A Structural Model of Segregation in Social Networks*

Angelo Mele†

July 16, 2013

Abstract

The main challenges in estimating strategic network formation models are the presence of multiple equilibria, and the fact that the number of possible network configurations increases exponentially with the number of players. I propose a dynamic model of strategic network formation with heterogeneous players, which converges to a unique stationary equilibrium. Hence, the structural preference parameters can be estimated using a single network observation. In addition, the model provides the first equilibrium micro-foundation of exponential random graphs. Because of the curse of dimensionality, the likelihood is computationally intractable. Therefore, I propose a Bayesian estimation strategy that samples from the posterior, interleaving parameter and network simulations, without evaluating the likelihood. I prove that the proposed algorithm converges to the correct posterior distribution. A mean-field analysis shows that the algorithm converges fast for practical applications. Estimation is tested with artificial and Add Health data, confirming evidence of homophily in high schools.

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

*I am grateful to Roger Koenker, Ron Laschever, Dan Bernhardt and George Deltas for their inputs at crucial stages of this work. I thank Federico Bandi, Alberto Bisin, Ethan Cole, Aureo de Paula, Steven Durlauf, Andrea Galeotti, Shweta Gaonkar, Sanjeev Goyal, Dan Karney, Darren Lubotsky, Antonio Mele, Luca Merlino, Ferdinando Monte, 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, University of Notre Dame, Johns Hopkins University - Carey Business School, University of Pennsylvania, Southern Methodist University, FED Board, UQAM, SBIES 2011, PET 2011, Conference on Networks in Cambridge, UK for helpful comments and suggestions. I am grateful to Andriy Norets and Matt Jackson for suggesting references on how to prove and bound convergence of the algorithms. Financial support from the R. Ferber Award, the R. W. Harbeson Memorial Dissertation Fellowship, and the NET Institute Summer Research Grant 2010 is gratefully acknowledged. All remaining errors are mine.

†*Address*: Johns Hopkins University - Carey Business School, 100 International Drive, Baltimore, MD 21202. *Email*: angelo.mele@jhu.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 personal networks is endogenous: individuals choose their peers and friends according to their socioeconomic characteristics and their relationships. As a consequence, in socially generated networks the agents are likely to interact with similar individuals (homophily), segregating along socioeconomic attributes.[2] This paper provides a model and estimation methods to study segregation and homophily in social networks.

The literature on strategic models of network formation provides a framework to interpret the observed network as the equilibrium of a game.[3] However, the estimation and identification of strategic models is challenging. First, network formation models usually have multiple equilibria. Second, there is a curse of dimensionality: the number of possible network configurations increases exponentially with the number of players. Finally, the data available to the econometrician usually consist of a single network snapshot. It is therefore crucial to develop a systematic framework to deal with each of these challenges.

The model presented in this work combines ingredients from the strategic and random network formation literature.[4] The random components allow me to derive a likelihood for the observed network, while preserving the equilibrium interpretation of the strategic model. Players have preferences over network realizations and individual characteristics. The utility function includes payoffs from direct links, reciprocated links, friends of friends and popularity. In each period, a player is randomly drawn from the population and meets another individual, according to a random meeting technology. Upon meeting, the player has the opportunity to revise his link. Before updating their linking strategy, the players receive a random shock to their preferences, unobserved by the econometrician. The dynamic of the model follows a stochastic best-response dynamics, and generates a sequence of directed networks.

The theoretical results provide conditions under which the model admits a potential function, simplifying the analysis of the equilibria of the network formation game.[5] The set of Nash equilibria of the model corresponds to the maxima of the potential function. Under mild restrictions on preferences, meeting technology and standard assumptions on preferences shocks, I show that the sequence of networks generated by the model is a Markov chain, and it converges to a unique stationary equilibrium distribution. The latter provides the likelihood of observing a specific network realization in the long-run.

[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] See Jackson (2008), Jackson and Wolinsky (1996), Bala and Goyal (2000) for examples.

[4] See Jackson (2008) for a review of network formation models.

[5] See Monderer and Shapley (1996)

Estimation of the posterior distribution for the structural preference parameters is challenging because of the curse of dimensionality: the likelihood is known up to a normalizing constant that cannot be computed even for very small networks.[6] Traditional MCMC samplers like Metropolis-Hastings and Gibbs are infeasible for this model. Therefore I propose an approximate Bayesian MCMC exchange algorithm that allows sampling from the posterior distribution without evaluating the likelihood.[7] The sampler interleaves parameter and network simulations in two steps: first, a parameter is proposed with a metropolis-hastings step; second, given the proposed parameter, another metropolis-hastings sampler simulates the network formation model to draw a network from the stationary distribution.[8]

I prove that the algorithm is ergodic and converges to the correct posterior distribution. The speed of convergence depends crucially on the convergence rate of the network simulations algorithm. Extending the approach of Bhamidi et al. (2011) to directed networks through mean-field approximations for the exponential family,[9] I show that in the special case of linear utilities the convergence properties of the model are good. The reason is that the model incorporates dependencies for up to two directed links (two-mixed-stars): these dependencies are well-behaved and the region of parameters corresponding to exponentially slow mixing of the simulations is negligible. The network simulation algorithm mixes in order $n^2 \log n$ steps for most parameter vectors, guaranteeing good mixing properties for the parameter simulations. Furthermore, I show that extensions of the model that incorporate additional dependencies with more than two links, may have convergence problems. Simulations confirm that the algorithm converges well and in reasonable time for quite large networks of 1000 players.[10]

Finally, this model provides a first step towards the equilibrium micro-foundation of the exponential family of random graph models. The exponential random graph model[11] (ERGM) assumes that the likelihood of observing a specific network configuration is an exponential function of several network statistics, e.g. the number of edges, the number of triangles, etc. The ERGM is widely used in practice, but to the best of my knowledge there has been no rigorous study on the restrictions that the model imposes on preferences, meeting technology and equilibrium. I show that *some* specifications of the ERGM can be

[6] For a small network with 10 players, a state-of-the-art supercomputer may take several years to evaluate the constant at a single parameter value. This makes traditional optimization algorithms infeasible.

[7] The original algorithm was developed by Murray et al. (2006). I propose an approximate version of the algorithm that does not require exact sampling, similar to Liang (2010)

[8] This simulation scheme, interleaving the parameter simulation and the network simulation, is similar to Bayesian methods for dynamic discrete choice models developed in Norets (2009).

[9] See Wainwright and Jordan (2008) for an extended overview of mean-field methods for exponential families.

[10] Without major modification of the network simulation algorithm, the estimation of very large networks requires a huge amount of simulations. For networks with $n \geq 10000$) players the deterministic approximations and asymptotic approximations provide a valid alternative approach. Diaconis and Chatterjee (2011) shows conditions under which the mean-field approximation for the ERGM becomes asymptotically exact, while Chandrasekhar and Jackson (2012) and Amir et al. (2012) focus on consistent estimators.

[11] See for example Snijders (2002), Wasserman and Pattison (1996), Diaconis and Chatterjee (2011), Koskinen (2004).

interpreted as the stationary equilibrium of my model in the special case of linear utilities. If the preferences are linear in parameters, the stationary distribution of networks observed in the long-run belongs to the discrete exponential family and has a likelihood similar to the ERGM. However, not all the ERGM specifications are micro-founded and in general not all the network statistics are compatible with the equilibrium micro-foundation.

The estimation is tested with artificial data and with Add Health data on school friendships in US[12] The simulations show that accurate posterior estimation can be performed in reasonable time. The Add Health estimation with few networks confirm the existence of homophily in US high schools.

This paper contributes to the recent literature on estimation of social network formation models. The most closely related work is Christakis et al. (2010). In their model myopic players meet sequentially and choose which links to form by maximizing current utility. The sequence of meetings is unobservable, and therefore must be integrated out in the likelihood. This computational challenge is addressed with an MCMC scheme that samples from the space of meeting sequences. To limit the computational burden they assume that individuals can meet only once, and linking decisions are permanent. In my model, I make assumptions on the meeting technology that guarantee existence of a closed form solution for the stationary equilibrium distribution of networks. Players meet often and have the opportunity to revise their links frequently. In addition, I provide complete characterizations of the strategic equilibrium and the convergence properties of the estimation algorithm.

The paper is also related to recent work in Chandrasekhar and Jackson (2012) and Amir et al. (2012): they provide a computationally tractable inference method for network formation models, showing consistency of maximum likelihood estimators. The asymptotic analysis is developed for large n, i.e. when the number of players goes to infinity.[13] My algorithm performs well in small samples and converges quite fast for networks of size $n \leq 1000$. For larger networks, the network simulation algorithm has to be modified to include larger steps and improve mixing.[14]

An alternative approach uses pairwise stability as equilibrium concept: this usually leads to multiplicity of equilibria[15] and may preclude point identification. Sheng (2012) proposes

[12]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.

[13]The alternative approach is to prove consistency when the number of network observations grows large. However, most network datasets contain only a single observation of a network.

[14]Mele (2011) proposes a simulated tempering algorithm that allows larger steps in the simulation. The algorithm is useful when the likelihood has multiple modes. Atchade and Wang (forthcoming) consider an adaptive MCMC exchange algorithm for exponential random graph models, showing asymptotic normality of the estimated posterior.

[15]See Comola (2010) for example.

computationally feasible moment inequalities to estimate a model of strategic network formation. Since pairwise stability of the network imposes pairwise stability of any subnetwork, she can derive bounds on the probability of observing any (small) subnetwork. This reduces the computational burden due to the curse of dimensionality. Related approaches are presented in DePaula et al. (2011) and Miyauchi (2012). The model presented here avoids multiple equilibria by implicitly using the sequential network formation process as an equilibrium selection mechanism. As a consequence, point identification is possible. My estimation strategy alleviates the curse of dimensionality by use of the potential function in simulations, and by avoiding exact evaluation of the likelihood function.

In my model the existence of equilibrium is guaranteed by the potential function characterization. While the theoretical literature on network formation has used the potential game characterization to guarantee existence and avoid cycles (Gilles and Sarangi (2004), Jackson and Watts (2002)), the contribution in this paper is to show that the potential function characterization reduces the computational complexity and allows identification of structural preference parameters. The equilibrium characterization is nested in a discrete choice framework: the random meeting technology and the preference shocks generate random deviations from the Nash equilibria. The use of noise to generate unique equilibrium predictions has been exploited in the theoretical literature.[16] The novelty of my approach is to exploit the random components of the model to derive a likelihood for the observed network.

The model can be extended to include additional utility components, unobserved heterogeneity and preference heterogeneity. The estimation algorithm needs to include an additional layer of simulation for the unobserved types. Bayesian data augmentation allows estimation of models that account for missing links (see Appendix C). Recently, several models have exploited the sequential framework or the potential game characterization to endogenize both network formation and actions. Badev (2013) extends the model presented in this paper to include a binary action (smoking), and separately identify the contribution of friendship formation and individual decisions in smoking. The equilibrium characterization generalizes the meeting technology: players can meet multiple agents, allowing more complex deviations from equilibrium than single link updates. Hsieh and Lee (2012) and Goldsmith-Pinkham and Imbens (forthcoming) use similar frameworks. The computational challenges of these models closely mirror the ones shown here and are solved with MCMC methods.[17]

The paper proceeds as follows. In section 2, I present the model's details and fully characterize the potential function and the unique stationary equilibrium. In addition, the model is shown to provide micro-foundations for the ERGMS and possible extensions of the preferences are discussed. Section 3 proposes an approximate exchange algorithm for estimation. I fully characterize the convergence properties of the algorithm. prove that the proposed

[16] See for example Jackson and Watts (2002), Blume (1993)

[17] Boucher (2013) shows that homophily has empirical power in a strategic model of network formation. His model also produces a unique equilibrium and allows identification of the relative importance of individual characteristics in determining the shape of the network.

algorithm converges to the correct posterior distribution. The performance of the algorithm is evaluated with artificial data and Add Health data in Section 4. Section 5 concludes. Appendix A contains proof for the theoretical section. Appendix B provides details for the estimation algorithm, and Appendix C show extensions with unobserved heterogeneity.

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) characteristics $X_i = \{X_{i1}, ..., X_{iA}\}$, e.g. gender, wealth, age, location, etc. Let the matrix $X = \{X_1, X_2, ..., X_n\}$ collect the vectors of characteristics for the population and let $\mathcal{X}$ denote 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 entry g_{ij} is equal to 1 if individual i nominates individual j as a friend and 0 otherwise; by convention $g_{ii} = 0$, for any i. The network G is *directed*, i.e. $g_{ij} = 1$ does not necessarily imply $g_{ji} = 1$. This modeling choice reflects the structure of the Add Health data, where friendship nominations are not necessarily mutual.[18]

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).

