 4 ***(Exponential Family Likelihood)***
Under Assumptions 1-4, the stationary distribution $\pi(g, X)$ belongs to the ***exponential family****, i.e., it can be written in the form*

$$\pi(g, X) = \frac{\exp[\theta' \mathbf{t}(g, X)]}{\sum_{\omega \in \mathcal{G}} \exp[\theta' \mathbf{t}(\omega, X)]}, \tag{30}$$

where $\theta = (\theta_u, \theta_m, \theta_v)'$ is a (column) vector of parameters and $\mathbf{t}(g, X)$ is a (column) vector of canonical statistics.

Proof. See Appendix A ■

The vector $\mathbf{t}(g, X) = (t_1(g, X), ..., t_K(g, X))$ is a vector of sufficient statistics for the network formation model. This vector can contain the number of links, the number of whites-to-whites links, the number of male-to-female links and so on. Interactions between different variables are possible, e.g. the number of black-males-to-white-females links, or interactions of individual controls with school-level controls.

This likelihood is very similar to the one of exponential random graph models. My theoretical model can be interpreted as providing the microfoundations for exponential random graph models. In this sense, we can interpret the ERGM as the stationary equilibrium of a strategic game of network formation with myopic agents following a stochastic best response dynamics, when utility are linear functions of the parameters.

3.4 Identification and Practical Implementation

The identification of parameters for the linear utility case follows from the theory of exponential families (Lehman, 1983). Identification is guaranteed as long as the sufficient statistics $t(g, X)$ are not linearly dependent. The nonlinear case is more complex and there are no general conditions that guarantee identification.[34] For this reason, I consider estimation of the model only in the linear case.

The Bayesian framework can help to achieve identification of the parameters in the nonlinear case, by using prior distributions. This is familiar in the DSGE estimation literature, where parameters are often ill-identified and prior distributions are used to produce more precise estimates (as long as the prior is reasonable). This possibility is not explored here, and it is left to future research.

The linear case also allows for specifications of the utility function involving network-level controls, when estimation is performed using multiple networks. This can be achieved by specifying parameters

$$\theta_p = \theta_{p0} + \sum_{c=1}^{C} \theta_{pc} Z_c \tag{31}$$

where Z_c is a network-level variable. This specification allow network fixed effects and interactions of network controls with individual controls. The estimation methodology presented above can be applied to this specification without any change. However, estimation of a model with random coefficients would require significant additional computational effort.

As noted above, it is possible to modify the precision of the estimates when there is some previous information that can be incorporated in the prior. I choose somewhat vague priors for the parameters, in order to extract most of the information from the data. I assume independent normal priors

$$p(\theta) = \mathcal{N}(\mathbf{0}, 3\mathbf{I}_P), \tag{32}$$

[34] Geyer (1992) provides some guidance in this matter. He provides conditions that guarantee convergence of the Monte Carlo Maximum Likelihood estimate to the exact MLE. However, to the best of my knowledge, there are no sufficient conditions that guarantee identification in this setting.

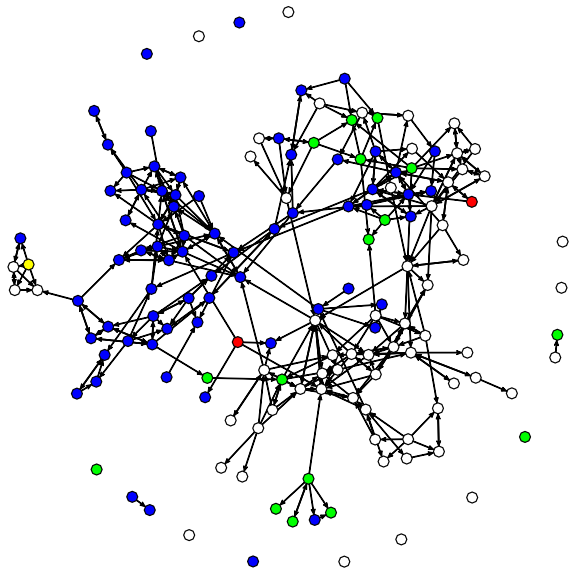
where P is the number of parameters.

The proposal distribution for the posterior simulation is

$$q_\theta(\cdot|\theta) = \mathcal{N}\left(\mathbf{0}, \delta\boldsymbol{\Sigma}\right), \tag{33}$$

where δ is a scaling factor and $\boldsymbol{\Sigma}$ is a covariance matrix. I use an adaptive procedure to

Figure 3: A School Network



white=Whites; blue = African Americans; yellow = Asians; green = Hispanics; red = Others

Note: The graphs represent the friendship network of a school extracted from AddHealth. Each dot represents a student, each arrow is a friend nomination. The colors represent racial groups.

determine a suitable $\boldsymbol{\Sigma}$. I start the iterations with $\boldsymbol{\Sigma} = \lambda\mathbf{I}_P$, where λ is a vector of standard deviations. I choose λ so that the sampler accepts at least 20%-25% of the proposed parameters, as is standard in the literature (Gelman et al., 2003; Robert and Casella, 2005). I run the chain and monitor convergence using standard methods. Once the chains have reached approximate convergence, I estimate the covariance matrix of the chains and use it as an approximate $\boldsymbol{\Sigma}$. The scaling factor is $\delta = 2.38^2/P$ as suggested in Gelman et al. (1996).

The network sampler uses a proposal $q_g(g|g')$, that selects a link to be updated at each period according to a discrete uniform distribution. The probability of network inversion is $p_{inv} = 0.01$.

The posterior distributions shown in the graphs are obtained with a simulation of 50000 Metropolis-Hastings updates of the parameters. These simulations start from values found after extensive experimentation with different starting values and burn-in periods, monitoring convergence using standard methods. For each parameter update, I simulate the network for 3000 iterations to collect a sample from the stationary distribution.

3.5 The Add Health Data

The *National Longitudinal Study of Adolescent Health* (Add Health) is a dataset containing information on a nationally representative sample of US schools. The survey started in 1994, when the 90118 participants were entering grades 7-12, and the project collected data in four successive waves.[35] Each student responded to an *in-school* questionnaire, and a subsample of 20745 was given an *in-home* interview to collect more detailed information about behaviors, characteristics and health status. In this paper I use only data from the *saturated sample* of Wave I, containing information on 16 schools. Each student in this sample completed both the in-school and in-home questionnaires. I exclude the two largest schools, 58 and 77: they have respectively 811 and 1664 students, while the third largest school has 159 students. To keep the sample as homogeneous as possible I prefer to not use these schools. My final sample includes 1139 students in 14 schools.

The *in-school* questionnaire collects the social network of each participant. Each student was given a school roster and was asked to identify up to five male and five female friends.[36] I use the friendship nominations as proxy for the social network in a school. The resulting network is *directed*: Paul may nominate Jim, but this does not imply that Jim nominates Paul.[37] The model developed in this paper takes this feature of the data into account.

A sub-sample of 20745 students was also given an *in-home* questionnaire, that collected most of the sensible data. I use data on racial group, grade and gender of individuals. A student with a missing value in any of these variables is dropped from the sample. Each students that declares to be of Hispanic origin is considered Hispanic. The remaining non-Hispanic

[35] More details about the sampling design and the representativeness are contained in Moody (2001) and the Add Health website http://www.cpc.unc.edu/projects/addhealth/projects/addhealth

[36] One can think that this limit could bias the friendship data, but only 3% of the students nominated 10 friends (Moody, 2001).

[37] Some authors do not take into account this feature of the data and they recode the friendships as mutual: if a student nominates another one, the opposite nomination is also assumed.

students are assigned to the racial group they declared. Therefore the racial categories are: White, Black, Asian, Hispanic and Other race. Other race contains Native Americans.

Descriptive statistics are in Table 1. The smallest school has 20 enrolled students while the largest used in estimation has 159 students. There is a certain amount of variation in the number of links: some schools are more social and form many links per capita, while other schools have very few friendship nominations. The ratio of boys to girls is balanced in almost all schools, except school 369, where female students are large majority.

Panel A summarizes the racial composition. Most schools are extremely racially homogeneous. School 1, 28, 126 and 175 are more diverse as reflected in the Racial Fragmentation index. This is an index that measure the degree of heterogeneity of a population. It is interpreted as the probability that two randomly chosen students in the school belong to different racial groups.[38] An index of 0 indicates that there is only one racial group and the population is perfectly homogeneous. Higher values of the index represents increasing levels of racial heterogeneity. Panel B summarizes the grade composition. Most schools offer all grades from 7th to 12th, with homogeneous population across grades. Several schools only have lower grades.

Panel C analyzes the racial and gender segregation of each school friendship network. The level of segregation is measured with the Freeman (1972) segregation index. If there is no segregation, the number of links among individuals of different groups does not depend on the group identity. The index measures the difference between the expected and actual number of links among individuals of different groups. An index of 0 means that the actual network closely resembles one in which links are formed at random. Higher values indicate more segregation. The index varies between 0 and 1, where the maximum corresponds to a network in which there are no cross-group links.

Since most schools are racially homogeneous, the measured segregation is zero. Schools with a racially diverse student population show high level of segregation for each racial group. On the other hand gender segregation is quite low and homogeneous across schools.

[38] If there are K racial groups and the share of each race is s_k, the index is

$$FRAG = 1 - \sum_{k=1}^{K} (s_k)^2 \tag{34}$$

Table 1: **Descriptive Statistics for the schools in the Saturated Sample**

School	1	2	3	7	8	28	58	77	81	88	106	115	126	175	194	369
Students	44	60	117	159	110	150	811	1664	98	90	81	20	53	52	43	52
Links	12	120	125	344	239	355	3290	3604	163	308	162	44	123	171	42	48
Females	0.5	0.517	0.419	0.44	0.5	0.587	0.473	0.483	0.531	0.522	0.531	0.55	0.491	0.538	0.512	0.654
A. Racial Composition																
Whites	0.5	0.95	0.983	0.981	0.973	0.42	0.978	0.055	0.98	0.989	0	1	0.472	0.769	0.977	0.942
Blacks	0.136	0	0	0.006	0.018	0.453	0.002	0.233	0	0	0.963	0	0.151	0.019	0	0
Asians	0	0	0	0	0.009	0.007	0.005	0.299	0.01	0	0	0	0.038	0.038	0	0
Hispanics	0.364	0.05	0.017	0.006	0	0.107	0.011	0.392	0.01	0	0.025	0	0.302	0.154	0.023	0.058
Others	0	0	0	0	0	0.013	0.004	0.02	0	0.011	0	0	0.038	0.019	0	0
Racial Fragm	0.599	0.095	0.034	0.037	0.053	0.606	0.044	0.699	0.04	0.022	0.072	0	0.661	0.382	0.045	0.109
B. Grade Composition																
7th Grade	0.159	0.2	0.128	0.145	0.227	0.173	0.002	0.001	0.112	0.144	0.506	0.4	0.491	0.462	0.488	0.538
8th Grade	0.159	0.217	0.154	0.157	0.2	0.173	0.004	0.003	0.153	0.178	0.481	0.6	0.472	0.538	0.488	0.462
9th Grade	0.114	0.2	0.12	0.214	0.136	0.2	0.289	0.004	0.153	0.122	0.012	0	0.038	0	0	0
10th Grade	0.273	0.133	0.205	0.157	0.182	0.167	0.277	0.346	0.214	0.167	0	0	0	0	0	0
11th Grade	0.136	0.167	0.179	0.164	0.118	0.14	0.223	0.345	0.265	0.211	0	0	0	0	0.023	0
12th Grade	0.159	0.083	0.214	0.164	0.136	0.147	0.205	0.301	0.102	0.178	0	0	0	0	0	0
C. Segregation																
Segr Whites	0	0	0	0	0	0.720	0.005	0.266	0	0	-	-	0.573	0.115	0	0
Segr Blacks	0	-	-	0	0	0.764	0	0.790	-	-	0	-	0.179	0	-	-
Segr Asian	-	-	-	-	0	0	0	0.744	0	-	-	-	0	0	-	-
Segr Hisp	0	0	0	0	-	0.429	0	0.691	-	-	0	-	0.227	0.025	0	0
Segr Other	-	-	-	-	-	0	0	0.026	-	0	-	-	0	0	-	-
Seg Gender	0.250	0.100	0.140	0.341	0.069	0.255	0.221	0.287	0.264	0.176	0.258	0.168	0.129	0.122	0.262	0.156

4 Empirical Results

4.1 Parameter Estimates

4.1.1 One school network

An important feature of the model is that it allows estimation using only one network observation. In this section, I estimate the model using data from school 28 of Add Health. The school has 150 enrolled students, 58.7% of whom are girls, with a total of 355 friend nominations. The clustering coefficient is 0.2906 and the racial fragmentation is 0.606. The racial composition is as follows: 42% whites, 45.3% blacks, 0.667% asians, 10.6% hispanics. Figure 3 shows the network of friendship nomination: each dot corresponds to a student, the color represents his racial group and an arrow is a friend nomination.