2.1.1 Preferences

The players have utility functions defined over network realizations $g \in \mathcal{G}$ and population characteristics $X \in \mathcal{X}$, and indexed by parameters $\theta \in \Theta$. The utility of player i from a network g and population attributes $X = (X_1, ..., X_n)$ at parameter θ is given by

$$\begin{aligned} U_i(g, X; \theta) &= \underbrace{\sum_{j=1}^{n} g_{ij} u_{ij}(\theta_u)}_{\text{direct friends}} + \underbrace{\sum_{j=1}^{n} g_{ij} g_{ji} m_{ij}(\theta_m)}_{\text{mutual friends}} \\ &+ \underbrace{\sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{jk} v_{ik}(\theta_v)}_{\text{friends of friends}} + \underbrace{\sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ki} w_{kj}(\theta_w)}_{\text{popularity}} \end{aligned} \tag{1}$$

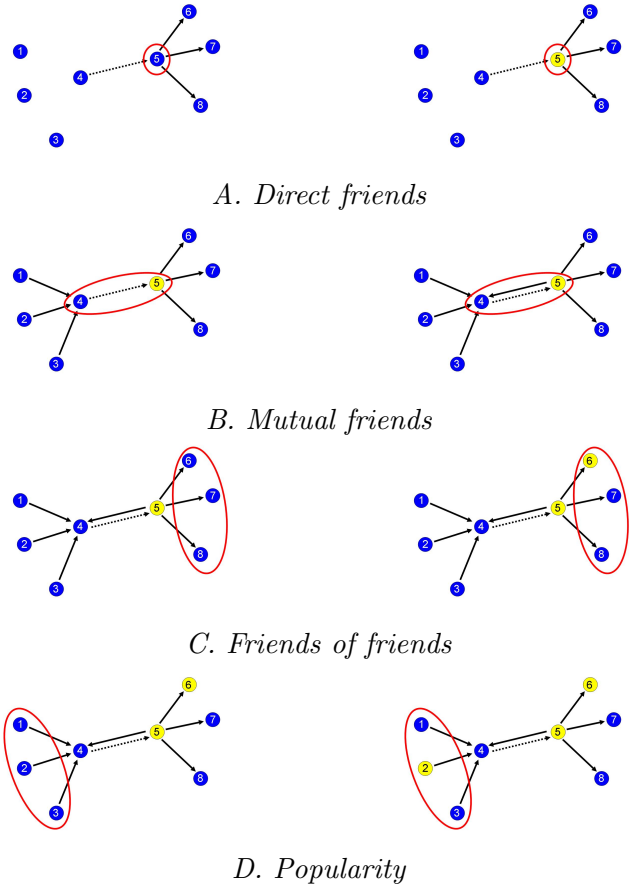
where $u_{ij}(\theta_u) \equiv u(X_i, X_j; \theta_u)$, $m_{ij}(\theta_m) \equiv m(X_i, X_j; \theta_m)$, $v_{ij}(\theta_v) \equiv v(X_i, X_j; \theta_v)$ and $w_{ij}(\theta_w) \equiv w(X_i, X_j; \theta_w)$ are (bounded) real-valued functions of the attributes. The utility of the network is the sum of the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

[18] Some authors refer to this data as *perceived* networks. See Wasserman and Faust (1994) for references.

additional link has four components.

When player i creates a link to agent j, he receives a *direct* net benefit $u_{ij}(\theta_u)$, including

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).

both costs and benefits from the relationship. The net benefit can possibly be negative, e.g. when only homophily enters payoffs of direct links, the net utility $u_{ij}(\theta_u)$ 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. A player can be of blue

or yellow type. The link that player 4 forms to individual 5 provides different direct payoff in the two networks, since the type of 5 is different: blue for the network on the left and yellow for the network on the right. In many models this component is specified as $u_{ij} = b_{ij} - c_{ij}$, where b_{ij} indicates the (gross) benefit 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.

A player receives additional utility $m_{ij}(\theta_m)$ if the link is mutual; a friendship has different value when the other party reciprocates. A student may consider an individual to be a friend, but that person may not. 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).

Players value the composition of friends of friends. When i is deciding whether to befriend j, she observes j's friends and their socioeconomic characteristics. Each of j's friend provides additional utility $v(X_i, X_k; \theta_v)$ 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.[19] 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 decrease in popularity is less pronounced.

2.1.2 Network Formation Process

The process of network formation follows a *stochastic best-response dynamics* (Blume (1993)), generating a Markov chain of networks. The main ingredients of this process are random meetings and utility maximization. The implicit assumption is that meetings are very frequent, and the players can revise their linking strategies often.

Meeting Technology. At the beginning of each period a player i is randomly selected from the population, and he meets individual j, according to a meeting technology. The *meeting process* is a stochastic sequence $m = \{m^t\}_{t=1}^{\infty}$ with support $\mathcal{I} \times \mathcal{I}$. The realizations

[19] A similar assumption is used in De Marti and Zenou (2009) where the agents' cost of linking depends 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.

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.[20]

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

$$\Pr\left(m^t = ij | g^{t-1}, X\right) = \rho\left(g^{t-1}, X_i, X_j\right) \tag{2}$$

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 meeting 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 general formulation includes meeting technologies with a bias for same-type individuals as in Currarini et al. (2009). The simplest example of meeting 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 g_{ij} to maximize his current utility, taking the existing network g^t_{-ij} as given. I assume that the agents do not take into account the effect of their linking strategy on the future evolution of the network. The players have *complete information*, since they can observe the entire network and the individual attributes of all agents.[21] 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 models unobservables that could influence the utility of an additional link, e.g. mood, gossips, fights, etc. Player i links agent j at time t if and only if it is a best response to the current network configuration, i.e. $g^t_{ij} = 1$ if and only if

$$U_i\left(g^t_{ij} = 1, g^{t-1}_{-ij}, X; \theta\right) + \varepsilon_{1t} \geq U_i\left(g^t_{ij} = 0, g^{t-1}_{-ij}, X; \theta\right) + \varepsilon_{0t}. \tag{3}$$

I assume that when the equality holds, the agent plays the status quo.[22] The network formation process generates a Markov chain of networks, with transition probabilities determined by the meeting process and agents' linking choices.

2.2 Equilibrium Analysis

I impose an additional assumption on the functional forms of the utility functions, which provides important equilibrium and identification restrictions. I assume that the utility $m_{ij}(\theta_m)$ obtained from mutual links is symmetric, and that the utility of an indirect link $v_{ij}(\theta_v)$ has the same functional form as the utility from the popularity effect $w_{ij}(\theta_v)$.

[20] Several models incorporate a meeting 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. Christakis et al. (2010) develop a dynamic model, where the sequence of meetings determines which players have the opportunity to form a link in each period.

[21] More precisely, to make a decision about linking, the player needs to observe his in-links and the out-links of his friends.

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

ASSUMPTION 1 ***(Preferences)*** *The preferences satisfy the following restrictions*

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

A direct consequence of the assumption is that the utility function becomes

$$\begin{aligned} U_i\left(g, X; \theta\right) &= \sum_{j=1}^{n} g_{ij} u_{ij}(\theta_u) + \sum_{j=1}^{n} g_{ij} g_{ji} m_{ij}(\theta_m) \\ &+ \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{jk} v_{ik}(\theta_v) + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ki} v_{kj}(\theta_v) \end{aligned} \tag{4}$$

The symmetry in $m_{ij}(\theta_m)$ does not imply that a mutual link between i and j gives both the same utility. If i and j have a mutual link, they receive the same common utility component $(m_{ij}(\theta_m))$ but they may receive different payoffs from direct or indirect links. Two individuals with the same exogenous characteristics $X_i = X_j$ who form a mutual link receive the same $u_{ij}(\theta_u)$ and $m_{ij}(\theta_m)$, but they may have different payoffs from the additional link because of the composition of their friends of friends and their popularity. Therefore, the first part of the assumption is crucial for identification of the utility from indirect links and popularity.

The second part of the assumption imposes an identifying restriction to the externality generated by i when creating a link to j: any individual k that has formed a link to i, has an additional indirect friend, i.e. j, who agent k values by an amount $w\left(X_k, X_j; (\theta_w)\right)$. When $w\left(X_k, X_j; (\theta_v)\right) = v\left(X_k, X_j; (\theta_v)\right)$, 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.[23,24]
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

[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.

[24] The first part of the assumption is a normalization of the utility 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.Similar restrictions are also encountered in spatial econometrics models (Besag, 1974) and in the literature on qualitative response models (Heckman, 1978; Amemiya, 1981)

state of the network are summarized by a ***potential function****,* $Q : \mathcal{G} \times \mathcal{X} \to \mathbb{R}$

$$Q(g, X; \theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} u_{ij}(\theta_u) + \sum_{i=1}^{n} \sum_{j>i}^{n} g_{ij} g_{ji} m_{ij}(\theta_m) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{k=1 \\ k \neq i,j}}^{n} g_{ij} g_{jk} v_{ik}(\theta_v), \tag{5}$$

and the network formation game is a Potential Game.

Proof. See Appendix A ■

The intuition for the result is simple.[25] Under the restrictions of Assumption 1, for any player i and any link g_{ij} we have

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

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, $U_i(g, X; \theta) - U_i(g', X; \theta)$, is exactly equal to the difference of the *potential* function evaluated at the two networks, $Q(g, X; \theta) - Q(g', X; \theta)$. 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 the analysis and the simulations. 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.[26]

To analyze the long run behavior of the model, I impose more structure on the meeting technology.[27]

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

$$\rho(g^{t-1}, X_i, X_j) > 0 \tag{6}$$

The meeting process is such that any player can be chosen and any pair of agents can meet. This assumption guarantees that any equilibrium network can be reached with positive

[25] See Monderer and Shapley (1996) for definitions and properties of potential games.

[26] 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. The computation of all profitable deviations for each player involves $n(n-1)2^{n(n-1)}$ operations: each player has $n-1$ possible deviations, there are n players and a total of $2^{n(n-1)}$ possible network configurations. As it is shown below (Proposition 2), when the game is a potential game, the computation of all Nash equilibria is equivalent to finding the local maxima of the potential function. This corresponds to evalutating the potential function for all the $2^{n(n-1)}$ possible network structures. The latter task involves fewer operations by a factor of $n(n-1)$, thus decreasing the computational burden.

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

probability. For example, a discrete uniform distribution satisfies this assumption.

A Nash equilibrium is a network in which any player has no profitable deviations from his current linking strategy, when randomly selected from the population. We can show that the set of Nash networks correspond to the local maxima of the potential function. Suppose that the current network is a Nash network. As a consequence, if a player deviates from the current linking strategy, he receives less utility.[28] 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.

In the absence of preference shocks, the consequences of assumptions 1 and 2 are that the model will evolve according to a Markov Chain, converging to one of the Nash networks with probability one (see formal details in Appendix A). Suppose a player 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.

I make the following standard 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.*

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

$$\Pr\left(g_{ij}^t = 1 \middle| g_{-ij}^{t-1}, X; \theta\right) = \Pr\left[\varepsilon_{0t} - \varepsilon_{1t} \leq U_i\left(1, g_{-ij}^{t-1}, X; \theta\right) - U_i\left(0, g_{-ij}^{t-1}, X; \theta\right)\right] \tag{7}$$

$$= \frac{\exp\left[u_{ij}(\theta_u) + g_{ji}^{t-1} m_{ij}(\theta_m) + \sum_{k \neq i,j} g_{jk}^{t-1} v_{ik}(\theta_v) + \sum_{k \neq i,j} g_{ki}^{t-1} v_{kj}(\theta_v)\right]}{1 + \exp\left[u_{ij}(\theta_u) + g_{ji}^{t-1} m_{ij}(\theta_m) + \sum_{k \neq i,j} g_{jk}^{t-1} v_{ik}(\theta_v) + \sum_{k \neq i,j} g_{ki}^{t-1} v_{kj}(\theta_v)\right]} \tag{8}$$

Under Assumptions 1-3, the network evolves as a Markov chain with transition probabilities given by the conditional choice probabilities (8) and the probability law of the meeting process m^t.

One can easily show that the sequence $[g^0, g^1, ..., g^t]$ is *irreducible* and *aperiodic*.[29] The following theorem summarizes the main theoretical result.

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

[29] 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* $\pi(g, X; \theta)$
2. *If the meeting probability of* i *and* j *does not depend on the existence of a link between them, i.e.,*

$$\rho\left(g^{t-1}, X_i, X_j\right) = \rho\left(g_{-ij}^{t-1}, X_i, X_j\right). \tag{9}$$

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

$$\pi\left(g, X; \theta\right) = \frac{\exp\left[Q\left(g, X; \theta\right)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X; \theta\right)\right]}, \tag{10}$$

where $Q\left(g, X; \theta\right)$ *is the potential function (5).*

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. Therefore a high proportion of the possible networks generated by the network formation game, will correspond to Nash networks.

The stationary distribution $\pi\left(g, X; \theta\right)$ includes a normalizing constant

$$c\left(\mathcal{G}, X; \theta\right) \equiv \sum_{\omega \in \mathcal{G}} \exp\left[Q\left(\omega, X; \theta\right)\right] \tag{11}$$

that guarantees that (10) is a proper probability distribution. Unfortunately, this normalizing constant greatly complicates estimation, since it cannot be evaluated exactly or approximated with precision. The details about how this problem is circumvented are presented in the empirical strategy section.

2.3 Micro-foundations for Exponential Random Graphs

The Exponential Random Graph model is a statistical model of random network formation, with complex dependencies among links. Exponential random graphs have been successfully

used to fit social network data, providing a useful benchmark for alternative models. However, as any random network formation model, they lack the equilibrium micro-foundations of the strategic literature.[30] The model proposed in this paper nests the ERGM as a special case. Therefore it provides a first attempt of equilibrium micro-foundations to these class of models.

Assume that the utility functions u, m and v depend *linearly* on the vectors of parameters. Define the functions $H : \mathbb{R}^A \times \mathbb{R}^A \to \mathbb{R}$.

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

$$\begin{aligned} u(X_i, X_j, \theta_u) &= = \theta_u' \mathbf{H}_u(X_i, X_j) \\ m(X_i, X_j, \theta_m) &= = \theta_m' \mathbf{H}_m(X_i, X_j) \\ v(X_i, X_j, \theta_v) &= = \theta_v' \mathbf{H}_v(X_i, X_j) \end{aligned}$$

The assumption of linearity does not exclude interactions among different characteristics, for example interactions of race and gender of both individuals. We can consider different specifications, including different sets of variables for direct, mutual and indirect links. Interactions of individual and network-level attributes are also possible.

The main consequence of the linearity assumption is that the stationary equilibrium of the model belongs to the discrete exponential family (Lehman (1983)).

PROPOSITION 2 ***(Exponential Family Likelihood)***
Under Assumptions 1-4, the stationary distribution $\pi(g, X; \theta)$ belongs to the ***exponential family****, i.e., it can be written in the form*

$$\pi(g, X; \theta) = \frac{\exp\left[\theta' \mathbf{t}(g, X)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[\theta' \mathbf{t}(\omega, X)\right]}, \tag{12}$$

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 may include the number of links, the number of whites-to-whites links, the number of male-to-female links and so on.

This likelihood is analogous to the one of exponential random graph models: we can interpret *some specification* of ERGMs as the stationary equilibrium of a strategic game

[30] 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 link occurs depends on the existence of other links. Wasserman and Pattison (1996) generalized the Markov random graphs to more general dependencies, developing the Exponential Random graph models. Snijders (2002) reviews these models and the related estimation techniques.

of network formation, where myopic agents follow a stochastic best response dynamics and utilities are linear functions of the parameters. The payoff specification in the model assumes that individuals receive utility from four components. This results into a ERGM model that includes links, mutual links and 2-mixed-stars. The inclusion of additional utility components will give rise to a different ERGM specification. In general additional components of the utility functions need to be restricted in analogous ways as in Assumption 1.[31] Therefore, not all the ERGM specifications are necessarily micro-founded or compatible with this model. Nonetheless, the assumptions provide a more transparent interpretation of the restrictions imposed by the ERGM to the primitives of the network formation mechanism.

2.4 Additional Utility Components

It is possible to modify the utility function (1) to include additional components. For example, one may be interested in studying preferences that include utility from cyclic triangles effects, i.e. individual i nominates j, j nominates k and k nominates i. The latter can be modeled as a component of the utility τ that varies with the characteristics of the three individuals involved in the relationships, i.e. $\tau(X_i, X_j, X_k; \theta_\tau)$ for all $i, j, k \in \mathcal{I}$. The utility is easily modified as

$$\begin{aligned} U_i(g, X; \theta) &= \sum_{j=1}^{n} g_{ij} u_{ij}(\theta_u) + \sum_{j=1}^{n} g_{ij} g_{ji} m_{ij}(\theta_m) \\ &+ \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}} g_{ki} v_{kj}(\theta_v) + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}} g_{jk} v_{ik}(\theta_v) + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}} g_{jk} g_{ki} \tau_{ijk}(\theta_\tau) \end{aligned}$$

However, to guarantee the existence of a potential function, we need an additional assumption: the function τ is such that $\tau_{ijk}(\theta_\tau) = \tau_{i'j'k'}(\theta_\tau)$ for any i', j', k' permutation of i, j, k. The potential is easily computed as

$$\begin{aligned} Q(g, X; \theta) &= \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} u_{ij}(\theta_u) + \sum_{i=1}^{n} \sum_{j>i} g_{ij} g_{ji} m_{ij}(\theta_m) \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}} g_{jk} v_{ik}(\theta_v) + \frac{1}{3} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k \neq i,j}} g_{jk} g_{ki} \tau_{ijk}(\theta_\tau) \end{aligned}$$

[31] See extensions to triangles below.

If we want to extend the specification and include an utility component that captures the additional utility from friends in common, we can modify the utility as

$$\begin{aligned} U_i(g,X;\theta) &= \sum_{j=1}^{n} g_{ij}u_{ij}(\theta_u) + \sum_{j=1}^{n} g_{ij}g_{ji}m_{ij}(\theta_m) + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k\neq i,j}} g_{jk}v_{ik}(\theta_v) \\ &+ \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k\neq i,j}} g_{ki}v_{kj}(\theta_v) + \sum_{j=1}^{n} g_{ij} \sum_{\substack{k=1 \\ k\neq i,j}} (g_{jk}+g_{kj})(g_{ki}+g_{ik})\tau_{ijk}(\theta_\tau) \end{aligned}$$

and compute the potential accordingly.

In general, it is possible to include additional utility components to (1) as long as we can find restrictions on the payoffs that guarantee the existence of a potential function.

3 Estimation Strategy

3.1 Computational Problem

The structural parameters are estimated using a Bayesian approach. Let $p(\theta)$ be the prior distribution, and let the likelihood function of the observed data (g,X) be the long-run stationary distribution of the model $\pi(g,X,\theta)$. The posterior distribution of θ is

$$p(\theta|g,X) = \frac{\pi(g,X,\theta)\,p(\theta)}{\int_\Theta \pi(g,X,\theta)\,p(\theta)\,d\theta}. \tag{13}$$

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 $p(\theta)$.

The posterior is usually estimated using a Metropolis-Hastings algorithm that avoids direct computation of the integral $\int_\Theta \pi(g,X,\theta)\,p(\theta)\,d\theta$. This algorithm generates a Markov chain of parameters $\{\theta_0,\theta_1,\theta_2,..\}$ whose unique invariant distribution is the posterior (13). 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 $s+1$ the new parameter θ_{s+1} is updated according to

$$\theta_{s+1} = \begin{cases} \theta' & \text{with prob. } \alpha(\theta,\theta') \\ \theta & \text{with prob. } 1-\alpha(\theta,\theta'), \end{cases} \tag{14}$$

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

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

However, this naive version of the Metropolis-Hastings algorithm cannot be used for this model. The acceptance probability $\alpha(\theta, \theta')$ depends on the ratio $c(\mathcal{G}, X, \theta) / c(\mathcal{G}, X, \theta')$, whose exact computation is infeasible even for very small networks. To be concrete, consider a network with $n = 10$ agents. According to (11), $c(\mathcal{G}, X, \theta) = \sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X, \theta)]$: to compute the constant at the current parameter θ we 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.[32]

3.2 Estimation Algorithm

To solve the estimation problem, I propose an approximate version of the *exchange algorithm* (see Murray et al. (2006)). The approximate algorithm uses a double Metropolis-Hastings step to avoid the computation of the normalizing constant $c(\mathcal{G}, X, \theta)$ in the likelihood, as in Liang (2010).[33] Several authors have proposed similar algorithms in the related literature on Exponential Random Graphs Models (ERGM).[34]

In this section I describe the algorithm for a single network, while in the appendix I provide the extension for multiple independent networks.[35] 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.

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 (13) as unique invariant distribution.

[32] 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.

[33] This improvement comes with a possible cost: the algorithm may produce MCMC chains of parameters 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. 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.

[34] 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).

[35] 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.

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.[36]

For a fixed parameter value θ, the algorithm simulates a Markov chain of networks whose unique invariant distribution is (10). 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*

$$g' \sim q_g(g'|g_r) \tag{15}$$

2. *Update the network according to*

$$g_{r+1} = \begin{cases} g' & \text{with prob. } \alpha_{mh}(g_r, g') \\ g_r & \text{with prob. } 1 - \alpha_{mh}(g_r, g') \end{cases} \tag{16}$$

where

$$\alpha_{mh}(g_r, g') = \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g_r, X, \theta)\right]} \frac{q_g(g_r|g')}{q_g(g'|g_r)}\right\} \tag{17}$$

At each iteration of ALGORITHM 1 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 $\alpha_{mh}(g_r, g')$ 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 convergence. Standard results[37] show that the chain generated by the algorithm converges uniformly to the likelihood

[36] The algorithm used in this paper is similar to the Metropolis-Hastings algorithm proposed in Snijders (2002). I also experimented with the Simulated Tempering algorithm proposed in Mele (2011). The latter may be useful when the stationary distribution of the network formation model has more than one mode. While it improves the mixing of the chain, 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 local Metropolis-Hastings updates for networks of moderate size. Therefore I use the latter in this paper.

[37] See Meyn and Tweedie (2009), Levin et al. (2008)

of the model. Diaconis and Stroock (1991) show that for any initial network g_0, the rate of convergence can be bounded as

$$2 \sup_{g \in \mathcal{G}} \left| \pi(g, \theta) - \mathcal{P}_\theta^{(r)}(g|g_0) \right| \leq \sqrt{\frac{1 - \pi(g_0, \theta)}{\pi(g_0, \theta)}} \lambda_*(\theta)^r \tag{18}$$

where $\mathcal{P}_\theta^{(r)}(g|g_0)$ is the r step transition of the algorithm, started at network g_0 and $\lambda_*(\theta) = \max\{\lambda_2(\theta), |\lambda_{2^{n(n-1)}}(\theta)|\}$ is the second highest eigenvalue of ALGORITHM 1's transition matrix.

The bound (18) suggests that the researcher can improve convergence through a careful choice of the initial network and proposal distributions.[38] The choice of the initial network is quite trivial in this model, and it is given by the observed network g. There are two reasons for this choice. First, in the high density region of the posterior the observed network g should have high probability according to the model. This decreases $\sqrt{\frac{1-\pi(g_0,\theta)}{\pi(g_0,\theta)}}$. Second, the result in Lemma 1 in Appendix B shows that this choice guarantees that the *approximate* and the *exact* exchange algorithm have the same acceptance ratio, for any length R of the simulation. Therefore, the proof of convergence to the correct posterior only need to show the convergence of the proposal distribution (see details in Appendix B).

3.2.2 Posterior Simulation

I propose an approximate 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. I instead exploit Lemma 1 (in Appendix B) to decrease the number of simulations. The Lemma implies that my approximate algorithm and the exact algorithm have the same probability of accepting the proposed parameter.[39]

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:

[38] In the previous version of the paper, I used 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 small probability p_{inv}, the sampler proposes a to invert the network matrix, i.e. $g' = \mathbf{1} - g$, and the proposal is accepted with probability $\alpha_{mh}(g, g')$. This move is suggested in Geyer (1992) and Snijders (2002). Snijders (2002) argues that this is particularly useful in case of a bimodal distribution. These proposals directly modify the transition matrix of ALGORITHM 1, decreasing the value of the second highest eigenvalue.

[39] The details require a careful use of the detailed balance condition for the network simulation model. See Appendix B.

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

$$\theta' \sim q_\theta(\cdot|\theta). \tag{19}$$

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

$$g' \sim \mathcal{P}_{\theta'}^{(R)}(g'|g). \tag{20}$$

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

$$\alpha_{ex}(\theta, \theta', g', g) = \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}\right\}. \tag{21}$$

The main advantage of this algorithm is that all quantities in the acceptance ratio (21) can be evaluated: there are no integrals or normalizing constants to compute. This simple modification of the original Metropolis-Hastings scheme makes estimation feasible.