The results for three alternative specifications of the model are presented in Table 2. I report the posterior means and standard deviations. Each estimate measures the marginal effect of the variable: for example, the parameter associated with the direct utility of *white-white* measures the marginal utility of a white individual forming a link to another white, other things being equal.

The first column contains posterior means and standard deviations of a specification in which the direct utility is a function of total number of links (constant), total number of links in which both are Whites, Blacks or Hispanic. This specification tests for the presence of differential homophily: each racial group may have different homophily levels. A positive coefficient for the variable white-white would indicate that white students have a bias towards same race friends. The remaining controls are for the number of reciprocated links (mutual constant) and for the number of indirect friends (friends of friends).

These results point to strong racial homophily effects for each racial group. Each additional link is costly as indicated by the negative coefficient of the constant. However, an additional link is more valuable if the pair belongs to the same racial group: all the homophily coefficients are positive. A mutual link increases utility as expected, while linking to an individual with many friends decreases it. The latter effect can be due to congestion: individuals with many links have less time to devote to each of their friends.[39]

Model 2 includes controls for the racial composition of mutual friends and friends of friends. This model confirms the existence of homophily in direct links, but also in mutual

[39]At the same time one should notice that the homophily effect for Hispanics is estimated with higher variability: this is because there are very few Hispanics in the dataset, and they form few links. A partial solution is to run more simulations. Alternatively one could estimate a model with multiple schools and exploit the variability among schools as a source of identification.

Table 2: Three Specifications, School 28

	Model 1		Model 2		Model 3	
	mean	s.d.	mean	s.d.	mean	s.d.
Direct utility (u_{ij})						
constant	-4.6448	0.4555	-4.1779	0.5330	-4.5947	0.6502
same gender					0.2199	0.4942
same grade					0.7720	0.5558
white-white	1.3013	0.4812	0.4012	0.7681	0.4624	0.8419
black-black	1.4942	0.4463	0.7709	0.7670	0.7132	0.7985
hispanic-hispanic	0.7628	1.1791	0.8504	1.4012	1.5408	1.1437
Mutual utility (m_{ij})						
constant	3.5171	0.5036	2.8197	1.0779	0.9503	1.3547
same gender					1.5864	1.0896
same grade					0.0060	1.0120
white-white			0.4614	1.2300	0.3804	1.1925
black-black			0.7945	1.2114	0.7624	1.1534
hispanic-hispanic			-0.2865	1.9812	0.3745	1.7842
Indirect utility (v_{ij})						
constant	-0.0745	0.0596	-0.2629	0.1353	-0.3628	0.1849
same gender					-0.0152	0.1835
same grade					0.3559	0.1665
white-white			0.3249	0.1879	0.3354	0.2027
black-black			0.2426	0.1825	0.2761	0.1767
hispanic-hispanic			-0.0404	0.7695	-0.3136	0.9793

Posterior mean and standard deviation for three alternative specifications of the model. The estimates are obtained with a sample of 50000 simulations for the parameters, and 3000 network simulations for each parameter proposal.

and indirect links. The only exception is for links that involve hispanics: mutual and indirect links decrease utility.

Model 3 includes controls for homophily in gender and grade. In this dataset more than 50% of all friendships are within the same grade. At the same time, it is known that gender differences are an important explanatory variable of interaction, especially among adolescents. The estimates show that there are homophily effects for both grade and gender.

4.1.2 Multiple networks

The algorithm and the estimation methodology are easily extended to the case with multiple independent networks. In this section, I report results from an estimation performed using

data from all the 14 schools in my sample. In the first column of Table 3, I report the results

Table 3: Estimation results, full sample

	School 28		Full Sample		Full Sample	
	mean	s.d.	mean	s.d.	mean	s.d.
Direct utility (u_{ij})						
constant	-4.5947	0.6502	-5.0269	0.1701	-4.9742	0.1842
same gender	0.2199	0.4942	0.1475	0.1069	0.1644	0.1065
same grade	0.7720	0.5558	1.9400	0.1364	1.9745	0.1165
white-white	0.4624	0.8419	0.3268	0.1561	0.5575	0.2017
black-black	0.7132	0.7985	0.0039	0.2485	-0.2858	0.2101
hispanic-hispanic	1.5408	1.1437	0.5230	0.4267	0.6662	0.3216
white-white * whites					-0.4289	0.1316
black-black * blacks					2.0846	0.3656
hisp-hisp * hisp					-1.0826	0.8320
Mutual utility (m_{ij})						
constant	0.9503	1.3547	2.9716	0.3910	2.8194	0.3756
same gender	1.5864	1.0896	1.1868	0.2479	1.1686	0.2430
same grade	0.0060	1.0120	-1.6454	0.2791	-1.7988	0.2230
white-white	0.3804	1.1925	0.2342	0.3230	0.5027	0.3257
black-black	0.7624	1.1534	0.4118	0.4275	0.6010	0.3428
hispanic-hispanic	0.3745	1.7842	-0.4523	0.8312	-0.3575	0.2487
Indirect utility (v_{ij})						
constant	-0.3628	0.1849	0.0263	0.0388	0.0141	0.0424
same gender	-0.0152	0.1835	-0.1223	0.0481	-0.1335	0.0470
same grade	0.3559	0.1665	0.0839	0.0281	0.0890	0.0273
white-white	0.3354	0.2027	0.0290	0.0314	0.0433	0.0339
black-black	0.2761	0.1767	-0.0206	0.0459	0.0010	0.0434
hispanic-hispanic	-0.3136	0.9793	0.1104	0.1712	0.1424	0.1565

for school 28 as a useful comparison. Not surprisingly the standard deviation of the marginal posteriors are smaller when compared to the estimation with only one network. In Column 2 there is evidence of racial homophily in the direct links. Other things equal, a student prefers to form links to students of the same gender, grade and race. The racial homophily is not present for blacks.

Data from multiple schools allow the inclusion of school level variables that may help in identifying the homophily effects. The third column presents results where the homophily effects are interacted with the proportions of each racial group in the school. As the white student body increases, White students receive lower utility from same race friends. Conversely, when the proportion of blacks in the school increases, African American students

value friends of the same racial group more. Hispanic preferences mirror those of whites.

It is important to highlight that the estimated marginal utilities for direct links are obtained controlling for the structure of the network. The homophily effects are therefore net of the network structure. Homophily effects are present in the mutual and indirect links. Interpreting these estimates is not as simple as with the direct utility. Therefore, I present several examples in Figure 4. In Panel A a network with 8 students is shown. The students are assumed to be all whites, male and enrolled in the same grade. Student 4 has to choose whether to form a new link to agent 5. To simplify the exposition, suppose that the utility is evaluated at the posterior mean. The probability that Agent 4 forms the link is 0.067, according to the estimate in column 3 of Table 3. Considering only the direct utility, this probability would be a little lower, 0.062. In Panel B agent 5 is now African American. If we were to consider only the direct effect of this change, the probability of the link would drop by 11.4%. When we consider the effect of the network structure (effect on the popularity and friends of friends), this change implies a decline in the probability of that link of 21.5%. The remaining graphs are variations of this simple example and all the percentage changes are measured with respect to the baseline network in Panel A. The most intriguing result is in Panels D and E. In Panel D, agent 5 is female. When considering only the direct effect, this would imply a decrease in utility and therefore in the probability of linking. However, the indirect and popularity effects counterbalance the decrease in direct utility, implying an increase in the linking probability. A similar mechanism appears in Panel E.

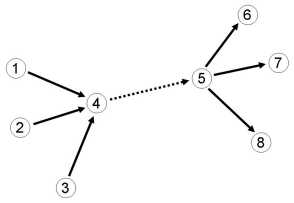
4.2 Policy Experiments

The estimated model can be used to predict how alternative policies affect network structure. Policy makers may be interested in pursuing policies that promote racial integration, or they may consider policies that create separate schools for boys and girls. My model can provide guidance.

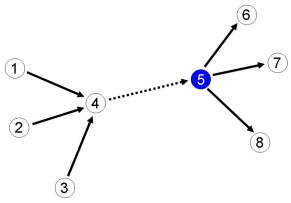
Consider evaluating the effectiveness of busing programs in promoting interracial integration. School 28 has an extremely segregated friendship network: if the school administration starts a busing program that modifies the composition of the school, does segregation increase or decrease?

Using the posterior distribution estimated in column 3 of Table 2, I simulate two policies. The first policy increases the African American enrollment by transferring 8 African-American students from a random school to school 28. The second reassigns 16 Hispanic students from the same random school to school 28. In both cases, I compute the segregation levels in the stationary equilibrium before and after the implementation of the policy. I

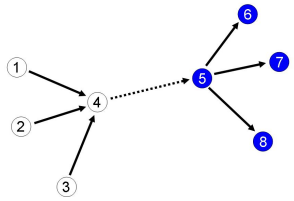
Figure 4: Change in the probability of forming a link



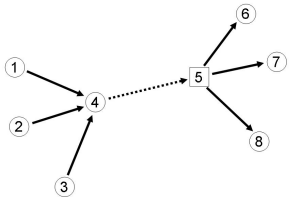
A. Baseline



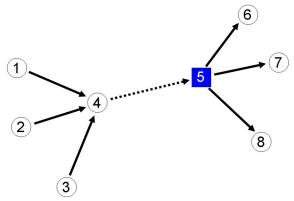
B. Agent 5 is black
Direct effect: −11.4%
Total effect: −21.5%



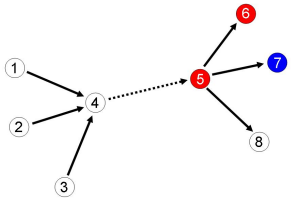
C. Agents 5, 6, 7 and 8 are black
Direct effect: −11.4%
Total Effect: −30.7%



D. Agent 5 is female
Direct effect: −14.3%
Total effect: 24.5%



E. Agent 5 is black female
Direct effect: −24.2%
Total Effect: −2.1%



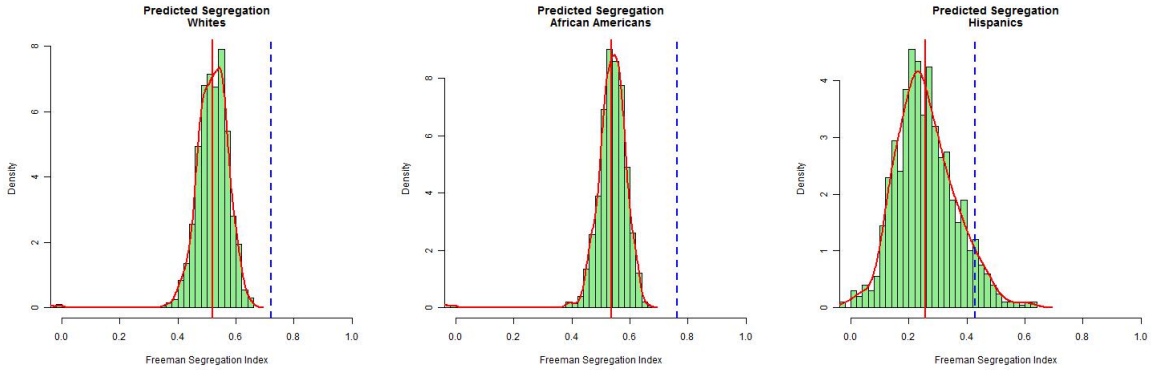
F. Agent 5 has diverse friends
Direct effect: −11.4%
Total effect: −27.7%

The network contains $n = 8$ agents. In each panel agent 4 is deciding whether to create a link to agent 5. Panel A is the baseline situation, where all the students are white. For simplicity assume they are all males enrolled in the same grade. The remaining panels show the change in the probability that the link is formed, when the structure of the network is altered. The direct effect is the change in probability (with respect to Panel A) arising only because of the change in the direct utility. The total effect is the change in the probability of linking when considering all the components of the utility function. In Panel B, agent 5 is black: if we consider only the effect on the direct utility the probability of a link among 4 and 5 goes down by 11%. When we consider the full utility of agent 4, the probability of the link decrease by almost twice as much. Similar results hold for the remaining panels.