The details of the algorithm, the relative proofs of convergence to the posterior and mixing rates are considered in Appendix B. The algorithm is easily extended to estimate the model using multiple networks using parallelization (see Appendix B). Here I explain intuitively why the sampler works, with the help of Figure 2.

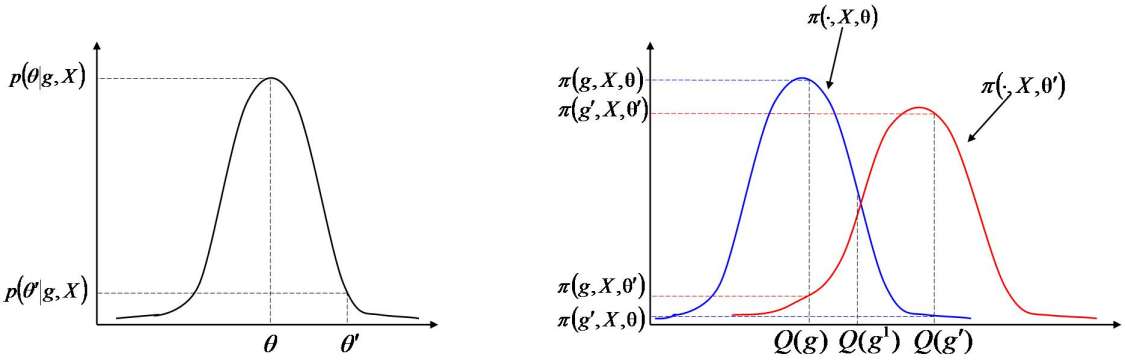
For ease of exposition, assume a relatively flat prior, i.e. $p(\theta)/p(\theta') \simeq 1$. Start the sampler at parameter θ, which has high posterior probability, given the data g. That is, there is good agreement between the data and the parameter, so it is likely that the data are generated from a model with parameter θ. This is displayed on the left panel of Figure 2. 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 Panel B of Figure 2. Let's start the network simulations using parameter θ'. The sequence of simulated networks will sequentially climb the new stationary distribution $\pi(\cdot, X, \theta')$, moving away from the stationary distribution $\pi(\cdot, X, \theta)$. This is indicated in Figure 2 (Panel B) 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. Therefore the ratio

$$\frac{\pi(g', X, \theta)}{\pi(g', X, \theta')} \tag{22}$$

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.

is small. Notice that the the product of the latter ratios

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

is small, and it is contained in (21). As a consequence the acceptance ratio of the exchange algorithm is low and the proposed parameter θ' is very likely to be rejected. If we start the sampler at θ' and propose θ, the update 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. This is the intuition behind the sampler. The formal statement about convergence 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* $\theta_0 \in \Theta$

$$\left\| \widetilde{P}_R^{(s)}(\theta_0, \cdot) - p(\cdot|g, X) \right\|_{TV} \leq \epsilon \tag{23}$$

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

$$\frac{1}{S} \sum_{s=1}^{S} h(\theta_s) \xrightarrow{P} \int_{\Theta} h(\theta) p(\theta|g, X) \, d\theta \tag{24}$$

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.

In general, for a fixed number of network simulations R the samples generated by the algorithm will converge to a posterior that is "close" to the correct posterior. As $R \to \infty$ the algorithm converges to the exact exchange algorithm of Murray et al. (2006), producing exact samples from the posterior distribution. However, an higher value of R would increase the computational cost and result in a higher rejection rate for the proposed parameters. The results in the next section provide some guidance on setting a suitable R, without compromising computational efficiency.

3.3 Convergence and Feasibility

The most expensive step of the approximate exchange algorithm is the network simulation. To gain some insights on the feasibility of estimation, I extend the analysis of Bhamidi et al. (2011) to directed networks and show that the model presented here has good convergence properties, at least in the linear utility specification.

Let's define the function $\varphi(\mu, \theta)$

$$\varphi(\mu, \theta) \equiv \frac{\exp\left[\frac{\psi_1(\theta_{(1)}, X)}{n(n-1)} + 2\frac{\psi_2(\theta_{(2)}, X)}{n(n-1)}\mu\right]}{1 + \exp\left[\frac{\psi_1(\theta_{(1)}, X)}{n(n-1)} + 2\frac{\psi_2(\theta_{(2)}, X)}{n(n-1)}\mu\right]} \tag{25}$$

and its derivative with respect to μ is $\varphi'(\mu, \theta)$. The equation is derived in Appendix through variational approximations for the exponential family (see Wainwright and Jordan (2008),

Jaakkola (2000) for an introduction). The function $\frac{\psi_1(\theta_{(1)},X)}{n(n-1)}$ is a function of the sufficient statistics of the model that involve one link, i.e. direct utility. The function $\frac{\psi_2(\theta_{(2)},X)}{n(n-1)}$ is a function of network statistics that involve interactions of two links, i.e. mutual and indirect utility. Intuitively the mean-field equation (25) is (approximately) the probability of updating a link if the network is generated as an Erdos-Renyi graph with link formation probability μ.

We partition the parameter space according to the following definition.

DEFINITION 1 *A parameter vector θ belongs to the* high temperature *region if the equation $\varphi(\mu,\theta)=\mu$ has a unique solution μ^* with $\varphi'(\mu,\theta)<1$. We write $\theta\in\mathcal{HT}$.*

A parameter vector θ belongs to the low temperature *region if the equation $\varphi(\mu,\theta)=\mu$ has at least two solutions μ^* with $\varphi'(\mu,\theta)<1$. We write $\theta\in\mathcal{LT}$.*

The main convergence result is proven in Bhamidi et al. (2011) and extended to the our model in the following proposition.

PROPOSITION 3 ***(Convergence rates)***
Assume $\frac{\psi_2(\theta_{(2)},X)}{n(n-1)}>0$.

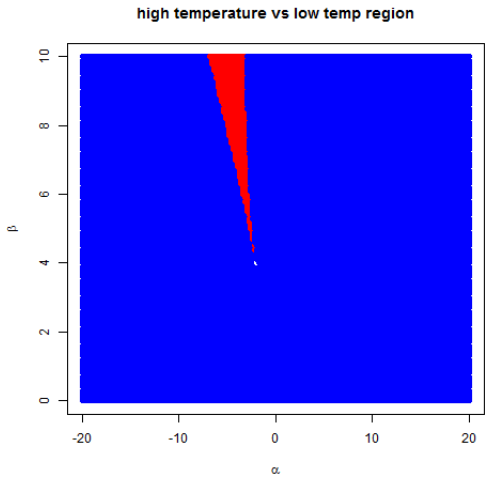
1. *If θ belongs to the* high temperature *region, i.e. $\theta\in\mathcal{HT}$, the chain of networks generated by the model mixes in order $n^2\log n$.*
2. *If θ belongs to the* low temperature *region, i.e. $\theta\in\mathcal{LT}$ then convergence is exponentially slow. This holds for any local dynamics, i.e. a dynamics that updates an $o(n)$ number of links per period.*

Proof. See Bhamidi et al. (2011), Thm. 5 and 6 ■

The intuition for the result is as follows. When the mean-field equation has a unique solution, there is a unique mode: the simulation of the model through local dynamics generates a Markov chain that concentrates around the mode in quadratic time. However, when the mean-field equation has multiple solutions, the stationary distribution has multiple modes and the simulation through a local chain will concentrate around one of the modes. To be able to switch between modes, the simulation needs to be able to make big steps, so a local dynamics is unable to escape from one of the modes.[40]

In summary, sampling with any local sampler is possible for parameters that lie in the high temperature phase, and convergence is in the order of $n^2\log n$ steps; for parameters belonging the low temperature phase, convergence is exponentially slow and the algorithm may be infeasible in practice.

Figure 3: High temperature and low temperature regions



The figure shows the high (blue) and low (red) temperature regions for the model in equation (26), with direct utility and reciprocity. Convergence is exponentially slow in the low temperature region, while mixing is of order $n^2 \log n$ in the high temperature region. The slow convergence are is negligible for this model.

Let's consider a simple example, where players receive utility from direct and reciprocated links only

$$U_i(g, X) = \alpha \sum_{j=1}^{n} g_{ij} + \beta \sum_{j=1}^{n} g_{ij} g_{ji} \tag{26}$$

Assume $\alpha \in \mathbb{R}$ and $\beta > 0$. The mean-field equation for the model is

$$\mu = \frac{\exp\left[\alpha + \beta\mu\right]}{1 + \exp\left[\alpha + \beta\mu\right]} \tag{27}$$

The high and low temperature regions for model (26) are shown in Figure 3, for reasonable parameter values. The figure shows that the region of exponentially slow convergence is extremely small: the model has good convergence properties.

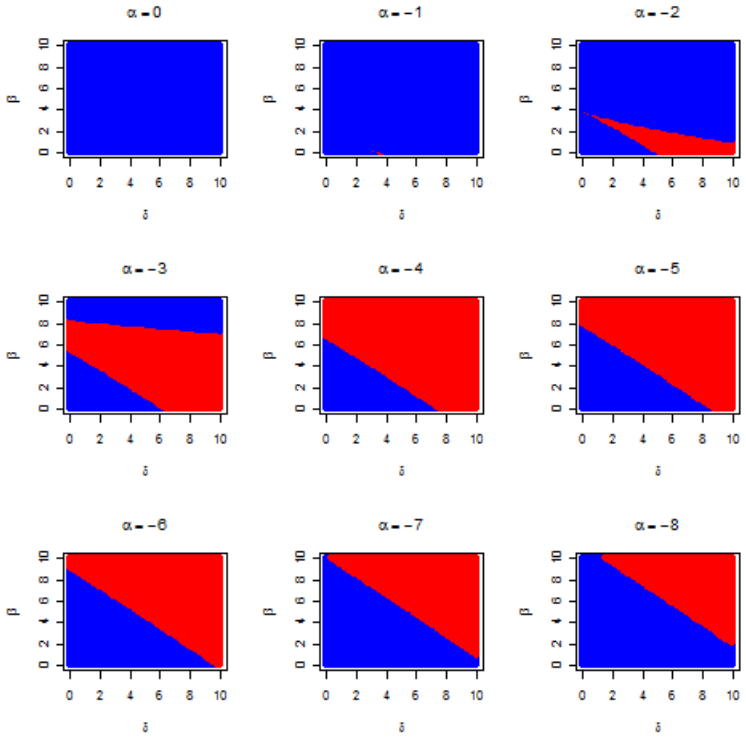
The good convergence properties hold for any model with dependencies up to two links. For example, the model including the indirect links and popularity effects on the utility function has a mean-field equation similar to (27).[41] However, when we introduce additional components of the utility function the convergence properties deteriorate fast. Consider and extended model with inclusion of cyclic triads

$$U_i(g, X) = \alpha \sum_{j=1}^{n} g_{ij} + \beta \sum_{j=1}^{n} g_{ij} g_{ji} + \delta \sum_{j=1}^{n} g_{ij} \sum_{k \neq j,i} g_{jk} g_{ki} \tag{28}$$

[40] Mele (2011) proposed a Simulated Tempering algorithm to partially offset this problem.

[41] See Appendix B for details.

Figure 4: High temperature and low temperature regions, model with cyclic triads



The figure shows the high (blue) and low (red) temperature regions for the model in equation (28) with direct utility, reciprocity and cyclic triads. Convergence is exponentially slow in the low temperature region, while mixing is of order $n^2 \log n$ in the high temperature region. The are of slow convergence is substantially large for this model, and estimation may not be feasible.

The mean-field equation for this specification is

$$\mu = \frac{\exp\left[\alpha + \beta\mu + \delta\mu^2\right]}{1 + \exp\left[\alpha + \beta\mu + \delta\mu^2\right]}$$

and the graph is reported in Figure 4 for several values of α. The low temperature region covers an extended area of the parameter space, suggesting that estimation of such a model with the exchange algorithm has convergence problems or may be infeasible.

For the more general mean-field equation (25) the regions look similar to Figure 3.[42] In summary, the original model presented in (1) has good convergence and mixing properties. Therefore estimation with the approximate exchange algorithm is feasible.[43]

[42]Substitute α with the function ψ_1 and β with the function ψ_2.

[43]However, when the number of players grows large we may need very long runs of the chains to obtain accurate estimates. The algorithm for network simulation would need to be modified drastically to accelerate convergence and update more than $o(n)$ links per iteration. It is possible to use a simulated tempering sampler

4 Results

The performance of the estimation method is tested using artificial data and friendship networks data from Add Health. All the computations with artificial data are performed in a standard desktop Dell Precision T3500 with 2 Intel Xeon CPUs at 3.47GHZ and 24GB of RAM.[44] All the replication files are available at https://jshare.johnshopkins.edu/amele1/bayesnet_replic.zip.

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. Additional details on the Add Health dataset, identification and practical implementation are in Appendix B.

4.1 Artificial Data

Ideally, we want to compare the results of the approximate algorithm with the exact algorithm. This is feasible for a special case, where preferences depend only on direct and mutual links (i.e. excluding friends of friends and popularity effects). For this model, described by equation (26), we can show that the constant is[45]

$$c(\theta) = \left(1 + 2e^{\alpha} + e^{2\alpha+\beta}\right)^{\frac{n(n-1)}{2}}$$

For this model we can compute the exact likelihood and we can therefore perform inference using the *exact* metropolis-hastings sampler. We then compare the results of the exact algorithm with the approximate exchange algorithm.

The results of the simulations are shown in Table 1. The data were generated by parameters $(\alpha, \beta) = (-2.0, 0.5)$. Notice that the parameters lie in the region of fast convergence. The number of network simulations per each proposed parameter are $R = \{1000,$

as in Mele (2011). An alternative is an adaptive version of the exchange algorithm complemented with larger proposed steps of the sampler, as proposed in Atchade and Wang (forthcoming).

[44] Because of the contractual constraints for the use of restricted Add Health data, I have to run the simulations in a Condor cluster, where the data are stored. The cluster processors are slightly slower than the desktop used for the artificial data simulations, but the performance is comparable.

[45] With some algebra

$$\begin{aligned} c(\theta) &= \sum_{g} \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}\alpha + \sum_{i=1}^{n}\sum_{j=i+1}^{n} g_{ij}g_{ji}\beta\right] = \sum_{g} \exp\left[\sum_{i=1}^{n}\sum_{j=i+1}^{n} (g_{ij}+g_{ji})\,\alpha + g_{ij}g_{ji}\beta\right] \\ &= \sum_{g}\prod_{i=1}^{n}\prod_{j=i+1}^{n} \exp\left[(g_{ij}+g_{ji})\,\alpha + g_{ij}g_{ji}\beta\right] = \prod_{i=1}^{n}\prod_{j=i+1}^{n}\sum_{g_{ij}=0}^{1}\sum_{g_{ji}=0}^{1} \exp\left[(g_{ij}+g_{ji})\,\alpha + g_{ij}g_{ji}\beta\right] \\ &= \prod_{i=1}^{n}\prod_{j=i+1}^{n} \left(e^{0} + e^{\alpha} + e^{\alpha} + e^{2\alpha+\beta}\right) = \left(1 + 2e^{\alpha} + e^{2\alpha+\beta}\right)^{\frac{n(n-1)}{2}} \end{aligned}$$

Table 1: Convergence of estimated posteriors, model with two parameters

$n = 100$	Exact Metropolis		R=1000		R=5000		R=10000		R=50000		R=100000					
	α	β	α	β	α	β	α	β	α	β	α	β				
mean	-1.923	0.286	-1.915	0.286	-1.925	0.285	-1.922	0.275	-1.921	0.284	-1.919	0.286				
median	-1.923	0.288	-1.919	0.296	-1.923	0.292	-1.921	0.273	-1.921	0.287	-1.919	0.286				
std. dev.	0.034	0.114	0.105	0.263	0.054	0.141	0.042	0.123	0.034	0.115	0.034	0.111				
mcse	0.000	0.002	0.007	0.033	0.001	0.005	0.001	0.004	0.000	0.003	0.000	0.003				
pctile 2.5%	-1.992	0.058	-2.115	-0.257	-2.034	-0.007	-2.006	0.034	-1.987	0.039	-1.985	0.069				
pctile 97.5%	-1.857	0.506	-1.705	0.767	-1.820	0.553	-1.842	0.514	-1.853	0.505	-1.851	0.512				
KS	NA	NA	0.275	0.205	0.114	0.057	0.066	0.057	0.032	0.015	0.060	0.022				
KL	NA	NA	0.041	0.027	0.013	0.186	0.039	0.075	0.040	0.062	0.006	0.088				

$n = 200$	Exact Metropolis		R=1000		R=5000		R=10000		R=50000		R=100000		R=1mil		R=10mil	
	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β
mean	-1.988	0.463	-1.975	0.463	-1.964	0.463	-1.979	0.465	-1.989	0.455	-1.988	0.463	-1.988	0.466	-1.987	0.459
median	-1.989	0.467	-1.974	0.509	-1.968	0.468	-1.978	0.465	-1.989	0.454	-1.989	0.464	-1.988	0.467	-1.987	0.457
std. dev.	0.017	0.061	0.048	0.275	0.042	0.113	0.033	0.073	0.019	0.053	0.017	0.059	0.017	0.057	0.016	0.055
mcse	0	0.003	0.002	0.075	0.002	0.012	0.001	0.005	0	0.002	0	0.003	0	0.003	0	0.002
pctile 2.5%	-2.021	0.335	-2.071	-0.21	-2.044	0.186	-2.042	0.32	-2.024	0.353	-2.024	0.339	-2.024	0.358	-2.017	0.354
pctile 97.5%	-1.954	0.572	-1.89	0.889	-1.872	0.687	-1.921	0.614	-1.949	0.56	-1.955	0.571	-1.955	0.582	-1.955	0.577
KS	NA	NA	0.343	0.34	0.381	0.135	0.25	0.057	0.067	0.105	0.015	0.039	0.039	0.045	0.062	0.086
KL	NA	NA	0.1	0.178	0.105	0.079	0.129	0.099	0.05	0.05	0.041	0.058	0.092	0.044	0.173	0.234
time	0.124s		14.539s		21.808s		30.451s		100.761s		193.722s		1762.202s		17370.945s	

$n = 500$	Exact Metropolis		R=1000		R=5000		R=10000		R=50000		R=100000		R=1mil		R=10mil	
	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β
mean	-2.018	0.551	-1.941	0.337	-2.014	0.562	-2.017	0.561	-2.017	0.552	-2.018	0.552	-2.018	0.55	-2.018	0.55
median	-2.018	0.552	-1.922	0.369	-2.012	0.562	-2.019	0.562	-2.016	0.553	-2.018	0.552	-2.018	0.551	-2.018	0.55
std. dev.	0.007	0.024	0.071	0.218	0.045	0.107	0.036	0.074	0.016	0.036	0.012	0.028	0.007	0.022	0.007	0.022
mcse	0	0	0.005	0.047	0.002	0.011	0.001	0.005	0	0.001	0	0	0	0	0	0
pctile 2.5%	-2.032	0.501	-2.074	-0.106	-2.105	0.335	-2.085	0.424	-2.05	0.479	-2.041	0.497	-2.032	0.508	-2.031	0.507
pctile 97.5%	-2.004	0.596	-1.838	0.666	-1.931	0.755	-1.942	0.707	-1.988	0.621	-1.994	0.606	-2.004	0.596	-2.005	0.592
KS	NA	NA	0.743	0.703	0.408	0.363	0.341	0.31	0.229	0.117	0.107	0.066	0.027	0.034	0.033	0.032
KL	NA	NA	0.466	0.196	0.121	0.081	0.081	0.019	0.026	0.009	0.02	0.008	0.061	0.036	0.049	0.041
time	0.187s		87.344s		95.831s		105.955s		181.413s		275.357s		2010.322s		19319.663s	

$n = 1000$	Exact Metropolis		R=1000		R=5000		R=10000		R=50000		R=100000		R=1mil		R=10mil	
	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β
mean	-2.001	0.481	-1.986	0.456	-1.974	0.459	-1.995	0.486	-1.999	0.479	-2.001	0.479	-2.001	0.481	-2.002	0.481
median	-2.001	0.48	-1.991	0.501	-1.979	0.461	-1.993	0.479	-2.001	0.48	-2.001	0.479	-2.001	0.48	-2.002	0.482
std. dev.	0.003	0.011	0.081	0.247	0.05	0.091	0.031	0.07	0.017	0.037	0.011	0.026	0.004	0.012	0.003	0.012
mcse	0	0	0.007	0.06	0.002	0.007	0.001	0.004	0	0.001	0	0	0	0	0	0
pctile 2.5%	-2.008	0.459	-2.148	-0.007	-2.047	0.284	-2.057	0.361	-2.029	0.404	-2.024	0.425	-2.01	0.457	-2.008	0.459
pctile 97.5%	-1.995	0.503	-1.835	0.813	-1.825	0.642	-1.938	0.64	-1.966	0.55	-1.979	0.529	-1.993	0.505	-1.995	0.504
KS	NA	NA	0.506	0.484	0.63	0.47	0.502	0.355	0.351	0.268	0.271	0.216	0.078	0.018	0.027	0.036
KL	NA	NA	0.3	0.261	0.528	0.041	0.297	0.083	0.45	0.21	0.137	0.047	0.021	0.031	0.034	0.014
time	0.234s		364.761s		371.563s		381.172s		459.578s		556.330s		2304.228s		19730.772	

5000, 10000, 50000, 100000, 1000000, 10000000}. We run each algorithm for $S = 10000$ parameters iterations, and we use the output to measure the Kolmogorov-Smirnov distance between the posterior estimated with the exact metropolis sampler $p(\theta|g, X)$ and the posterior estimated with the approximated algorithm with R network simulations $p_R(\theta|g, X)$

$$KS\left(p_R(\theta_i|g, X), p(\theta_i|g, X)\right) = \sup_{\theta_i \in \Theta_i} \left| \int_{-\infty}^{\theta_i} p_R(\theta_i|g, X) - \int_{-\infty}^{\theta_i} p(\theta_i|g, X) \right|$$

We also compute the Kullback-Leibler divergence

$$KL\left(p_R(\theta_i|g, X), p(\theta_i|g, X)\right) = \int_{\Theta_i} \log \left[\frac{p_R(\theta_i|g, X)}{p(\theta_i|g, X)} \right] p_R(\theta_i|g, X) d\theta_i$$

The table reports posterior mean, median, standard deviation, Monte Carlo standard errors for the posterior mean (mcse), 95% credibility intervals, Kolmogorov-Smirnov statistics, Kullback-Leibler divergence and time for computation.

The exact metropolis is reported in the first column of the table. The approximate exchange algorithm works very well for small to moderate networks. For a small network with $n = 100$ players, a reasonable degree of accuracy can be reached with as low as $R = 5000$ network simulations per parameter. This is far below the $n^2 \log n = 46052$ suggested by the theoretical convergence result. The latter seems to be a general pattern: for each size of the network a good level of accuracy is obtained by using less than $n^2 \log n$ iterations.

If the researcher is interested in a Bayesian point estimate, the number of simulations required to obtain an accurate estimate of the posterior means is quite small. However, a small R translates into overestimated variability of the posterior. If the researcher is interested in fully Bayesian inference, the number of simulations should be high.

Simulations from over-dispersed starting values converge to the same posterior distribution. Convergence is quite fast to the high density region of the posterior.[46] Computations can be faster if we embed sparse matrix algebra routines in the codes. The results in Table 1 are obtained with codes that do no use sparse matrix algebra, thus representing a worst case scenario in computational time.

4.2 Add Health data

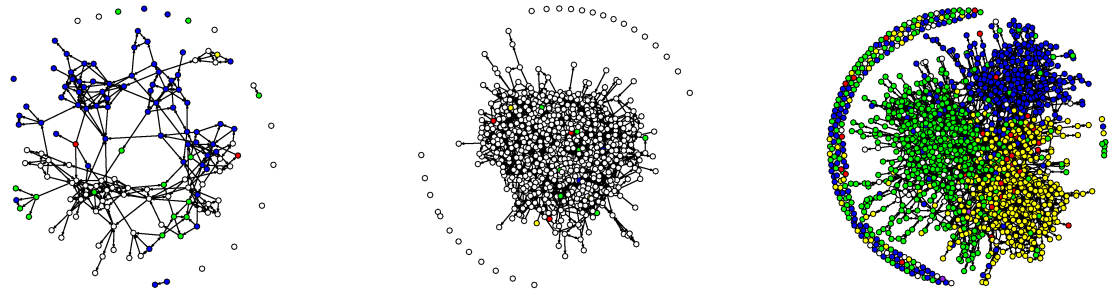
The *National Longitudinal Study of Adolescent Health* (Add Health) is a dataset containing information on a nationally representative sample of US schools. I use data from the first wave of the survey, collected in 1994 when the 90118 participants were entering grades 7-12.[47] Each student responded to an *in-school* questionnaire, and a subsample of 20745 was given

[46] This result is common with the class of exchange algorithms. See Caimo and Friel (2010), Atchade and Wang (forthcoming) for examples.