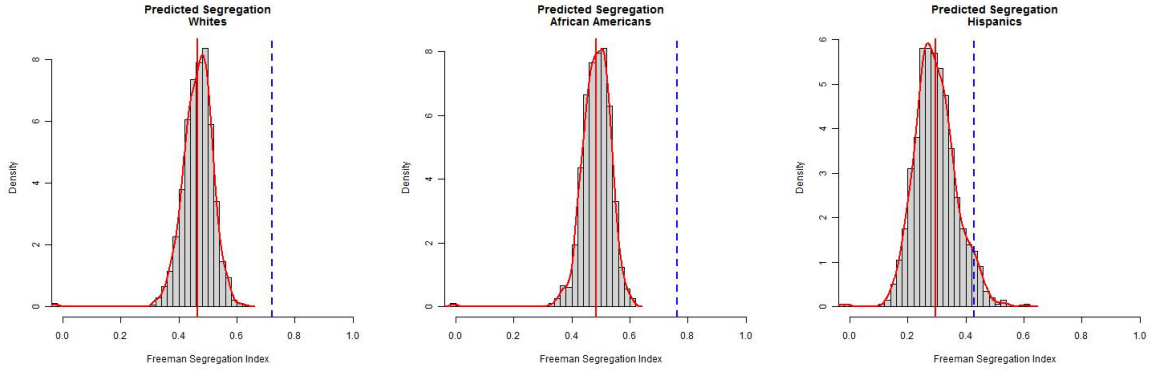
use Freeman's segregation index (see Freeman (1972)) to measure segregation for the three relevant groups: Whites, African-Americans and Hispanics.

The results are reported in Figure 5. Panel A shows the segregation level without policy

Figure 5: Policy Experiments



Panel A. Busing program transporting 8 African American students to School 28.



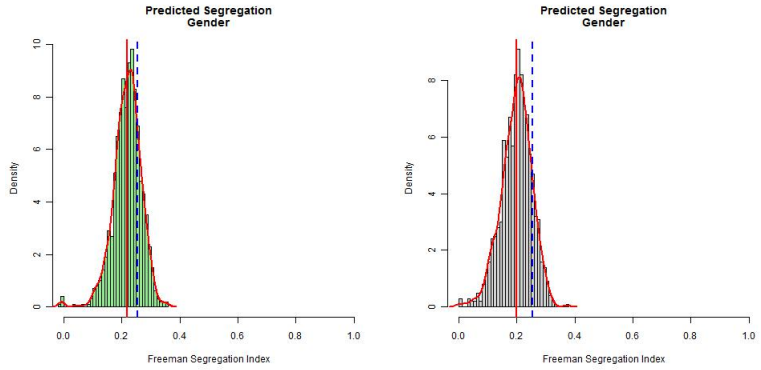
Panel B. Busing program transporting 16 Hispanic students to School 28.

The graphs show the distribution and average of Freeman's Segregation Index for the 3 racial groups after the policy is implemented (red solid) and the segregation before the policy (blue dashed). The graphs also show the histogram of the simulated segregation and a kernel smoothed density. The graphs in Panel A row shows a reassignment of 8 African-American students to school 28. The graphs in Panel B refer to a policy that reassigns 16 Hispanic students to school 28.

(blue) and the distribution after the implementation of the policy (red) when we reassign 8 African-Americans to school 28. For all the racial groups the expected segregation goes down. The probability of an increase in racial segregation is null for Whites and African Americans and it is minimal for Hispanics (0.06). Panel B shows that the second policy has similar results. Figure 6 analyzes the effects of the policies on gender segregation. The policy successfully reduces both racial and gender expected segregation. The probability of an increase in gender segregation is 0.213 and 0.131 for the two policies respectively.

These examples might suggest that policies that modify the racial composition within

Figure 6: Policy Experiments



Panel A. Policy 1 Panel B. Policy 2

The graphs show the distribution and average of Freeman's Segregation Index for the gender segregation after the policy is implemented (red solid) and the segregation before the policy (blue dashed). The graphs also show the histogram of the simulated segregation and a kernel smoothed density. The graphs in Panel A row shows a reassignment of 8 African-American students to school 28. The graphs in Panel B refer to a policy that reassigns 16 Hispanic students to school 28.

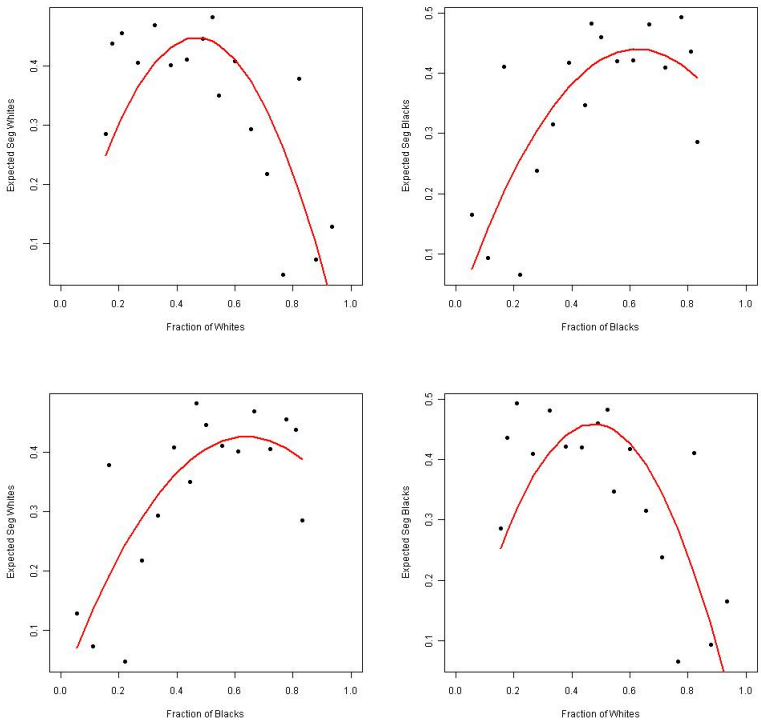
schools reduce segregation in the social network of friendship. However, this is not always the case.

I simulate several swaps of students among school 88 and 106. These are two schools with an homogeneous student population: 98.9% whites and 96.3% African American respectively. The simulated policies take several (white) students from school 88 and enroll them in school 106, while the same number of (black) students in school 106 are enrolled in school 88. This allows me to modify the ratio of Whites and African Americans in the two schools and predict the levels of segregation.

The results of these simulations are reported in Figure 7. The relationship between proportion of a racial group and the expected segregation levels has an inverted-U shape. The graph suggests that the implementation of a policy that modifies the fraction of whites from .9 to .8 will increase segregation on average by .2. The main lesson from this graph is that equalizing the racial shares between the two schools is a bad idea if integration is one of the policymaker's goals. An alternative concern for busing programs is that a recent decision of the Supreme Court[40] declared unconstitutional the use of race to determine children assignment to schools. Therefore, school district administrators who want to promote racial

[40]Parents Involved in Community Schools vs Seattle School District No. 1, 551 U.S. 701 (2007), http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=us&vol=000&invol=05-908.

Figure 7: Policy Experiments, School 88



The graphs shows the results of policy experiments in which students are swapped between school 88 and school 106. The expected segregation in the stationary equilibrium after the policy is plotted against the fraction of each racial group. Each dot represents a different simulated policy. The red solid line is the fitted value of a regression where the expected segregation is a function of fraction of the racial group and fraction of the racial group squared.

integration have to find alternative ways to assign students to schools. For example, one may be tempted to create single-gender schools. Table 4 presents the results from such a policy

Table 4: Same gender schools, school 28

	Current	Female	Male
White	0.7202	0.2768	0.3507
African Americans	0.7636	0.2791	0.3752
Hispanic	0.4288	0.0970	0.2221

using school 28. I create two schools, one with only male students and one with only female students. The results are clear: the expected racial segregation decreases in both schools.

This could provide an alternative to busing programs based on race.

5 Conclusions

This paper develops and estimates a dynamic model of strategic network formation with heterogeneous agents. The paper contributes to the economic literature on network formation in two ways. First, while most strategic models have multiple equilibria, I establish the existence of a unique stationary equilibrium, which characterizes the likelihood of observing a specific network structure in the data. As a consequence, I can estimate and identify the structural parameters using only one observation of the network at a single point in time. Second, I propose a Bayesian Markov Chain Monte Carlo algorithm that drastically reduces the computational burden for estimating the posterior distribution. In this model, the likelihood function cannot be evaluated or approximated with precision: a state-of-the-art supercomputer would take several years to evaluate the likelihood once. To overcome this problem, I propose an algorithm that generates samples from the posterior distribution and avoids the evaluation of the likelihood. Using the properties of the stationary equilibrium, I reduce the computational burden even further and I am able to study high dimensional models.

The model can be used to infer the effect of different policies on network structure. To illustrate this point, I explore different desegregation policies in US schools. The model provides predictions about the expected levels of segregation implied by busing programs: there is an inverted U-shape relationship between the share of a racial group in the school and the expected segregation level. These results suggest that these policies must be carefully designed to avoid unexpected outcomes. My model can be used to guide the design of such programs.

The model is easily extended to incorporate unobserved heterogeneity in individual quality and preferences. Additionally, the Bayesian estimation strategy can be used to estimate models with missing links. These improvements come with an increase in the computational costs, but also provide a more realistic model. Both these extensions are reviewed and explained in Appendix C.

My methodology can be used in different settings. Models of social interactions with sequential moves as in Nakajima (2007) share the same simple equilibrium characterization presented in this work. In these models individuals interact in an exogenous network and their actions are optimally chosen given the action of their neighbors. The estimation techniques developed here are easily adapted to these settings.

The methodology can also be applied to the class of autologistic models in spatial econometrics.[41] These are models for spatial binary data that account for the spatial dependence among variables. The likelihood of these models has the exponential form with normalizing constant but their estimation has relied on approximate methods: Maximum Pseudolikelihood (Besag, 1974) or Markov Chain Monte Carlo Maximum Likelihood (Geyer and Thompson, 1992). My estimation strategy provides a valid alternative from a Bayesian point of view.[42]

References

Acemoglu, D., M. Dahleh, I. Lobel and A. Ozdaglar (2011), 'Bayesian learning in social networks', *Review of Economic Studies* **forthcoming**.

Amemiya, Takeshi (1981), 'Qualitative response models: A survey', *Journal of Economic Literature* **19**(4), 1483–1536.

Andrieu, C. and G. O. Roberts (2009), 'The pseudo-marginal approach for efficient monte carlo computations', *Annals of Statistics* **37**(2), 697–725.

Bala, Venkatesh and Sanjeev Goyal (2000), 'A noncooperative model of network formation', *Econometrica* **68**(5), 1181–1229.

Bandiera, Oriana and Imran Rasul (2006), 'Social networks and technology adoption in northern mozambique', *Economic Journal* **116**(514), 869–902.

Besag, Julian (1974), 'Spatial interaction and the statistical analysis od lattice systems', *Journal of the Royal Statistical Society Series B (Methodological)* **36**(2), 192–236.

Bhamidi, Shankar, Guy Bresler and Allan Sly (forthcoming), 'Mixing time of exponential random graphs', *Annals of Applied Probability* .

Blume, Lawrence E. (1993), 'The statistical mechanics of strategic interaction', *Games and Economic Behavior* **5**(3), 387–424.

Caimo, Alberto and Nial Friel (2010), 'Bayesian inference for exponential random graph models', *Social Networks* **forthcoming**.

[41] Besag (1974) provides a description of these models and a simple approximate estimation strategy.

[42] In principle, any model with a potential that admits an exponential likelihood with normalizing constant can be estimated using my method.

Chib, Siddartha and Srikanth Ramamurthy (2009), 'Tailored random block', *Journal of Econometrics* .

Christakis, Nicholas, James Fowler, Guido W. Imbens and Karthik Kalyanaraman (2010), An empirical model for strategic network formation. Harvard University.

Comola, Margherita (2008), The network structure of informal arrangements: Evidence from rural tanzania, PSE Working Papers 2008-74, PSE (Ecole normale suprieure).

Conley, Timothy and Christopher Udry (forthcoming), 'Learning about a new technology: Pineapple in ghana', *American Economic Review* .

Cooley, Jane (2010), Desegregation and the achievement gap: Do diverse peers help? working paper.

Currarini, Sergio, Matthew O. Jackson and Paolo Pin (2009), 'An economic model of friendship: Homophily, minorities, and segregation', *Econometrica* **77**(4), 1003–1045.

Currarini, Sergio, Matthew O. Jackson and Paolo Pin (2010), 'Identifying the roles of race-based choice and chance in high school friendship network formation', *the Proceedings of the National Academy of Sciences* **107**(11), 48574861.

De Giorgi, Giacomo, Michele Pellizzari and Silvia Redaelli (2010), 'Identification of social interactions through partially overlapping peer groups', *American Economic Journal: Applied Economics* **2**(2).

De Marti, Joan and Yves Zenou (2009), Ethnic identity and social distance in friendship formation, Cepr discussion papers, C.E.P.R. Discussion Papers.

Diaconis, Persi and D. Stroock (1991), 'Geometric bounds for eigenvalues of markov chains', *Annals of Applied Probability* **1**(1), 36–61.

Diaconis, Persi and Sourav Chatterjee (2011), Estimating and understanding exponential random graph models.

Echenique, Federico, Roland G. Fryer and Alex Kaufman (2006), 'Is school segregation good or bad?', *American Economic Review* **96**(2), 265–269.

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

Freeman, L. (1972), 'Segregation in social networks', *Sociological Methods and Research* **6**, 411–427.

Galeotti, Andrea (2006), 'One-way flow networks: the role of heterogeneity', *Economic Theory* **29**(1), 163–179.

Gelman, A., G. O. Roberts and W. R. Gilks (1996), 'Efficient metropolis jumping rules', *Bayesian Statistics* **5**, 599–608.

Gelman, A., J. Carlin, H. Stern and D. Rubin (2003), *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC.

Geyer, Charles and Elizabeth Thompson (1992), 'Constrained monte carlo maximum likelihood for depedendent data', *Journal of the Royal Statistical Society, Series B (Methodological)* **54**(3), 657–699.

Geyer, Charles J. (1992), 'Practical markov chain monte carlo', *Statistical Science* **7**, 473–511.

Gilles, Robert P. and Sudipta Sarangi (2004), Social network formation with consent, Discussion paper, Tilburg University, Center for Economic Research.

Golub, Benjamin and Matthew Jackson (2011), 'Network structure and the speed of learning: Measuring homophily based on its consequences', *Annals of Economics and Statistics* .

Heckman, James J. (1978), 'Dummy endogenous variables in a simultaneous equation system', *Econometrica* **46**(4), 931–959.

Jackson, Matthew and Allison Watts (2002), 'The evolution of social and economic networks', *Journal of Economic Theory* **106**(2), 265–295.

Jackson, Matthew and Asher Wolinsky (1996), 'A strategic model of social and economic networks', *Journal of Economic Theory* **71**(1), 44–74.

Jackson, Matthew O. (2008), *Social and Economics Networks*, Princeton.

Jackson, Matthew O. and Brian W. Rogers (2007), 'Meeting strangers and friends of friends: How random are social networks?', *American Economic Review* **97**(3), 890–915.

Koskinen, Johan H. (2008), The linked importance sampler auxiliary variable metropolis hastings algorithm for distributions with intractable normalising constants. MelNet Social Networks Laboratory Technical Report 08-01, Department of Psychology, School of Behavioural Science, University of Melbourne, Australia.

Laschever, Ron (2009), The doughboys network: Social interactions and labor market outcomes of world war i veterans. working paper.

Lehman, E. L. (1983), *Theory of Point Estimation*, Wiley and Sons.

Levin, David. A, Yuval Peres and Elizabeth L. Wilmer (2008), *Markov Chains and Mixing Times*, American Mathematical Society.

Liang, Faming (2010), 'A double metropolis-hastings sampler for spatial models with intractable normalizing constants', *Journal of Statistical Computing and Simulation* **forthcoming**.

Liang, Faming, Chuanhai Liu and Raymond J. Carroll (2010), *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*, John Wiley and Sons, Ltd.

Marjoram, Paul, John Molitor, Vincent Plagnol and Simon Tavar (2003), 'Markov chain Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences of the United States of America* **100**(26), 15324–15328.

Mayer, Adalbert and Steven L. Puller (2008), 'The old boy (and girl) network: Social network formation on university campuses.', *Journal of Public Economics* **92**(1-2), 329–347.

Mele, Angelo (2010), Segregation in social networks: Monte carlo maximum likelihood estimation. Working Paper.

Meyn, Sean and Richard L. Tweedie (2009), *Markov Chains and Stochastic Stability*, Cambridge University Press.

Monderer, Dov and Lloyd Shapley (1996), 'Potential games', *Games and Economic Behavior* **14**, 124–143.

Moody, James (2001), 'Race, school integration, and friendship segregation in america', *American Journal of Sociology* **103**(7), 679–716.

Murray, Iain A., Zoubin Ghahramani and David J. C. MacKay (2006), 'Mcmc for doubly-intractable distributions', *Uncertainty in Artificial Intelligence* .

Nakajima, Ryo (2007), 'Measuring peer effects on youth smoking behavior', *Review of Economic Studies* **74**(3), 897–935.

Robert, Christian P. and George Casella (2005), *Monte Carlo Statistical Methods (Springer Texts in Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Snijders, Tom A.B (2002), 'Markov chain monte carlo estimation of exponential random graph models', *Journal of Social Structure* **3**(2).

Tamer, Elie (2003), 'Incomplete simultaneous discrete response model with multiple equilibria', *The Review of Economic Studies* **70**(1), 147–165.

Topa, Giorgio (2001), 'Social interactions, local spillovers and unemployment', *Review of Economic Studies* **68**(2), 261–295.

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.

Wasserman, Stanley and Philippa Pattison (1996), 'Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*', *Psychometrika* **61**(3), 401–425.

A Proofs

Proof of Proposition 1
The potential is a function Q from the space of actions to the real line such that $Q\left(g_{ij}, g_{-ij}, X\right) - Q\left(g'_{ij}, g_{-ij}, X\right) = U_i\left(g_{ij}, g_{-ij}, X\right) - U_i\left(g'_{ij}, g_{-ij}, X\right)$, for any ij.[43] A simple computation shows that, for any ij

$$\begin{aligned} Q\left(g_{ij}=1, g_{-ij}, X\right) - Q\left(g_{ij}=0, g_{-ij}, X\right) &= u_{ij} + g_{ji} m_{ij} + \sum_{\substack{k=1 \\ k\neq i,j}}^{n} g_{jk} v_{ik} + \sum_{\substack{k=1 \\ k\neq i,j}}^{n} g_{ki} v_{kj} \\ &= U_i\left(g_{ij}=1, g_{-ij}, X\right) - U_i\left(g_{ij}=0, g_{-ij}, X\right) \end{aligned}$$

[43] For more details and definitions see Monderer and Shapley (1996).

therefore Q is the potential of the network formation game. The welfare function is computed as

$$\begin{aligned}
W(g,X) &= \sum_{i=1}^{n} U_i(g,X) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}u_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}g_{ji}m_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} g_{ij}g_{jk}v_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} g_{ij}g_{ki}v_{kj} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}u_{ij} + 2\sum_{i=1}^{n}\sum_{j>i}^{n} g_{ij}g_{ji}m_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} g_{ij}g_{jk}v_{ik} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} g_{ij}g_{ki}v_{kj} \\
&= Q(g,X) + \sum_{i=1}^{n}\sum_{j>i}^{n} g_{ij}g_{ji}m_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n} g_{ij}g_{ki}v_{kj}
\end{aligned}$$

Proof of Proposition 2
1) The existence of Nash equilibria follows directly from the fact that the network formation game is a potential game with finite strategy space. (see Monderer and Shapley (1996) for details)
2) The set of Nash equilibria is defined as the set of g^* such that, for every i and for every $g_{ij} \neq g^*_{ij}$

$$U_i\left(g^*_{ij}, g^*_{-ij}, X\right) \geq U_i\left(g_{ij}, g^*_{-ij}, X\right)$$

Therefore, since Q is a potential function, for every $g_{ij} \neq g^*_{ij}$

$$Q\left(g^*_{ij}, g^*_{-ij}, X\right) \geq Q\left(g_{ij}, g^*_{-ij}, X\right)$$

Therefore g^* is a maximizer of Q. The converse is easily checked by the same reasoning.
3) Suppose $g^t = g^*$. Since this is a Nash equilibrium, no player will be willing to change her linking decision when her turn to play comes. Therefore, once the chain reaches a Nash equilibrium, it cannot escape from that state.
4) The probability that the potential will increase from t to $t+1$ is

$$Pr\left[Q\left(g^{t+1}, X\right) \geq Q\left(g^t, X\right)\right] =$$

$$\begin{aligned}
&= \sum_i \sum_j Pr\left(m^{t+1} = ij\right) \underbrace{Pr\left[U_i\left(g^{t+1}_{ij}, g^t_{-ij}, X\right) \geq U_i\left(g^t_{ij}, g^t_{-ij}, X\right) \middle| m^{t+1} = ij\right]}_{=1 \text{ because agents play Best Response, conditioning on } m^{t+1}} \\
&= \sum_i \sum_j \rho_{ij} = 1.
\end{aligned}$$

By part 3) of the proposition, a Nash network is an absorbing state of the chain. Therefore any probability distribution that puts probability 1 on a Nash network is a stationary distribution. For any initial network, the chain will converge to one of the stationary distributions. It follows that in the long run the model will be in a Nash network, i.e. for any $g^0 \in \mathcal{G}$

$$\lim_{t\to\infty} Pr\left[g^t \in NE \middle| g^0\right] = 1.$$

Proof of Theorem 1

1. The sequence of networks $\left[g^0, g^1, ...\right]$ generated by the network formation game is a markov chain. Inspection of the transition probability proves that the chain is irreducible and aperiodic, therefore it is ergodic. The existence of a unique stationary distribution then follows from the ergodic theorem (see Gelman et al. (1996) for details).

2. A sufficient condition for stationarity is the *detailed balance* condition. In our case this requires

$$P_{gg'}\pi_g = P_{g'g}\pi_{g'} \tag{35}$$

where

$$\begin{aligned} P_{gg'} &= \Pr\left(g^{t+1} = g' \middle| g^t = g\right) \\ \pi_g &= \pi\left(g^t = g\right) \end{aligned}$$

Notice that the transition from g to g' is possible if these networks differ by only one element g_{ij}. Otherwise the transition probability is zero and the detailed balance condition is satisfied. Let's consider the nonzero probability transitions, with $g = (1, g_{-ij})$ and $g' = (0, g_{-ij})$. Define $\Delta Q \equiv Q(1, g_{-ij}, X) - Q(0, g_{-ij}, X)$.

$$\begin{aligned} P_{gg'}\pi_g &= \Pr\left(m^t = ij\right) Pr\left(g_{ij} = 0 \middle| g_{-ij}\right) \frac{\exp\left[Q\left(1, g_{-ij}, X\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X\right)\right]} \\ &= \rho\left(g_{-ij}, X_i, X_j\right) \times \frac{1}{1 + \exp\left[\Delta Q\right]} \times \frac{\exp\left[Q\left(1, g_{-ij}, X\right) + Q\left(0, g_{-ij}, X\right) - Q\left(0, g_{-ij}, X\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X\right)\right]} \\ &= \rho\left(g_{-ij}, X_i, X_j\right) \times \frac{1}{1 + \exp\left[\Delta Q\right]} \times \frac{\exp\left[Q\left(1, g_{-ij}, X\right) - Q\left(0, g_{-ij}, X\right)\right] \exp\left[Q\left(0, g_{-ij}, X\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X\right)\right]} \\ &= \rho\left(g_{-ij}, X_i, X_j\right) \frac{\exp\left[\Delta Q\right]}{1 + \exp\left[\Delta Q\right]} \frac{\exp\left[Q\left(0, g_{-ij}, X\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X\right)\right]} \\ &= \Pr\left(m^t = ij\right) \Pr\left(g_{ij} = 1 \middle| g_{-ij}\right) \frac{\exp\left[Q\left(0, g_{-ij}, X\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X\right)\right]} \\ &= P_{g'g}\pi_{g'} \end{aligned}$$

So the distribution (14) satisfies the detailed balance condition. Therefore it is a stationary distribution for the network formation model. From part 1) of the proposition, we know that the process is ergodic and it has a unique stationary distribution. Therefore $\pi(g, X)$ is also the unique stationary distribution.