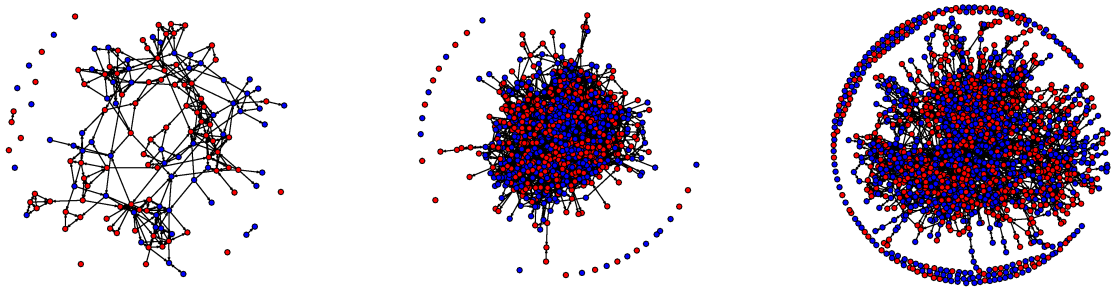
[47] More details about the sampling design and the representativeness are contained in Moody (2001) and the Add Health website http://www.cpc.unc.edu/projects/addhealth/projects/addhealth

Figure 5: School friendship networks data

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



A. School 28 B. School 58 C. School 77



blue = Males; red = Females

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.

an *in-home* interview to collect more detailed information about behaviors, characteristics and health status.

In this paper I use data only from 3 schools to explain how the estimation works.[48] I use schools 28, 58 and 77 from the saturated sample.[49] The *in-school* questionnaire collects the social network of each participant. Each student was given a school roster and was

[48] A previous version of the paper contains estimates for a more complex model, with several additional controls (income, physical appearance, personality) and data from multiple schools, using a parallel implementation of the approximate exchange algorithm. The estimated models are used to perform counterfactual analysis and study the effect of desegregation programs in the US.

[49] The saturated sample is a group of 16 schools for which the Add Health staff was able to collect all informations. These data have no missing values.

asked to identify up to five male and five female friends.[50] 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 necessarily imply that Jim nominates Paul.[51] The model developed in this paper takes this feature of the data into account. The friendship network of the three schools are shown in Figure 5. School 28 is relatively small, with 150 students, and a relatively heterogeneous racial composition. School 58 is larger (811 students) and racially homogeneous. School 77 is a large school with 1664 students and heterogeneous racial composition.

I estimate a model with 18 parameters for school 28. The results are contained in Table 2. I report the posterior mean, standard deviation and 5th, 50th and 95th quantiles. The marginal posteriors are shown in Figure 6. This specification studies homophily in direct, mutual and indirect utility.

Table 2: School 28, structural estimates

	mean	std. dev	5 pctile	median	95 pctile
A. Direct Utility (u_{ij})					
constant	-5.0118	0.2972	-5.5157	-5.0073	-4.5212
same gender	0.2613	0.1842	-0.0371	0.2607	0.5651
same grade	1.0199	0.1933	0.7037	1.0202	1.3401
white-white	0.5694	0.3053	0.0660	0.5653	1.0779
black-black	0.8931	0.3076	0.3931	0.8917	1.4045
hisp-hisp	2.0850	0.5844	1.1295	2.0882	3.0351
B. Mutual Utility (m_{ij})					
constant	2.3907	0.6053	1.3846	2.3996	3.3626
same gender	1.3834	0.4720	0.6372	1.3660	2.2014
same grade	-0.7662	0.4363	-1.4726	-0.7680	-0.0601
white-white	-0.3113	0.5982	-1.2830	-0.3169	0.6750
black-black	-0.0791	0.5864	-1.0551	-0.0784	0.8645
hisp-hisp	-1.4928	1.1213	-3.3859	-1.4649	0.3130
C. Indirect Utility (v_{ij})					
constant	-0.3077	0.0797	-0.4410	-0.3067	-0.1780
same gender	-0.0653	0.0678	-0.1774	-0.0651	0.0436
same grade	0.3010	0.0544	0.2127	0.3002	0.3903
white-white	0.3478	0.0798	0.2173	0.3486	0.4782
black-black	0.2620	0.0758	0.1375	0.2633	0.3839
hisp-hisp	0.0317	0.3800	-0.6400	0.0625	0.6027

Panel A of Table 2 shows the estimates for the direct utility component. We have evidence of homophily for race, gender and grade. The marginal posteriors should be interpreted as marginal utilities. For example, the variable white-white, indicates the marginal utility of a white student forming a link to a white student, *other things being equal*.

The estimates for mutual utility in Panel B show that a reciprocated link provides positive

[50] One can think that this limit could bias the friendship data, but only 3% of the students nominated 10 friends (Moody, 2001). Moreover, the estimation routine could be easily extended to deal with missing links, as reported in Appendix C.

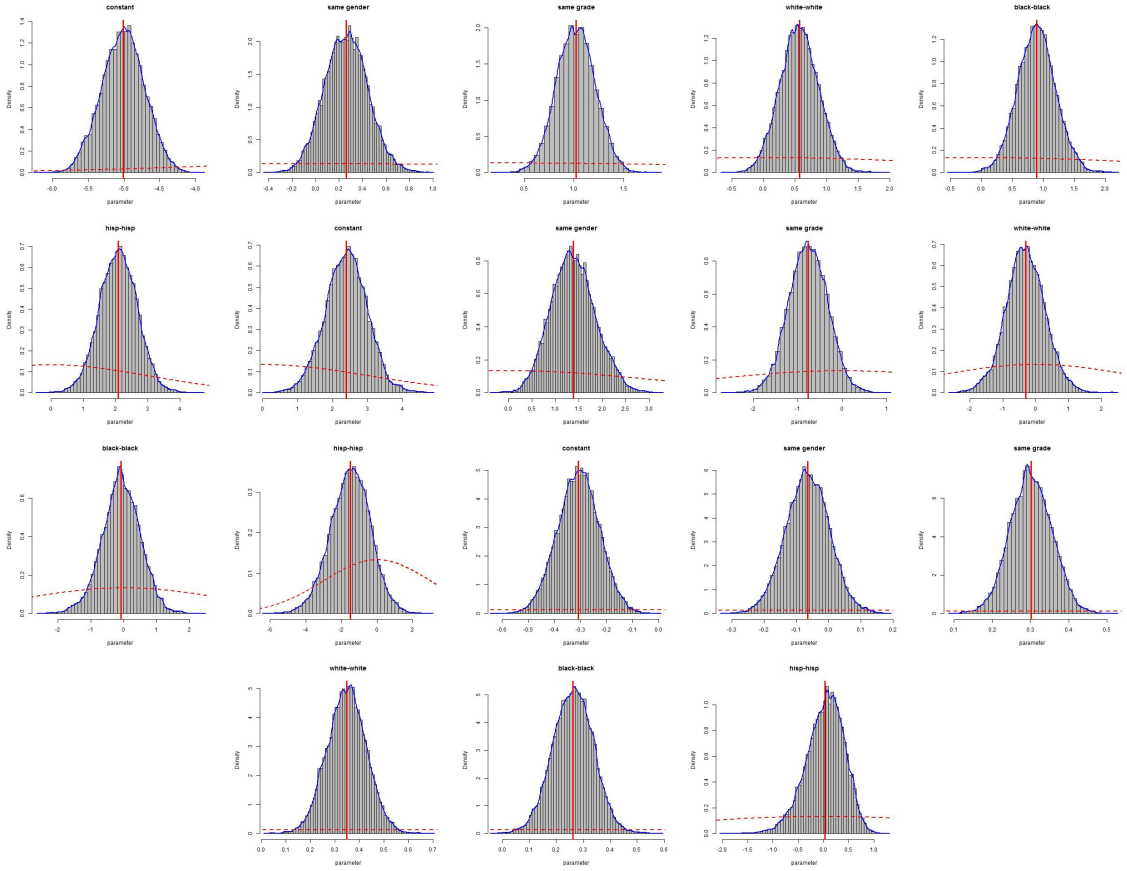
[51] Some authors do not take into account this feature of the data and they recode the friendships as mutual: if a student nominates another one, the opposite nomination is also assumed.

utility. This is reinforced if the reciprocation comes from a student of the same gender. However, a mutual link from a person of the same grade provides negative marginal utility on average. The race coefficients are negative, but their marginal posterior is highly variable.

The estimates for indirect and popularity effects show that students prefer friendships with individuals with fewer friends, probably for a congestion effect. There is evidence of homophily for grade and race, but not for hispanics.

The estimates for school 58 are contained in Table 3. The school contains 98% white

Figure 6: Posterior Distribution, School 28



Estimated posterior distribution for the full structural model. Each graph shows the histogram of the simulation output. The vertical red line indicates the posterior mean, while the red dotted line is the prior distribution. The estimates are obtained with a sample of 50000 parameter simulations and 50000 network simulations for each proposed parameter. I use simulations from 5 over-dispersed chains, discarding the first 10000 iterations for burn-in.

students. In this very parsimonious specification, I study the differential homophily for gender (the omitted category is female-female). Males and females have the same propensity to form links to same gender students. However, guys have a slightly higher propensity to

form links to girls than girls have to form links to guys. Reciprocated links increase utility, but there is congestion in indirect links.

The last set of estimates, for school 77, is shown in Table 4. This is a large school

Table 3: School 58, structural estimates

	mean	std. dev	5 pctile	median	95 pctile
constant	-6.0332	0.1795	-6.3338	-6.0340	-5.7381
male-male	-0.0426	0.1489	-0.2867	-0.0438	0.1983
male-female	-0.3216	0.1658	-0.5985	-0.3217	-0.0438
female-male	-0.4749	0.1740	-0.7665	-0.4762	-0.1909
same grade	1.4308	0.1297	1.2157	1.4289	1.6454
mutual	4.8094	0.1353	4.5902	4.8108	5.0283
indirect	-0.0180	0.0151	-0.0426	-0.0176	0.0067

with 1664 students and a very heterogeneous population. There is evidence of homophily for gender, grade and race (excluding whites). Reciprocated links and indirect links increase utility levels.

These results confirm previous evidence on homophily using the same data.[52] In terms

Table 4: School 77, structural estimates

	mean	std. dev	5 pctile	median	95 pctile
constant	-8.1694	0.2892	-8.6482	-8.1705	-7.7038
same gender	0.4405	0.2164	0.0907	0.4361	0.8028
same grade	1.1318	0.2303	0.7560	1.1316	1.5135
white-white	0.5291	1.1812	-1.6677	0.6548	2.2397
black-black	0.8497	0.3424	0.2732	0.8539	1.4005
hisp-hisp	0.7372	0.2485	0.3244	0.7426	1.1345
mutual	5.5428	0.2667	5.0978	5.5438	5.9817
indirect	0.0356	0.0331	-0.0183	0.0354	0.0894

of computational feasibility, the estimates for school 58 were obtained with 50000 parameter simulations for each of 5 over-dispersed chains. I run the network simulations for 100000 steps for each proposed parameter. The over-dispersed starting values were obtained from a burn-in of 100000 simulations, used also to optimize the proposal distribution. The estimates in Table 3 are obtained in less than 15 hours of CPU time. The same number of simulations is used for Table 4, with a CPU time of about 100 hours.[53]

[52] See Currarini et al. (2010) and Moody (2001) for example.

[53] A reason for this drastic increase in the computational time is that the codes do not use any sparse matrix algebra routines, which would speed up computations especially for these large networks. Therefore, these computation times are upper bounds and can be improved.

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, my algorithm generates samples from the posterior distribution and avoids the evaluation of the likelihood. First, a new parameter is proposed according to a Metropolis-Hastings step. Second, another Metropolis-Hastings sampler simulates a network from the stationary distribution of model.

I prove that the algorithm converges to the correct posterior distribution. However, the network simulation step is computationally very demanding and may require an infeasible number of simulations to mix well.[54] Using mean-field approximations for the exponential family, I show that the network simulation algorithm convergence is of order $n^2 \log(n)$ for most parameter vectors. In other words, the area of exponentially slow mixing is negligible. The reason is that the model incorporates link dependencies up to two links, which satisfy the conditions for fast mixing.[55]

The estimation strategy is evaluated using artificial data and friendship network data from the Add Health dataset. The performance of the algorithm is good and convergence to the posterior distribution is shown to be fast for practical implementations. The models estimated using Add Health data confirm the presence of homophily in US high schools.

The model can be easily extended to incorporate unobserved heterogeneity in individual quality and preferences. The Bayesian estimation strategy can be adapted to estimate models with missing links, using data augmentation techniques. These improvements come with a substantial increase in the computational burden, but also provide a more realistic model. Both these extensions are reviewed and explained in Appendix C.

The methodology introduced in this work can be used in different settings. Models of social interactions with sequential moves as in Nakajima (2007) and Badev (2013) share the same simple equilibrium characterization presented in this work. In these models individuals interact in a network and their actions are optimally chosen given the action of their neighbors. The estimation techniques developed here are easily adapted to these settings.[56]

[54] See Bhamidi et al. (2011) and Diaconis and Chatterjee (2011) for more detail.

[55] However, if the model were to be extended to include higher order dependencies, the region of exponentially slow mixing would be larger and the estimation could become computationally infeasible for some specifications.

[56] In principle, several models that admit a characterization as potential games could be estimated using the algorithm proposed in this paper.

The methodology can be applied to the class of autologistic models in spatial econometrics.[57] The latter are models for spatial binary data that explicitly model the spatial dependence among variables. The likelihood of these models has the same exponential form with normalizing constant derived in this paper, 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 perspective.

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.

Amir, Eyal, Wen Pu and Dorothy Espelage (2012), Approximating partition functions for exponential-family random graph models, NIPS Conference 2012.

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

Atchade, Yves and Jing Wang (forthcoming), 'Bayesian inference of exponential random graph models for large social networks', *Communications in Statistics - Simulation and Computation* .

Athey, Susan and Guido Imbens (2007), 'Discrete choice models with multiple unobserved choice characteristics', *International Economic Review* **48**(4), 1159–1192.

Badev, Anton (2013), Discrete games in endogenous networks: Theory and policy.

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 (2011), 'Mixing time of exponential random graphs', *The Annals of Applied Probability* **21**(6), 2146–2170.

Bishop, Christopher (2006), *Pattern recognition and machine learning*, Springer, New York.

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

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

Boucher, Vincent (2013), Structural homophily.

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

Chandrasekhar, Arun and Matthew Jackson (2012), Tractable and consistent exponential random graph models.

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 (2010), The network structure of mutual support: Evidence from rural tanzania, PSE Working Papers 2008-74, PSE (Ecole normale superieure).

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.

DePaula, Aureo, Seth Richards-Shubik and Elie Tamer (2011), Inference approaches with network data.

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.

Fernandez-Villaverde, Jesus, Pablo Guerron-Quintana and Juan F. Rubio-Ramirez (2010), *Handbook of Applied Bayesian Analysis*, Oxford University Press, chapter The New Macroeconometrics: a Bayesian Approach.

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.

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.

Goldsmith-Pinkham, Paul and Guido W. Imbens (forthcoming), 'Social networks and the identification of peer effects', *Journal of Business and Economic Statistics* .

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.

Hsieh, Chih-Sheng and Lung-Fei Lee (2012), A structural modeling approach for network formation and social interactions with applications to students' friendship choices and selectivity on activities.

Jaakkola, Tommi (2000), *Advanced mean field methods: theory and practice*, MIT PRESS, chapter Tutorial on Variational Approximation Methods.

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.

Koskinen, Johan (2004), Bayesian analysis of exponential random graphs - estimation of parameters and model selection, Research report 2004:2, Department of Statistics, Stockholm University.

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.

Mele, Angelo (2011), 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.

Miyauchi, Yuhei (2012), Structural estimation of a pairwise stable network formation with nonnegative externality.

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.

Norets, Andriy (2009), 'Inference in dynamic discrete choice models with serially correlated unobserved state variables', *Econometrica* **77**(5), 1665–1682.

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

Rossi, P., R. McCulloch and G. Allenby (1996), 'The value of purchase history data in target marketing', *Marketing Science* **15**(4), 321–340.

Sheng, Shuyang (2012), Identification and estimation of network formation games.

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.

Wainwright, M.J. and M.I. Jordan (2008), 'Graphical models, exponential families, and variational inference', *Foundations and Trends@ in Machine Learning* **1**(1-2), 1–305.

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.[58] A simple computation shows

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

that, for any ij

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

therefore Q is the potential of the network formation game.

Proof of Proposition 2
The proof consists of showing that $Q\left(g,X\right)$ can be written in the form $\theta'\mathbf{t}\left(g,X\right)$. 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 $t_{up}\left(g,X\right)\equiv\sum_i\sum_j g_{ij}H_{up}\left(X_i,X_j\right)$, $\theta_u=\left(\theta_{u1},...,\theta_{uP}\right)'$ and $\mathbf{t}_u\left(g,X\right)=\left(t_{u1}\left(g,X\right),...,t_{uP}\left(g,X\right)\right)'$. Analogously define $\theta_m=\left(\theta_{m1},\theta_{m2},...,\theta_{mL}\right)'$ and $\mathbf{t}_m\left(g,X\right)=\left(t_{m1}\left(g,X\right),t_{m2}\left(g,X\right),...,t_{mL}\left(g,X\right)\right)'$ and $\theta_v=\left(\theta_{v1},\theta_{v2},...,\theta_{vS}\right)'$ and $\mathbf{t}_v\left(g,X\right)=\left(t_{v1}\left(g,X\right),t_{v2}\left(g,X\right),...,t_{vS}\left(g,X\right)\right)'$. It follows that

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

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

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

Model without preference shocks: characterization of Nash networks

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.

$$\mathcal{N}(g) \equiv \{g' : g' = (g'_{ij}, g_{-ij}), \text{ for all } g'_{ij} \neq g_{ij}, \text{ for all } i, j \in \mathcal{I}\}. \tag{29}$$

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 4 ***(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*

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.*

$$\mathcal{NE}(\mathcal{G}, X, U) = \left\{ g^* : g^* = \arg \max_{g \in \mathcal{N}(g^*)} Q(g, X) \right\} \tag{30}$$

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. 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

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

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

$$= \sum_i \sum_j \rho_{ij} = 1.$$

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 $[g^0, g^1, ...]$ 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

$$P_{gg'}\pi_g = P_{g'g}\pi_{g'} \tag{31}$$

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 (10) 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.

B Computational Details

B.1 Network Simulation

The algorithm used to simulate the network (**ALGORITHM 1**) produces samples from the stationary equilibrium of the model.

1. The network simulation algorithm satisfies the detailed balance condition for the stationary distribution 10. 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 used in the text was derived by Diaconis and Stroock (1991), for reversible finite chains.

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,

$$\mathcal{P}_{\theta'}^{(R)}(g'|g) = \mathcal{P}_{\theta'}(g^1|g)\mathcal{P}_{\theta'}(g^2|g^1)\cdots\mathcal{P}_{\theta'}(g'|g^{R-1}), \tag{32}$$

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 $\pi(\cdot, X, \theta')$ using a Metropolis-Hastings algorithm starting at the network g observed in the data. Then*

$$\frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)} = \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]} \tag{33}$$

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 (32). 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 $\pi(g, X, \theta')$, therefore we have

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

It follows that

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

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 (33) 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 (13).

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'*

$$\theta' \sim q_\theta(\cdot|\theta) \tag{34}$$

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

$$g' \sim \pi\left(\cdot|X, \theta'\right) \tag{35}$$

3. *Compute the acceptance ratio*

$$\begin{aligned} \alpha_{ex}\left(\theta, \theta', g', g\right) &= \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p\left(\theta'\right)}{p\left(\theta\right)} \frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]} \frac{c(\theta)c(\theta')}{c(\theta)c(\theta')}\right\} \\ &= \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p\left(\theta'\right)}{p\left(\theta\right)} \frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}\right\} \end{aligned} \tag{36}$$

4. *Update the parameter according to*

$$\theta_{t+1} = \begin{cases} \theta' & \textit{with prob. } \alpha_{ex}\left(\theta, \theta', g', g\right) \\ \theta & \textit{with prob. } 1 - \alpha_{ex}\left(\theta, \theta', g', g\right) \end{cases} \tag{37}$$

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

$$\widetilde{\alpha}_{ex}\left(\theta, \theta', g', g\right) = \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p\left(\theta'\right)}{p\left(\theta\right)} \frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)} \frac{\mathcal{P}_{\theta'}^{(R)}(g|g')}{\mathcal{P}_{\theta'}^{(R)}(g'|g)}\right\} \tag{38}$$

$$= \min\left\{1, \frac{\exp\left[Q(g', X, \theta)\right]}{\exp\left[Q(g, X, \theta)\right]} \frac{p\left(\theta'\right)}{p\left(\theta\right)} \frac{q_\theta\left(\theta|\theta'\right)}{q_\theta\left(\theta'|\theta\right)} \frac{\exp\left[Q(g, X, \theta')\right]}{\exp\left[Q(g', X, \theta')\right]}\right\} \tag{39}$$

$$= \alpha_{ex}\left(\theta, \theta', g', g\right) \tag{40}$$

This result implies that to prove the convergence of the approximate algorithm to the exact algorithm, there is no need to prove convergence of $\widetilde{\alpha}_{ex}\left(\theta, \theta', g', g\right)$ to $\alpha_{ex}\left(\theta, \theta', g', g\right)$. 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 $Q\left(d\vartheta|\theta\right) = q_{\theta}\left(\vartheta|\theta\right)\nu\left(d\vartheta\right)$. The transition kernel of the exact exchange algorithm can be written as

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

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

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

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

$$\eta\left(\theta\right) := 1 - \int_{\Theta}\left[\sum_{g'\in\mathcal{G}} \mathcal{P}_{\vartheta}^{(R)}\left(g'|g\right) \alpha_{ex}\left(\theta, \vartheta, g', g\right)\right] Q\left(\theta, d\vartheta\right) \tag{41}$$

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

$$P\left[\theta, \vartheta|g, X\right] p\left(\theta|g, X\right) = \Pr\left[\theta|\vartheta, g, X\right] p\left(\vartheta|g, X\right) \tag{42}$$

The detailed balance condition is sufficient condition for the Markov chain generated by the algorithm to have stationary distribution the posterior (13) (for details see Robert and Casella (2005) or Gelman et al. (2003)).

LEMMA 2 *The exchange algorithm produces a Markov chain with invariant distribution (13).*

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

$$\Pr[\vartheta|\theta,g,X] = q_\theta(\vartheta|\theta)\frac{\exp[Q(g',X,\vartheta)]}{c(\mathcal{G},X,\vartheta)}\alpha_{ex}(\theta,\vartheta,g',g)\,. \tag{43}$$

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 (42) 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 $\vartheta \in \Theta$

$$\frac{\mathcal{P}_{\vartheta}^{(R)}(g'|g)}{\pi(g',\vartheta)} > 0 \quad \textit{for any } g' \in \mathcal{G} \tag{44}$$

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*

$$\lim_{s\to\infty} \left\| \widetilde{P}_R^{(s)}(\theta_0,\cdot) - \widetilde{p}(\theta) \right\|_{TV} = 0 \tag{45}$$

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.[59]

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* $R > R_0$,

$$\left|\widetilde{P}_R\phi(\theta) - P\phi(\theta)\right| < 2\epsilon \tag{46}$$

Proof. The transition of the exchange algorithm is

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

while the transition kernel for the approximate algorithm is

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

and therefore the difference is

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

Consider the quantity

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

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

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

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

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

$$|S| \leq |2S_0| = 2\epsilon \tag{47}$$

■

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.[60] For any function $\phi : \Theta \to [-1, 1]$ we have

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

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

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 $s \geq s(\theta_0, \varepsilon)$

$$\left| P^{(s)} \phi(\theta_0) - p(\phi) \right| \leq \varepsilon \tag{48}$$

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)

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

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 $R > R_0(\theta_0, \varepsilon)$

$$\left| \widetilde{P}_R^{(s_0)} \phi(\theta_0) - P^{(s_0)} \phi(\theta_0) \right| \leq 2 s_0 \varepsilon \tag{49}$$

this implies

$$\left| \widetilde{P}_R^{(s)} \phi(\theta_0) - p(\phi) \right| \leq (2 s_0 + 1) \varepsilon \tag{50}$$

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 and Feasibility of Network Simulation

The results in Bhamidi et al. (2011) are based on an equation that is derived from a mean-field variational approximation of the exponential family.[61] The likelihood for the model developed in this paper, in the special case of linear utilities, belongs to the exponential family and we can use a mean-field approximation to derive the corresponding equation for the directed network case. All the results of Bhamidi et al. (2011) about convergence rates follow with this simple modification.