Proof of Proposition 4

The proof consists of showing that $Q(g, X)$ can be written in the form $\theta' \mathbf{t}(g, X)$. Consider the first part of the potential

$$\begin{aligned} \sum_i \sum_j g_{ij} u_{ij} &= \sum_i \sum_j g_{ij} \sum_{p=1}^{P} \theta_{up} H_{up}\left(X_i, X_j\right) \\ &= \sum_{p=1}^{P} \theta_{up} \sum_i \sum_j g_{ij} H_{up}\left(X_i, X_j\right) \\ &\equiv \sum_{p=1}^{P} \theta_{up} t_{up}\left(g, X\right) \\ &= \theta_u' \mathbf{t}_u\left(g, X\right) \end{aligned}$$

where $t_{up}(g,X) \equiv \sum_i \sum_j g_{ij} H_{up}(X_i, X_j)$, $\theta_u = (\theta_{u1}, \theta_{u2}, ..., \theta_{uP})'$ and $\mathbf{t}_u(g,X) = (t_{u1}(g,X), t_{u2}(g,X), ..., t_{uP}(g,X))'$. Analogously define $\theta_m = (\theta_{m1}, \theta_{m2}, ..., \theta_{mL})'$ and $\mathbf{t}_m(g,X) = (t_{m1}(g,X), t_{m2}(g,X), ..., t_{mL}(g,X))'$ and $\theta_v = (\theta_{v1}, \theta_{v2}, ..., \theta_{vS})'$ and $\mathbf{t}_v(g,X) = (t_{v1}(g,X), t_{v2}(g,X), ..., t_{vS}(g,X))'$. It follows that

$$\begin{aligned}
\sum_i \sum_{j>i} g_{ij} g_{ji} m_{ij} &= \sum_i \sum_{j>i} g_{ij} g_{ji} \sum_{l=1}^{L} \theta_{ml} H_{ml}(X_i, X_j) \\
&= \sum_{l=1}^{L} \theta_{ml} \sum_i \sum_{j>i} g_{ij} g_{ji} H_{ml}(X_i, X_j) \\
&= \sum_{l=1}^{L} \theta_{ml} t_{ml}(g,X) \\
&= \theta_m' \mathbf{t}_m(g,X)
\end{aligned}$$

and

$$\begin{aligned}
\sum_i \sum_j g_{ij} \sum_{k \neq i,j} g_{jk} v_{ij} &= \sum_i \sum_j g_{ij} \sum_{k \neq i,j} g_{jk} \sum_{s=1}^{S} \theta_{vs} H_{vs}(X_i, X_k) \\
&= \sum_{s=1}^{S} \theta_{vs} \sum_i \sum_j g_{ij} \sum_{k \neq i,j} g_{jk} H_{vs}(X_i, X_k) \\
&= \sum_{s=1}^{S} \theta_{vs} t_{vs}(g,X) \\
&= \theta_v' \mathbf{t}_v(g,X)
\end{aligned}$$

Therefore $Q(g,X)$ can be written in the form $\theta' \mathbf{t}(g,X)$, where $\theta = (\theta_u, \theta_m, \theta_v)'$ and $\mathbf{t}(g,X) = [\mathbf{t}_u(g,X), \mathbf{t}_m(g,X), \mathbf{t}_v(g,X)]'$

$$\begin{aligned}
Q(g,X) &= \theta_u' \mathbf{t}_u(g,X) + \theta_m' \mathbf{t}_m(g,X) + \theta_v' \mathbf{t}_v(g,X) \\
&= \theta' \mathbf{t}(g,X)
\end{aligned}$$

and the stationary distribution is

$$\pi(g,X) = \frac{\exp[\theta' \mathbf{t}(g,X)]}{\sum_{\omega \in \mathcal{G}} \exp[\theta' \mathbf{t}(\omega,X)]}.$$

B Computational Details

B.1 Network Simulation

The algorithm used to simulate the network produces samples from the stationary equilibrium of the model. This is the result of Proposition 3.

Proof of Proposition 3
Proof.

1. The network simulation algorithm satisfies the detailed balance condition for the stationary distribu-

tion 14. Indeed for any given θ

$$\begin{aligned}
\Pr\left(g'|g,X,\theta\right)\pi\left(g,X,\theta\right) &= q_g\left(g'|g\right)\min\left\{1,\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{\exp\left[Q\left(g,X,\theta\right)\right]}\frac{q_g\left(g|g'\right)}{q_g\left(g'|g\right)}\right\}\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)} \\
&= \min\left\{q_g\left(g'|g\right)\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)},\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)}q_g\left(g|g'\right)\right\} \\
&= q_g\left(g|g'\right)\min\left\{\frac{q_g\left(g'|g\right)}{q_g\left(g|g'\right)}\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)},\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)}\right\} \\
&= q_g\left(g|g'\right)\min\left\{\frac{q_g\left(g'|g\right)}{q_g\left(g|g'\right)}\frac{\exp\left[Q\left(g,X,\theta\right)\right]}{\exp\left[Q\left(g',X,\theta\right)\right]},1\right\}\frac{\exp\left[Q\left(g',X,\theta\right)\right]}{c\left(\mathcal{G},X,\theta\right)} \\
&= \Pr\left(g|g',X,\theta\right)\pi\left(g',X,\theta\right)
\end{aligned}$$

This concludes the proof.

2. The algorithm generates a Markov Chain of network with finite state space. The chain is irreducible and aperiodic and therefore it is uniformly ergodic (see Theorem 4.9, page 52 in Levin et al. (2008)).

3. The bound to the convergence rate was derived by Diaconis and Stroock (1991), for reversible finite chains (see Proposition 3 in their paper).

■

The algorithm has a very useful property that can be exploited in the posterior simulation to reduce the computational burden. Adapting the suggestion in Liang (2010), define $\mathcal{P}_{\theta'}^{(R)}(g'|g)$ as the transition probability of a Markov chain that generates g' with R Metropolis-Hastings updates of the network simulation algorithm, starting at the observed network g and using the proposed parameter θ'. Then,

$$\mathcal{P}_{\theta'}^{(R)}(g'|g) = \mathcal{P}_{\theta'}(g^1|g)\mathcal{P}_{\theta'}(g^2|g^1)\cdots\mathcal{P}_{\theta'}(g'|g^{R-1}), \tag{36}$$

where $\mathcal{P}_{\theta'}(g^j|g^i) = q_g(g^j|g^i)\alpha_{mh}(g^i,g^j)$ is the transition probability of the network simulation algorithm above. Since the Metropolis-Hastings algorithm satisfies the detailed balance condition, we can prove the following

LEMMA 1 *Simulate a network g' from the stationary distribution $\pi\left(\cdot,X,\theta'\right)$ using a Metropolis-Hastings algorithm starting at the network g observed in the data. Then*

$$\frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)} = \frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]} \tag{37}$$

for all R, $g,g'\in\mathcal{G}$ and for any $\theta'\in\Theta$.

Proof. Let $\mathcal{P}_{\theta'}^{(R)}(g'|g)$ be defined as in (36). This is the transition probability of the chain that generates g' with R Metropolis-Hastings updates, starting at the observed network g and using the proposed parameter θ'. Notice that the Metropolis-Hastings algorithm satisfies the detailed balance for $\pi\left(g,X,\theta'\right)$, therefore we have

$$\begin{aligned}
\mathcal{P}_{\theta'}^{(R)}(g|g')\pi\left(g',X,\theta'\right) &= \mathcal{P}_{\theta'}(g_{R-1}|g')\mathcal{P}_{\theta'}(g_{R-2}|g_{R-1})\cdots\mathcal{P}_{\theta'}(g|g_1)\pi\left(g',X,\theta'\right) \\
&= \mathcal{P}_{\theta'}(g_1|g)\mathcal{P}_{\theta'}(g_2|g_1)\cdots\mathcal{P}_{\theta'}(g'|g_{R-1})\pi\left(g,X,\theta'\right) \\
&= \mathcal{P}_{\theta'}^{(R)}(g'|g)\pi\left(g,X,\theta'\right)
\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)} &= \frac{\pi\left(g,X,\theta'\right)}{\pi\left(g',X,\theta'\right)} \\ &= \frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]}\frac{c\left(\mathcal{G},X,\theta'\right)}{c\left(\mathcal{G},X,\theta'\right)} \\ &= \frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]}.\end{aligned}$$

This concludes the proof. ■

One should notice that as long as the algorithm is started from the network g observed in the data (which is assumed to be a draw from the stationary equilibrium of the model), the equality in (37) is satisfied for any R.

The approximate exchange algorithm presented in this paper removes the requirement of exact sampling by exploiting the property of the stationary equilibrium characterization, described in Lemma 1.

B.2 Posterior Simulation

In this section I provide the technical details for the algorithm proposed in the empirical part of the paper. The first set of results show that the exchange algorithm generate (approximate) samples from the posterior distribution (16).

The original exchange algorithm developed in Murray et al. (2006) is slightly different from the one used here. The main modification is in Step 2: the original algorithm requires an *exact* sample from the stationary equilibrium of the model.

ALGORITHM 3 ***(EXACT EXCHANGE ALGORITHM)***
Start at current parameter $\theta_t = \theta$ and network data g.

1. *Propose a new parameter vector θ'*

$$\theta' \sim q_\theta(\cdot|\theta) \tag{38}$$

2. *Draw an exact sample network g' from the likelihood*

$$g' \sim \pi\left(\cdot|X,\theta'\right) \tag{39}$$

3. *Compute the acceptance ratio*

$$\begin{aligned}\alpha_{ex}\left(\theta,\theta',g',g\right) &= \min\left\{1,\frac{\exp\left[Q(g',X,\theta)\right]}{\exp\left[Q(g,X,\theta)\right]}\frac{p\left(\theta'\right)}{p\left(\theta\right)}\frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)}\frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]}\frac{c(\theta)c(\theta')}{c(\theta)c(\theta')}\right\} \\ &= \min\left\{1,\frac{\exp\left[Q(g',X,\theta)\right]}{\exp\left[Q(g,X,\theta)\right]}\frac{p\left(\theta'\right)}{p\left(\theta\right)}\frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)}\frac{\exp\left[Q(g,X,\theta')\right]}{\exp\left[Q(g',X,\theta')\right]}\right\}\end{aligned} \tag{40}$$

4. *Update the parameter according to*

$$\theta_{t+1} = \begin{cases}\theta' & \text{with prob. } \alpha_{ex}\left(\theta,\theta',g',g\right) \\ \theta & \text{with prob. } 1-\alpha_{ex}\left(\theta,\theta',g',g\right)\end{cases} \tag{41}$$

The difference between this algorithm and the approximate one is in step 2. The exact and approximate algorithms use the same acceptance ratio $\alpha_{ex}(\theta, \theta', g', g)$, a consequence of LEMMA 1. Indeed the acceptance ratio for the approximate algorithm is

$$\begin{aligned}
\widetilde{\alpha}_{ex}(\theta, \theta', g', g) &= \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g, X, \theta)]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)}\right\} && (42)\\
&= \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g, X, \theta)]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp[Q(g, X, \theta')]}{\exp[Q(g', X, \theta')]}\right\} && (43)\\
&= \alpha_{ex}(\theta, \theta', g', g) && (44)
\end{aligned}$$

This result implies that to prove the convergence of the approximate algorithm to the exact algorithm, there is no need to prove convergence of $\widetilde{\alpha}_{ex}(\theta, \theta', g', g)$ to $\alpha_{ex}(\theta, \theta', g', g)$. The convergence of step 2 of the algorithm is sufficient.

B.2.1 Preliminary Lemmas for THEOREM 2

The convergence of the approximate exchange algorithm to the correct posterior distribution is proven in 4 steps.

1. First we prove that the exact exchange algorithm converges to the correct posterior (LEMMA 2)
2. Second, we prove that the approximate algorithm has a stationary distribution and it is ergodic (LEMMA 3, similar to the one in Liang 2010)
3. Third, we prove that the transition kernel of the approximate and exact algorithms are arbitrarily close for a large enough number of network simulations (LEMMA 4)
4. Fourth, we combine previous results to prove that the approximate algorithm converges to the correct posterior

A similar proof strategy is contained in Liang et al. (2010) and Andrieu and Roberts (2009).

Let $Q(d\vartheta|\theta) = q_\theta(\vartheta|\theta)\nu(d\vartheta)$. The transition kernel of the exact exchange algorithm can be written as

$$\begin{aligned}
P(\theta, d\vartheta) &= \left[\sum_{g' \in \mathcal{G}} \pi(g', \vartheta)\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta)\\
&+ \delta_\theta(d\vartheta)\left\{1 - \int_\Theta \left[\sum_{g' \in \mathcal{G}} \pi(g', \vartheta)\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta)\right\}
\end{aligned}$$

and the transition kernel of the approximate exchange algorithm can be written as

$$\begin{aligned}
\widetilde{P}_R(\theta, d\vartheta) &= \left[\sum_{g' \in \mathcal{G}} \mathcal{P}_\vartheta^{(R)}(g'|g)\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta)\\
&+ \delta_\theta(d\vartheta)\left\{1 - \int_\Theta \left[\sum_{g' \in \mathcal{G}} \mathcal{P}_\vartheta^{(R)}(g'|g)\alpha_{ex}(\theta, \vartheta, g', g)\right] Q(\theta, d\vartheta)\right\}
\end{aligned}$$

Let $\eta(\theta)$ be the average rejection probability for the approximate algorithm, i.e.

$$\eta(\theta) := 1 - \int_{\Theta} \left[\sum_{g' \in \mathcal{G}} \mathcal{P}_{\vartheta}^{(R)}(g'|g)\, \alpha_{ex}(\theta, \vartheta, g', g) \right] Q(\theta, d\vartheta) \tag{45}$$

The next lemma proves that the transition kernel satisfies the detailed balance condition for the posterior distribution. For any pair of parameters $(\theta, \vartheta) \in \Theta$ we have

$$P[\theta, \vartheta|g, X]\, p(\theta|g, X) = \Pr[\theta|\vartheta, g, X]\, p(\vartheta|g, X) \tag{46}$$

The detailed balance condition is sufficient condition for the Markov chain generated by the algorithm to have stationary distribution the posterior (16) (for details see Robert and Casella (2005) or Gelman et al. (2003)).

LEMMA 2 *The exchange algorithm produces a Markov chain with invariant distribution (16).*

Proof. Define $\mathcal{Z} \equiv \int_{\Theta} \pi(g|X, \theta)\, p(\theta)\, d\theta$. In the algorithm the probability $\Pr[\vartheta|\theta, g, X]$ of transition to θ_j, given the current parameter θ and the observed data (g, X), can be computed as

$$\Pr[\vartheta|\theta, g, X] = q_{\theta}(\vartheta|\theta) \frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \alpha_{ex}(\theta, \vartheta, g', g)\,. \tag{47}$$

This is the probability $q_{\theta}(\vartheta|\theta)$ of proposing ϑ times the probability of generating the new network g' from the model's stationary distribution, $\frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)}$ and accepting the proposed parameter $\alpha_{ex}(\theta, \vartheta, g', g)$. Therefore the left-hand side of (46) can be written as

$$\begin{aligned}
\Pr[\vartheta|\theta, g, X]\, p(\theta|g, X) &= q_{\theta}(\vartheta|\theta) \frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \alpha_{ex}(\theta, \vartheta, g', g)\, p(\theta|g, X) \\
&= q_{\theta}(\vartheta|\theta) \frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \alpha_{ex}(\theta, \vartheta, g', g) \frac{\frac{\exp[Q(g, X, \theta)]}{c(\mathcal{G}, X, \theta)} p(\theta)}{\mathcal{Z}} \\
&= q_{\theta}(\vartheta|\theta) \frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \\
&\times \min\left\{1, \frac{\exp[Q(g', X, \theta)]}{\exp[Q(g, X, \theta)]} \frac{p(\vartheta)}{p(\theta)} \frac{q_{\theta}(\theta|\vartheta)}{q_{\theta}(\vartheta|\theta)} \frac{\exp[Q(g, X, \vartheta)]}{\exp[Q(g', X, \vartheta)]}\right\} \\
&\times \frac{\frac{\exp[Q(g, X, \theta)]}{c(\mathcal{G}, X, \theta)} p(\theta)}{\mathcal{Z}}
\end{aligned}$$

$$= \min\left\{ q_{\theta}(\vartheta|\theta) \frac{\exp[Q(g', X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \frac{\exp[Q(g, X, \theta)]}{c(\mathcal{G}, X, \theta)} \frac{p(\theta)}{\mathcal{Z}}, q_{\theta}(\theta|\vartheta) \frac{\exp[Q(g', X, \theta]}{c(\mathcal{G}, X, \theta)} \frac{\exp[Q(g, X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \frac{p(\vartheta)}{\mathcal{Z}} \right\}$$

$$\begin{aligned}
&= q_{\theta}(\theta|\vartheta) \frac{\exp[Q(g', X, \theta)]}{c(\mathcal{G}, X, \theta)} \frac{\exp[Q(g, X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \frac{p(\vartheta)}{\mathcal{Z}} \times \\
&\times \min\left\{1, \frac{\exp[Q(g', X, \vartheta)]}{\exp[Q(g, X, \vartheta)]} \frac{p(\theta)}{p(\vartheta)} \frac{q_{\theta}(\vartheta|\theta)}{q_{\theta}(\theta|\vartheta)} \frac{\exp[Q(g, X, \theta)]}{\exp[Q(g', X, \theta)]}\right\} \\
&= q_{\theta}(\theta|\vartheta) \frac{\exp[Q(g', X, \theta)]}{c(\mathcal{G}, X, \theta)} \alpha(\vartheta, \theta, g', g) \frac{\exp[Q(g, X, \vartheta)]}{c(\mathcal{G}, X, \vartheta)} \frac{p(\vartheta)}{\mathcal{Z}} \\
&= q_{\theta}(\theta|\vartheta) \frac{\exp[Q(g', X, \theta)]}{c(\mathcal{G}, X, \theta)} \alpha(\vartheta, \theta, g', g) p(\vartheta|g, X) \\
&= \Pr[\theta|\vartheta, g, X]\, p(\vartheta|g, X)
\end{aligned}$$

The latter step proves the detailed balance for a generic network g'. Since the condition is satisfied for any network g', detailed balance follows from summing over all possible networks. ■

LEMMA 3 *(The approximate algorithm is ergodic)*
Assume the exact exchange algorithm is ergodic and that for any $\vartheta \in \Theta$

$$\frac{\mathcal{P}_{\vartheta}^{(R)}(g'|g)}{\pi(g',\vartheta)} > 0 \quad \textit{for any } g' \in \mathcal{G} \tag{48}$$

Then for any $R \in \mathbb{N}$ *such that for any* $\theta \in \Theta$, $\rho(\theta) > 0$, *the transition kernel of the approximate algorithm* $\widetilde{P}_R$ *is also irreducible and aperiodic, and there exists a stationary distribution* $\widetilde{p}(\theta)$ *such that*

$$\lim_{s\to\infty} \left\| \widetilde{P}_R^{(s)}(\theta_0,\cdot) - \widetilde{p}(\theta) \right\|_{TV} = 0 \tag{49}$$

Proof. The exact algorithm with transition kernel P is an irreducible and aperiodic Markov chain. To prove that the approximate algorithm with transition kernel $\widetilde{P}_R$ defines an ergodic Markov chain, it is sufficient to prove that the set of accessible states of P are also included in those of $\widetilde{P}_R$. The proof proceeds by induction. Formally, we need to show that for any $s \in \mathbb{N}$, $\theta \in \Theta$ and $A \in \mathcal{B}(\Theta)$ such that $P^{(s)}(\theta, A) > 0$, implies $\widetilde{P}_R^{(s)}(\theta, A) > 0$.
Notice that for any $\theta \in \Theta$ and $A \in \mathcal{B}(\Theta)$,

$$\begin{aligned}
\widetilde{P}_R(\theta, A) &= \int_A \left[\sum_{g'\in\mathcal{G}} \mathcal{P}_{\vartheta}^{(R)}(g'|g)\, \alpha_{ex}(\theta,\vartheta,g',g) \right] q_\theta(\vartheta|\theta)\, d\vartheta + \mathbb{I}(\theta \in A)\, \eta(\theta) \\
&\geq \int_A \left[\sum_{g'\in\mathcal{G}} \min\left\{ 1, \frac{\mathcal{P}_{\vartheta}^{(R)}(g'|g)}{\pi(g',\vartheta)} \right\} \pi(g',\vartheta)\, \alpha_{ex}(\theta,\vartheta,g',g) \right] q_\theta(\vartheta|\theta)\, d\vartheta + \mathbb{I}(\theta \in A)\, \eta(\theta) > 0
\end{aligned}$$

where the last inequality comes from $\frac{\mathcal{P}_{\vartheta}^{(R)}(g'|g)}{\pi(g',\vartheta)} > 0$ for any $g' \in \mathcal{G}$ and $\vartheta \in \Theta$.
This proves that the statement is true when $s = 1$. By induction we assume that it is true up to $s = n \geq 1$ and for some $\theta \in \Theta$ chose $A \in \mathcal{B}(\Theta)$ such that $P^{(n+1)}(\theta, A) > 0$ and assume that

$$\int_\Theta \widetilde{P}_R^{(n)}(\theta, d\vartheta)\, \widetilde{P}_R(\vartheta, A) = 0$$

This implies that $\widetilde{P}_R(\vartheta, A) = 0$, $\widetilde{P}_R^{(n)}(\theta,\cdot)$-a.s.; by the induction assumption at $s = 1$ it follows that $P(\vartheta, A) = 0$, $\widetilde{P}_R^{(n)}(\theta,\cdot)$-a.s.
From this and the induction assumption at $s = n$, $P(\vartheta, A) = 0$, $P^{(n)}(\theta,\cdot)$-a.s. (assume not, then $P(\vartheta, A) > 0$, $P^{(n)}(\theta,\cdot)$-a.s. which by induction would imply $\widetilde{P}_R(\vartheta, A) > 0$, which is a contradiction). The latter step contradicts $P^{(n+1)}(\theta, A) > 0$ and the result follows. ■

The next step consists of proving that the transition kernel of the approximate algorithm $\widetilde{P}_R(\theta,\vartheta)$ and the exact algorithm $P(\theta,\vartheta)$ are arbitrarily close for a large enough number of network simulations R. Formally we prove a statement which is equivalent to proving convergence in total variation norm.[44]

[44] See Levin et al. (2008), proposition 4.5, page 49.

LEMMA 4 *(Convergence of the exact and approximate transition kernels)*
Let $\epsilon \in (0,1]$. There exists a number of simulations $R_0 \in \mathbb{N}$ such that for any function $\phi : \Theta \to [-1,1]$ and any $R > R_0$,

$$\left| \widetilde{P}_R \phi(\theta) - P\phi(\theta) \right| < 2\epsilon \tag{50}$$

Proof. The transition of the exchange algorithm is

$$\begin{aligned} P\left(\phi(\theta), \phi(\vartheta)\right) &= \int_\Theta \phi(\vartheta) \left[\sum_{g' \in \mathcal{G}} \pi\left(g', \vartheta\right) \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \\ &+ \phi(\theta) \left[1 - \int_\Theta \left[\sum_{g' \in \mathcal{G}} \pi\left(g', \vartheta\right) \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \right] \end{aligned}$$

while the transition kernel for the approximate algorithm is

$$\begin{aligned} \widetilde{P}_R\left(\phi(\theta), \phi(\vartheta)\right) &= \int_\Theta \phi(\vartheta) \left[\sum_{g' \in \mathcal{G}} \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \\ &+ \phi(\theta) \left[1 - \int_\Theta \left[\sum_{g' \in \mathcal{G}} \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \right] \end{aligned}$$

and therefore the difference is

$$\begin{aligned} S &= P\left(\phi(\theta), \phi(\vartheta)\right) - \widetilde{P}_R\left(\phi(\theta), \phi(\vartheta)\right) \\ &= \int_\Theta \phi(\vartheta) \left[\sum_{g' \in \mathcal{G}} \left[\pi\left(g', \vartheta\right) - \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \right] \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \\ &- \phi(\theta) \int_\Theta \left[\sum_{g' \in \mathcal{G}} \left[\pi\left(g', \vartheta\right) - \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \right] \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \end{aligned}$$

Consider the quantity

$$\begin{aligned} S_0 &= \int_\Theta \left[\sum_{g' \in \mathcal{G}} \left[\pi\left(g', \vartheta\right) - \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \right] \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \\ &\leq \int_\Theta \left[\sum_{g' \in \mathcal{G}} \left| \pi\left(g', \vartheta\right) - \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \right| \alpha_{ex}\left(\theta, \vartheta, g', g\right) \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \end{aligned}$$

and since $\alpha_{ex}\left(\theta, \vartheta, g', g\right) \leq 1$ for any $(\theta, \vartheta) \in \Theta \times \Theta$ and $(g', g) \in \mathcal{G} \times \mathcal{G}$, we have

$$\begin{aligned} S_0 &\leq \int_\Theta \left[\sum_{g' \in \mathcal{G}} \left| \pi\left(g', \vartheta\right) - \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \right| \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \\ &= \int_\Theta \left[2 \sup_{g' \in \mathcal{G}} \left| \pi\left(g', \vartheta\right) - \mathcal{P}_\vartheta^{(R)}\left(g'|g\right) \right| \right] q_\theta\left(\vartheta|\theta\right) d\vartheta \end{aligned}$$

The convergence of the network simulation algorithm implies that for any $\varepsilon > 0$, there exists an $R_0(\vartheta, \varepsilon) \in \mathbb{N}$ such that for any $R > R_0(\vartheta, \varepsilon)$ and for any $g \in \mathcal{G}$

$$2 \sup_{g' \in \mathcal{G}} \left| \pi(g', \vartheta) - \mathcal{P}_{\vartheta}^{(R)}(g'|g) \right| \leq \varepsilon$$

Pick $R_0(\varepsilon) = \max_{\vartheta \in \Theta} \{R_0(\vartheta, \varepsilon)\}$. Then for any $\epsilon \in (0, 1]$, there is an $R_0(\epsilon) \in \mathbb{N}$ such that for any $R > R_0(\epsilon)$ and for any $g \in \mathcal{G}$

$$S_0 \leq \int_{\Theta} \epsilon q_{\theta}(\vartheta|\theta) \, d\vartheta = \epsilon$$

This implies that

$$|S| \leq |2S_0| = 2\epsilon \tag{51}$$

■

The next theorem is the main result for the convergence. It states that the approximate exchange algorithm converges to the correct posterior distribution, provided that the number of network simulations and parameter samples are big enough.

B.2.2 Proof of THEOREM 2

. **Proof.** The main idea is to decompose the total variation in two components

$$\begin{aligned} \left\| \widetilde{P}_R^{(s)}(\theta_0, \cdot) - p(\cdot|g, X) \right\|_{TV} &= \left\| \widetilde{P}_R^{(s)}(\theta_0, \cdot) - P^{(s)}(\theta_0, \cdot) + P^{(s)}(\theta_0, \cdot) - p(\cdot|g, X) \right\|_{TV} \\ &\leq \left\| \widetilde{P}_R^{(s)}(\theta_0, \cdot) - P^{(s)}(\theta_0, \cdot) \right\|_{TV} + \left\| P^{(s)}(\theta_0, \cdot) - p(\cdot|g, X) \right\|_{TV} \end{aligned}$$

and prove that each component converges. We will use the same idea, but rewrite the total variation in a more convenient form.[45] For any function $\phi : \Theta \to [-1, 1]$ we have

$$\begin{aligned} \left| \widetilde{P}_R^{(s)} \phi(\theta_0) - p(\phi) \right| &= \left| \widetilde{P}_R^{(s)} \phi(\theta_0) - P^{(s)} \phi(\theta_0) + P^{(s)} \phi(\theta_0) - p(\phi) \right| \\ &\leq \left| \widetilde{P}_R^{(s)} \phi(\theta_0) - P^{(s)} \phi(\theta_0) \right| + \left| P^{(s)} \phi(\theta_0) - p(\phi) \right| \end{aligned}$$

The second component converges because the exact exchange algorithm is ergodic, as stated in Lemma. For any $\varepsilon > 0$ there is number of simulation steps $s(\theta_0, \varepsilon)$, such that for any $s \geq s(\theta_0, \varepsilon)$

$$\left| P^{(s)} \phi(\theta_0) - p(\phi) \right| \leq \varepsilon \tag{52}$$

For the remaining of the proof, I will set $s_0 := s(\theta_0, \varepsilon)$. I use the telescoping sum decomposition in Andrieu and Roberts (2009) (page 15, adapted from last formula)

$$\begin{aligned} \left| \widetilde{P}_R^{(s_0)} \phi(\theta_0) - P^{(s_0)} \phi(\theta_0) \right| &= \left| \sum_{l=0}^{s_0 - 1} \left[P^{(l)} \widetilde{P}_R^{(s_0 - l)} \phi(\theta_0) - P^{(l+1)} \widetilde{P}_R^{(s_0 - (l+1))} \phi(\theta_0) \right] \right| \\ &= \left| \sum_{l=0}^{s_0 - 1} P^{(l)} \left(\widetilde{P}_R - P \right) \widetilde{P}_R^{(s_0 - (l+1))} \phi(\theta_0) \right| \end{aligned}$$

[45] See Levin et al. (2008), proposition 4.5, page 49.

Now we can apply s_0 times the result of LEMMA 4 (as in Liang et al. (2010) and Andrieu and Roberts (2009)) to prove that there exists an $R_0(\theta_0, \varepsilon) \in \mathbb{N}$ such that for any $R > R_0(\theta_0, \varepsilon)$

$$\left| \widetilde{P}_R^{(s_0)} \phi(\theta_0) - P^{(s_0)} \phi(\theta_0) \right| \leq 2 s_0 \varepsilon \tag{53}$$

this implies

$$\left| \widetilde{P}_R^{(s)} \phi(\theta_0) - p(\phi) \right| \leq (2 s_0 + 1) \varepsilon \tag{54}$$

We conclude the proof by choosing $\varepsilon = \epsilon / (2 s_0 + 1)$.

This proves that the approximate exchange algorithm is ergodic, therefore the law of large number holds, and the second part of the theorem is proven. ■

B.3 Convergence Experiments

In this section, I provide an overview of the convergence properties of the algorithm using examples with artificial data. Assume a toy model with three parameters, with an utility function of the following form

$$U_i(g, X) = \sum_{j=1}^{n} g_{ij} \theta_1 + \sum_{j=1}^{n} g_{ij} g_{ji} \theta_2 + \sum_{j=1}^{n} g_{ij} \sum_{k \neq i,j;k=1}^{n} g_{jk} \theta_3 + \sum_{j=1}^{n} g_{ij} \sum_{k \neq i,j;k=1}^{n} g_{ki} \theta_3 \tag{55}$$

The artificial data are generated using the vector of parameters

$$\theta = (-2.0, 0.5, 0.01) \tag{56}$$

To obtain the network dataset for the estimation, the network simulation algorithm is started at a random network and then ran for 1 million iterations. The initial random network is generated by assuming each link is independent and the probability of a link is $p = .2$. The last iteration of this long simulation is used as dataset in all the estimation exercises below. I report results for a network with $n = 50$ agents, but I ran the same simulations using a network with $n = 30$ and $n = 100$, with similar results.

To check if the exchange algorithm converges to the right region of the parameter space, the parameter simulations are started from 5 different starting values

$$\begin{aligned} \theta^1 &= (-2.0, 0.5, 0.01) \\ \theta^2 &= (-10.0, 5.0, 1.0) \\ \theta^3 &= (10.0, -5.0, -1.0) \\ \theta^4 &= (-3.0, -0.05, 0.3) \\ \theta^5 &= (-20.0, 15.0, -0.3) \end{aligned}$$

The first is the parameter vector that generates the data, while the others are overdispersed initial values. In Figure 8 I display the convergence of the simulations to the high density region of the posterior. In this example the number of network simulations per parameter proposal is $R = 3000$.[46] The solid horizontal black line represents the parameter that generated the data. Each color represents a simulation started at one of the initial values above. After 2000 iterations all the chains have converged to the region of the posterior that contains the data generating parameters. In Figure 9 I show the autocorrelation functions for the same example. In this example the autocorrelation disappears after 200 lags. This is mainly due to the small amount of parameters in this toy model. High dimensional models show more persistent autocorrelation of the chains. In Figure 10 I show the same convergence properties of Figure 8 by plotting two parameters in each graph. I show 3 snapshots of the simulations: at 500, 1000 and 2000 iterations. The dashed lines intersect at the parameter values that generated the data. After 500 iterations (Panel A) almost all chains have converged to the high density region. The purple chain converges after 2000 iterations: this is because this chain corresponds to the 5th starting value, which is the quite far from the parameter that generated the

network. In summary, convergence in this toy model is quite fast. For higher dimensional models convergence is slower, but reasonable, in the order of 50 or 100 thousands iterations. One possible strategy is to use a small R for the initial simulations: when the chain reaches approximate convergence we can increase the number of network simulations and estimate the posterior with higher precision.

B.4 Parallel estimation with multiple networks

When data from multiple independent networks are available the estimation routines are easily adapted. Assume the researcher has data from C networks: let g_c and X_c denote the network matrix and the individual controls for network c, $c = 1, ..., C$. The aggregate data are denoted as $g = \{g_1, ..., g_c\}$ and $X = \{X_1, ..., X_c\}$.

Assuming each network is drawn from the stationary equilibrium of the model, each network has distribution

$$\pi(g_c, X_c, \theta) = \frac{\exp[Q(g_c, X_c, \theta)]}{\sum_{\omega \in \mathcal{G}_c} \exp[Q(\omega_c, X_c, \theta)]} \tag{57}$$

Since each network is independent, the likelihood of the data (g, X) can be written as

$$\begin{aligned} \pi(g, X, \theta) &= \prod_{c=1}^{C} \pi(g_c, X_c, \theta) = \prod_{c=1}^{C} \left\{ \frac{\exp[Q(g_c, X_c, \theta)]}{c(\mathcal{G}_c, X_c, \theta)} \right\} \\ &= \frac{\exp\left[\sum_{c=1}^{C} Q(g_c, X_c, \theta)\right]}{\prod_{c=1}^{C} c(\mathcal{G}_c, X_c, \theta)} = \frac{\exp\left[\sum_{c=1}^{C} Q(g_c, X_c, \theta)\right]}{\mathcal{C}(\mathcal{G}, X, \theta)} \end{aligned}$$

where $\mathcal{G} = \bigcup_{c=1}^{C} \mathcal{G}_c$ and $X = \{X_1, ..., X_C\}$. The likelihood for multiple independent networks is of the same form as the likelihood for one network observation. The structure of this likelihood makes parallelization extremely easy: each network can be simulated independently using the network simulation algorithm; at the end of the simulation we collect the last network and compute the potential; then we compute the sum of potentials and use it to compute the probability of update.
Therefore, the algorithm is modified as follows

ALGORITHM 4 ***(Parallel FAST EXCHANGE ALGORITHM)***
Fix the number of simulations R. Store each network data (g_c, X_c) in a different processor/core. At each iteration t, with current parameter $\theta_t = \theta$ and network data g

1. *Propose a new parameter θ' from a distribution $q_\theta(\cdot|\theta)$*

$$\theta' \sim q_\theta(\cdot|\theta) \tag{58}$$

2. *For each processor c, start **ALGORITHM 1** at the observed network g_c, iterating for R steps using parameter θ' and collect the last simulated network g'_c*

$$g'_c \sim \mathcal{P}^{(R)}_{\theta'}(g'_c|g_c) \tag{59}$$

3. *Update the parameter according to*

$$\theta_{t+1} = \begin{cases} \theta' & \text{with prob. } \alpha_{pex}(\theta, \theta') \\ \theta & \text{with prob. } 1 - \alpha_{pex}(\theta, \theta') \end{cases}$$

[46] Similar results hold for different R values.

Table 5: Convergence Experiments

Starting value 1						
	true		R=1000	R=2000	R=3000	R=5000
θ_1	-2.000	mean	-2.0165	-2.0643	-2.077	-2.0838
		s.d.	0.2629	0.2018	0.1845	0.1635
		mc s.e.	0.0125	0.0069	0.0063	0.0051
θ_2	0.500	mean	0.5387	0.6083	0.6207	0.6158
		s.d.	0.5519	0.4435	0.4144	0.4076
		mc s.e.	0.0338	0.0294	0.0189	0.0279
θ_3	0.010	mean	0.0043	0.0121	0.0147	0.0175
		s.d.	0.0262	0.0201	0.0187	0.0165
		mc s.e.	0.0002	0.0001	0.0001	0.0001
Starting value 2						
	true		R=1000	R=2000	R=3000	R=5000
θ_1	-2.000	mean	-2.0131	-2.0651	-2.0688	-2.0673
		s.d.	0.2643	0.2013	0.1814	0.1655
		mc s.e.	0.0137	0.0067	0.0057	0.0046
θ_2	0.500	mean	0.5542	0.6181	0.6149	0.6571
		s.d.	0.5506	0.4425	0.4228	0.4046
		mc s.e.	0.0363	0.0279	0.029	0.022
θ_3	0.010	mean	0.0041	0.0119	0.0143	0.0157
		s.d.	0.0267	0.0201	0.0185	0.0167
		mc s.e.	0.0002	0.0001	0.0001	0.0001
Starting value 3						
	true		R=1000	R=2000	R=3000	R=5000
θ_1	-2.000	mean	-2.0287	-2.0583	-2.0656	-2.0686
		s.d.	0.2548	0.2072	0.1883	0.164
		mc s.e.	0.0099	0.0081	0.0085	0.0043
θ_2	0.500	mean	0.5723	0.6028	0.6275	0.6593
		s.d.	0.5418	0.4473	0.4084	0.3844
		mc s.e.	0.034	0.0224	0.0283	0.0207
θ_3	0.010	mean	0.0058	0.0113	0.0128	0.016
		s.d.	0.0255	0.0211	0.0203	0.0167
		mc s.e.	0.0002	0.0001	0.0001	0.0001
Starting value 4						
	true		R=1000	R=2000	R=3000	R=5000
θ_1	-2.000	mean	-2.016	-2.0727	-2.0884	-2.0724
		s.d.	0.2574	0.2033	0.1842	0.1625
		mc s.e.	0.01	0.0064	0.007	0.0051
θ_2	0.500	mean	0.5612	0.5993	0.6354	0.6576
		s.d.	0.5436	0.4442	0.4163	0.4044
		mc s.e.	0.0346	0.027	0.0252	0.0256
θ_3	0.010	mean	0.0047	0.0128	0.0158	0.0162
		s.d.	0.0254	0.0205	0.0181	0.0165
		mc s.e.	0.0002	0.0001	0.0001	0.0001
Starting value 5						
	true		R=1000	R=2000	R=3000	R=5000
θ_1	-2.000	mean	-2.0309	-2.056	-2.0823	-2.0794
		s.d.	0.2522	0.2059	0.1803	0.1648
		mc s.e.	0.0113	0.007	0.0056	0.0051
θ_2	0.500	mean	0.5668	0.6246	0.654	0.6539
		s.d.	0.5464	0.4389	0.416	0.3966
		mc s.e.	0.0399	0.0244	0.0249	0.0213
θ_3	0.010	mean	0.0061	0.0104	0.0153	0.0168
		s.d.	0.0253	0.0209	0.0183	0.0169
		mc s.e.	0.0002	0.0001	0.0001	0.0001

where

$$\alpha_{pex}(\theta,\theta') = \min\left\{1, \frac{\exp\left[\sum_{c=1}^{C} Q(g_c', X_c, \theta)\right]}{\exp\left[\sum_{c=1}^{C} Q(g_c, X_c, \theta)\right]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp\left[\sum_{c=1}^{C} Q(g_c, X_c, \theta')\right]}{\exp\left[\sum_{c=1}^{C} Q(g_c', X_c, \theta')\right]}\right\} \tag{60}$$

The speed of the algorithm depends on the largest network in the data. Since each parameter update requires the result of each processor simulation there is some idle time, since small networks are simulated much faster.

B.5 Freeman Segregation Index

The Freeman segregation index measures the degree of segregation in a population with two groups (Freeman, 1972). Assume there are two groups, A and B. Let n_{AB} be the total number of links that individuals of group A form to individuals of group B. Let n_{BA}, n_{BB} and n_{AA} be analogously defined. The original index developed by Freeman (1972) is defined as

$$FSI = \frac{\mathbb{E}[n_{AB}] + \mathbb{E}[n_{BA}] - (n_{AB} + n_{BA})}{\mathbb{E}[n_{AB}] + \mathbb{E}[n_{BA}]} \tag{61}$$

When the link formation does not depend on the identity of individuals, then the links should be randomly distributed with respect to identity. Therefore, the index measures the difference between the expected and actual number of links among individuals of different groups, as a fraction of the expected links. An index of 0 means that the actual network closely resembles one in which links are formed at random. Higher values indicate more segregation. In this paper segregation is measured using the index[47]

$$SEG = \max\{0, FSI\} \tag{62}$$

The index varies between 0 and 1, where the maximum corresponds to a network in which there are no cross-group links.

To complete the derivation of the index, the expected number of cross-group links is computed as

$$\begin{aligned} \mathbb{E}[n_{AB}] &= \frac{(n_{AA} + n_{AB})(n_{AB} + n_{BB})}{n_{AA} + n_{AB} + n_{BA} + n_{BB}} \\ \mathbb{E}[n_{BA}] &= \frac{(n_{BA} + n_{BB})(n_{AA} + n_{BA})}{n_{AA} + n_{AB} + n_{BA} + n_{BB}} \end{aligned}$$

C Extensions

It is possible to incorporate unobserved heterogeneity or random coefficients in the model. However this would significantly increase the computational cost of estimation. The simplest way to introduce unobserved heterogeneity is to model the preference shock ε_{ij} as incorporating individual random effects. In our application of school friendship networks the unobserved quality of the student could be interpreted as "coolness" or personality or attractiveness. The decision of the player to form a link is modified as follows

$$U_i(g_{ij} = 1, g_{-ij}, X) + \eta_i + \eta_j + \nu_{ij1} \geq U_i(g_{ij} = 0, g_{-ij}, X) + \eta_i + \nu_{ij0} \tag{63}$$

where ν_{ij} is an i.i.d. shock with logistic distribution and the vector $\eta = \{\eta_1, ..., \eta_n\}$ is drawn at time 0 from a known distribution $W(\eta)$. In this formulation I assume that the players observe the random effect η but the econometrician does not. Notice that the random effect of player i cancels out, while the choice of linking j is conditional on the random effect of player j (which is present only when the link is formed).

Conditioning on the realization of the vector $\eta \in \Upsilon$, the potential function is modified as follows

$$\mathcal{Q}(g, X, \theta; \eta) = Q(g, X, \theta) + \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\eta_j \tag{64}$$

[47] The index (61) varies between -1 and 1. However, the interpretation of the index when it assumes negative values is not clear. Therefore Freeman (1972) suggests to use only when it is nonnegative, to measure the presence of segregation

To compute the unconditional likelihood we need to integrate out the unobserved vector η to obtain

$$\pi(g, X, \theta) = \int_{\Upsilon} \frac{\exp[\mathcal{Q}(g, X, \theta; \eta)]}{\sum_{\omega \in \mathcal{G}} \exp[\mathcal{Q}(\omega, X, \theta; \eta)]} dW(\eta) \tag{65}$$

The integral above can be computed using Monte Carlo techniques, as it is standard in the IO literature or labor economics. However, the model does not allow standard Monte Carlo, because of the normalizing constant.

A more feasible strategy is to use data augmentation and Markov Chain Monte Carlo methods as in the discrete choice literature (Rossy Mc Cullugh and Allenby 1996, Atehy and Imbens ??).

Conditioning on the realization of the unobserved component η, we can use the exchange algorithm to sample from the posterior distribution of θ. Conditioning on the proposed θ we can use a metropolis hastings step to sample the unobserved component η.

Given an initial (θ, η) at simulation s, we propose a new θ' and use the exchange algorithm to accept or reject the proposal. Given the new value of θ_{s+1}, we propose a new vector of unobserved components η' and accept using a Metropolis Hastings step. The probability of η, conditioning on (θ, g, X) is

$$\Pr(\eta | g, X, \theta) = \frac{W(\eta)\, \pi(g, X, \theta; \eta)}{\pi(g, X, \theta)} \tag{66}$$

The Metropolis-Hastings step proceeds by proposing a new η' from a distribution $q_\eta(\eta' | \eta)$, which is accepted with probability

$$\alpha_\eta(\eta, \eta', g, \theta_s) = \left\{ 1, \frac{W(\eta')\, \pi(g, X, \theta; \eta')\, q_\eta(\eta | \eta')}{W(\eta)\, \pi(g, X, \theta; \eta)\, q_\eta(\eta' | \eta)} \right\} \tag{67}$$

Similar ideas apply to random coefficients. One possibility is to estimate a specification with random coefficient at the network level.

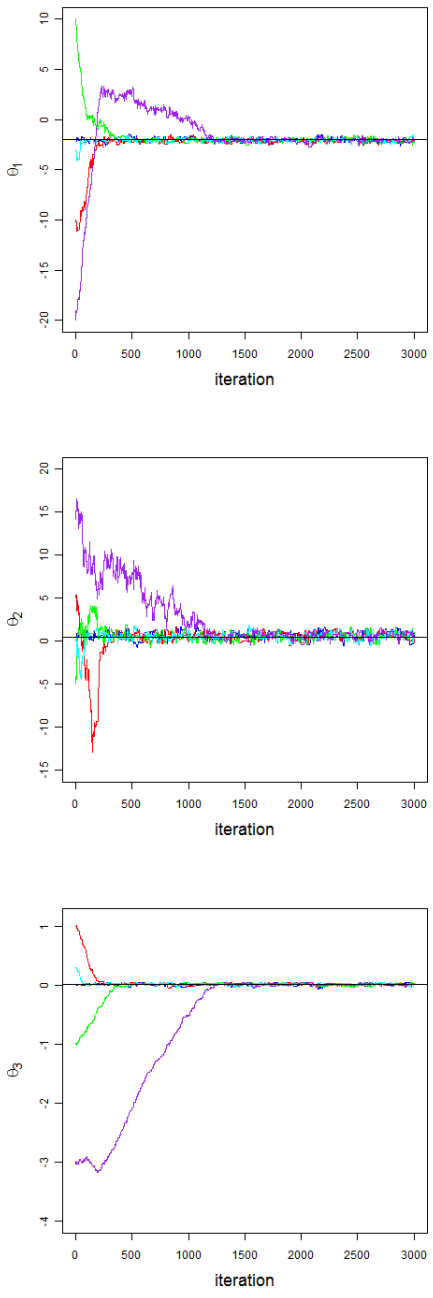
$$\theta_p = \theta_{p0} + \sum_{c=1}^{C} \theta_{pc} Z_c + \xi_c \tag{68}$$

where Z_c is a network-level variable and $\xi_c \sim \mathcal{N}(0, \sigma_\xi)$. The main computational burden

The estimation method is flexible enough to allow for estimation when there are missing links. The Add Health dataset raises concern about missing links, since the original questionnaire asked the students to report up to 5 male and 5 female friends. If a student has more than 5 male friends, those are missing from the dataset.

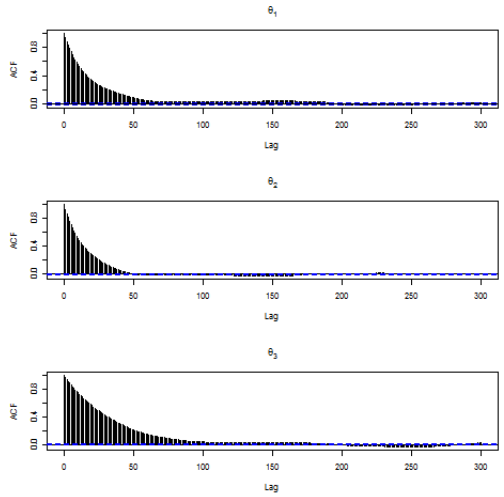
Using the Bayesian algorithm provided in this paper one could easily deal with the missing links. The algorithm cab be modified to include an additional simulation step that generates the missing links g_{mis}, given the observed network data g_{obs} and the current parameter. Then the algorithm proceeds as before using the augmented data $g = \{g_{mis}, g_{obs}\}$.

Figure 8: Convergence to the high density posterior region



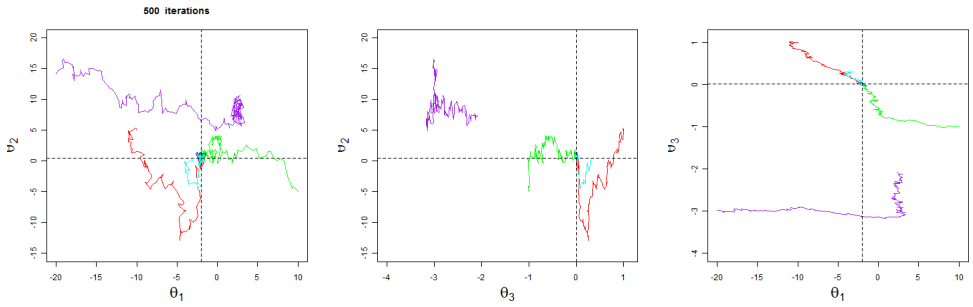
Each graph shows convergence to the high density region of the posterior distribution. The curves with different colors represent chains started at overdispersed initial values. The solid black line represent the parameter that generated the data. Convergence is very fast and we can use the initial 2000 iterations as burn-in. In this example the network has $n = 50$ agents and the number of network simulations per proposal is $R = 3000$.

Figure 9: Convergence to the high density posterior region

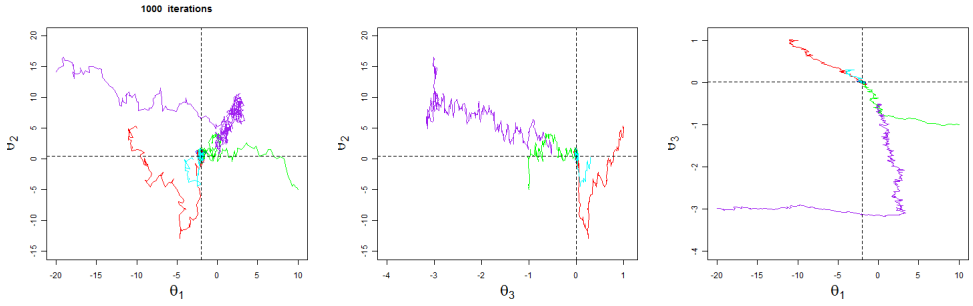


Each graph is the autocorrelation function of the chains generated by the exchange algorithm.

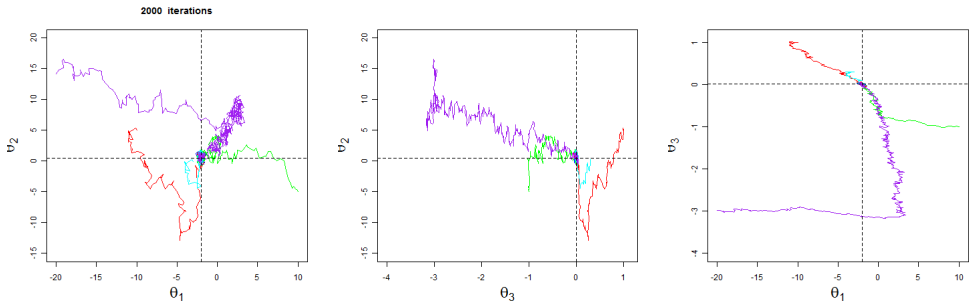
Figure 10: Convergence to the high density posterior region



Panel A. 500 iterations



Panel B. 1000 iterations



Panel C. 2000 iterations

Three snapshots of the simulations at 500, 1000 and 2000 iterations of the fast exchange algorithm. The true parameter value is indicated by the intersection of the dashed lines. After 500 iterations only few chains have converged close to the true parameters. After 1000 the remaining chains have almost reached the high density region of the posterior. At 2000 iterations the algorithm has reached approximate convergence for all the chains.