[61] An introduction to variational methods is contained in Wainwright and Jordan (2008) and Bishop (2006). Variational approximations are a set of deterministic approximation methods that transform the problem of computing the intractable normalizing constant into a maximization problem that can be solved iteratively. The method has several advantages with respect to simulation methods based on MCMC: 1) it converges fast to the final approximation; 2) convergence can be checked with a single scalar, instead of monitoring all the parameters; 3) the procedure can be understood as minimizing the Kullback-Leibler divergence between the true distribution and the approximating distribution. The main limitation is that the error of approximation is fixed and cannot be made smaller by increasing the number of iterations: this is the main difference between deterministic approximations and Monte Carlo method, where the error can be decreased by increasing the number of simulations.

Let's consider the exponential family distribution of the model, where the sufficient statistics are suitably re-scaled.[62] We will introduce some additional notation with respect to the main text. Let v_p denote the number of vertices involved in the sufficient statistics $\phi_p(g, X)$ and let e_p be the number of the edges involved in the sufficient statistics $\phi_p(g, X)$.

$$\pi(g, X; \theta) = \frac{\exp\left[\sum_{p=1}^{P} \theta_p \frac{\phi_p(g,X)}{n^{v_p-2}}\right]}{c(\mathcal{G}, X, \theta)}$$

and define $t_p(g, X) = \frac{\phi_p(g,X)}{n^{v_p-2}}$. The likelihood of the model is

$$\pi(g, X; \theta) = \frac{\exp\left[\theta^T t(g, X)\right]}{\sum_{\omega \in \mathcal{G}} \exp\left[\theta^T t(\omega, X)\right]} \tag{51}$$

with log-likelihood

$$\ell(g, X; \theta) = \ln \pi(g, X; \theta) = \theta^T t(g, X) - \kappa(\mathcal{G}, X, \theta) \tag{52}$$

where $\kappa(\mathcal{G}, X, \theta)$ is defined as

$$\kappa(\mathcal{G}, \theta) = \ln c(\mathcal{G}, \theta) = \ln \sum_{\omega \in \mathcal{G}} \exp\left[\theta^T t(\omega, X)\right] \tag{53}$$

The variational approach consists of restating the problem of computing $\kappa(\mathcal{G}, X, \theta)$ as a maximization problem. A simple application of the Jensen's inequality provides a lower bound to the log constant (53). Let $q(g, X)$ be an arbitrary distribution of network configurations.

$$\begin{aligned}
\kappa(\mathcal{G}, X, \theta) &= \ln \sum_{\omega \in \mathcal{G}} \exp\left[\theta^T t(\omega, X)\right] && (54)\\
&= \ln \sum_{\omega \in \mathcal{G}} q(\omega, X) \frac{\exp\left[\theta^T t(\omega, X)\right]}{q(\omega, X)} && (55)\\
&\geq \sum_{\omega \in \mathcal{G}} q(\omega, X) \ln \left\{ \frac{\exp\left[\theta^T t(\omega, X)\right]}{q(\omega, X)} \right\} && (56)\\
&= \sum_{\omega \in \mathcal{G}} q(\omega, X) \theta^T t(\omega, X) - \sum_{\omega \in \mathcal{G}} q(\omega, X) \ln q(\omega, X) && (57)\\
&= \theta^T \mathbb{E}_q\left[t(\omega, X)\right] + \mathcal{H}(q) && (58)
\end{aligned}$$

where $\mathcal{H}(q) = -\sum_{\omega \in \mathcal{G}} q(\omega, X) \ln q(\omega, X)$ is the entropy of distribution q. We can find the tightest lower bound by maximizing (58) with respect to the distribution q.

$$\kappa^*(\mathcal{G}, X, \theta) = \max_{q \in \mathcal{Q}} \left\{ \theta^T \mathbb{E}_q\left[t(\omega, X)\right] + \mathcal{H}(q) \right\} \tag{59}$$

[62] See Bhamidi et al. (2011) and Diaconis and Chatterjee (2011). The re-scaling is done to guarantee that in the large n limit the sufficient statistics have the same magnitude.

It can be shown that this problem is equivalent to minimize the Kullback-Leibler divergence between the likelihood π and the approximating distribution q. However, the unconstrained optimization problem (59) does not have a closed form solution for most problems. The optimization is thus performed in a constrained set of distributions q that guarantee explicit solution. The mean-field approximation consists of constraining the problem in (59) to the set of completely factorized distributions q. In our context, the latter corresponds to the set of all Erdos-Renyi graphs with n vertices. Our mean-field approximation will therefore give us the Erdos-Renyi graph that best approximates the model.

Let μ be the probability that a link is formed in the Erdos-Renyi graph. Then it is easy to show that the entropy $\mathcal{H}(q)$ is

$$\mathcal{H}(q) = -n(n-1)\left[\mu \log \mu + (1-\mu)\log(1-\mu)\right] \tag{60}$$

and the expectation can be computed as

$$\begin{aligned}
\mathbb{E}_q\left[\theta^T t(\omega, X)\right] &= \mathbb{E}_q\left\{\sum_{i=1}^n\sum_{j=1}^n g_{ij}\theta_u' H_u\left(X_i, X_j\right)\right\} + \mathbb{E}_q\left\{\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n g_{ij}g_{ji}\theta_m' H_m\left(X_i, X_j\right)\right\} + \\
&+ \mathbb{E}_q\left\{\sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i,j}^n g_{ij}g_{jk}\frac{\theta_v' H_v\left(X_i, X_k\right)}{n}\right\} \\
&= \sum_{i=1}^n\sum_{j=1}^n \mathbb{E}_q(g_{ij})\theta_u' H_u\left(X_i, X_j\right) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}_q(g_{ij}g_{ji})\theta_m' H_m\left(X_i, X_j\right) + \\
&+ \sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i,j}^n \mathbb{E}_q(g_{ij}g_{jk})\frac{\theta_v' H_v\left(X_i, X_k\right)}{n} \\
&= \sum_{i=1}^n\sum_{j=1}^n \mu\theta_u' H_u\left(X_i, X_j\right) + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \mu^2\theta_m' H_m\left(X_i, X_j\right) + \sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i,j}^n \mu^2\frac{\theta_v' H_v\left(X_i, X_k\right)}{n} \\
&= \left[\mu\psi_1(\theta_{(1)}, X) + \mu^2\psi_2(\theta_{(2)}, X)\right]
\end{aligned}$$

where we have defined the functions $\psi_1(\theta_{(1)}, X)$ and $\psi_2(\theta_{(2)}, X)$ as

$$\begin{aligned}
\psi_1(\theta_{(1)}, X) &\equiv \sum_{i=1}^n\sum_{j=1}^n \theta_u' H_u\left(X_i, X_j\right) \\
\psi_2(\theta_{(2)}, X) &\equiv \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \theta_m' H_m\left(X_i, X_j\right) + \frac{n-2}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k\neq i,j}^n \theta_v' H_v\left(X_i, X_k\right)
\end{aligned}$$

The mean field equation is the first order condition of the maximization problem

$$\max_{q\in\mathcal{Q}_{MF}}\left\{\psi_1(\theta_{(1)}, X)\mu + \psi_2(\theta_{(2)}, X)\mu^2 - n(n-1)\left[\mu\log\mu + (1-\mu)\log(1-\mu)\right]\right\} \tag{61}$$

with first order conditions w.r.t μ

$$\mu = \frac{\exp\left[\frac{\psi_1(\theta_{(1)},X)}{n(n-1)} + 2\frac{\psi_2(\theta_{(2)},X)}{n(n-1)}\mu\right]}{1+\exp\left[\frac{\psi_1(\theta_{(1)},X)}{n(n-1)} + 2\frac{\psi_2(\theta_{(2)},X)}{n(n-1)}\mu\right]} \tag{62}$$

The equation is approximately the probability of transition of the Glauber dynamics of our model when the network is a draw from an Erdos-Renyi graph with parameter μ. Bhamidi et al. (2011) use an analogous equation to determine the convergence properties of the undirected case.

The special case in equation (26) in the text corresponds to $\psi_1(\theta_{(1)}, X) = n(n-1)\alpha$ and $\psi_2(\theta_{(2)}, X) = n(n-1)\frac{\beta}{2}$. If we include the utility from indirect links and popularity (with no explanatory variables X) we have $\psi_1(\theta_{(1)}, X) = n(n-1)\alpha$ and $\psi_2(\theta_{(2)}, X) = n(n-1)\frac{\beta}{2} + \frac{n-2}{n}n(n-1)\gamma$.

If we include additional components to the utility function the mean-field equation must be modified accordingly. The general equation is

$$\varphi(\mu,\theta) \equiv \frac{\exp\left[\sum_{d=1}^{D} e_d \frac{\psi_d(\theta_{(d)},X)}{n(n-1)}\mu^{e_d-1}\right]}{1+\exp\left[\sum_{d=1}^{D} e_d \frac{\psi_d(\theta_{(d)},X)}{n(n-1)}\mu^{e_d-1}\right]} \tag{63}$$

where e_d is the number of edges involved in the statistics corresponding to function $\psi_d(\theta_{(d)}, X)$. For example, when we include cyclic triads, we have a mean-field equation with a quadratic term

$$\mu = \frac{\exp\left[\frac{\psi_1(\theta_{(1)},X)}{n(n-1)} + 2\frac{\psi_2(\theta_{(2)},X)}{n(n-1)}\mu + 3\frac{\psi_3(\theta_{(3)},X)}{n(n-1)}\mu^2\right]}{1+\exp\left[\frac{\psi_1(\theta_{(1)},X)}{n(n-1)} + 2\frac{\psi_2(\theta_{(2)},X)}{n(n-1)}\mu + 3\frac{\psi_3(\theta_{(3)},X)}{n(n-1)}\mu^2\right]} \tag{64}$$

where $\psi_3(\theta_{(3)}, X) = \frac{n-2}{n}n(n-1)\frac{\delta}{3}$.

B.4 Additional convergence simulations

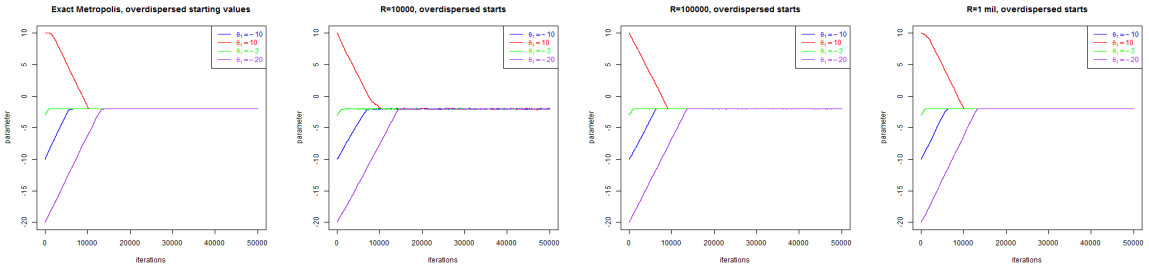
The results contained in Table 1 are obtained from a model with two parameters

$$U_i(g,X) = \theta_1 \sum_j g_{ij} + \theta_2 \sum_j g_{ij}g_{ji} \tag{65}$$

For this model the normalizing constant can be computed in closed form and therefore it is possible to compare the performance of the exact metropolis-hastings algorithm with the approximate exchange algorithm. All the figures reported in this appendix correspond to a network with $n = 1000$ players.

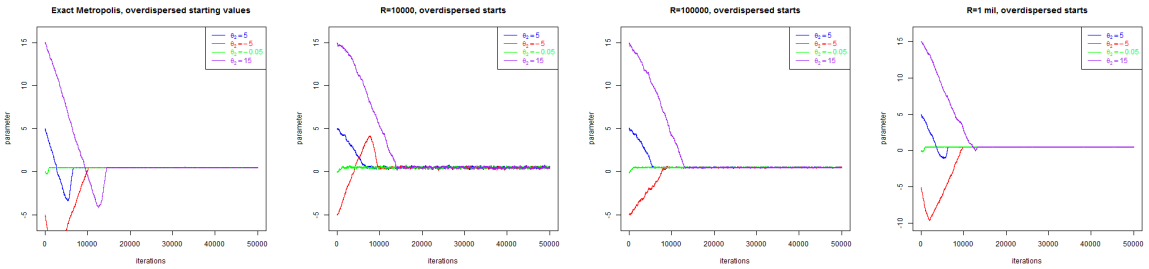
Figure 7: Overdispersed starting values, $n = 1000$

Convergence of θ_1



(a) Exact MH (b) $R = 10000$ (c) $R = 100000$ (d) $R = 1$ mil

Convergence of θ_2



(e) Exact MH (f) $R = 10000$ (g) $R = 100000$ (h) $R = 1$ mil

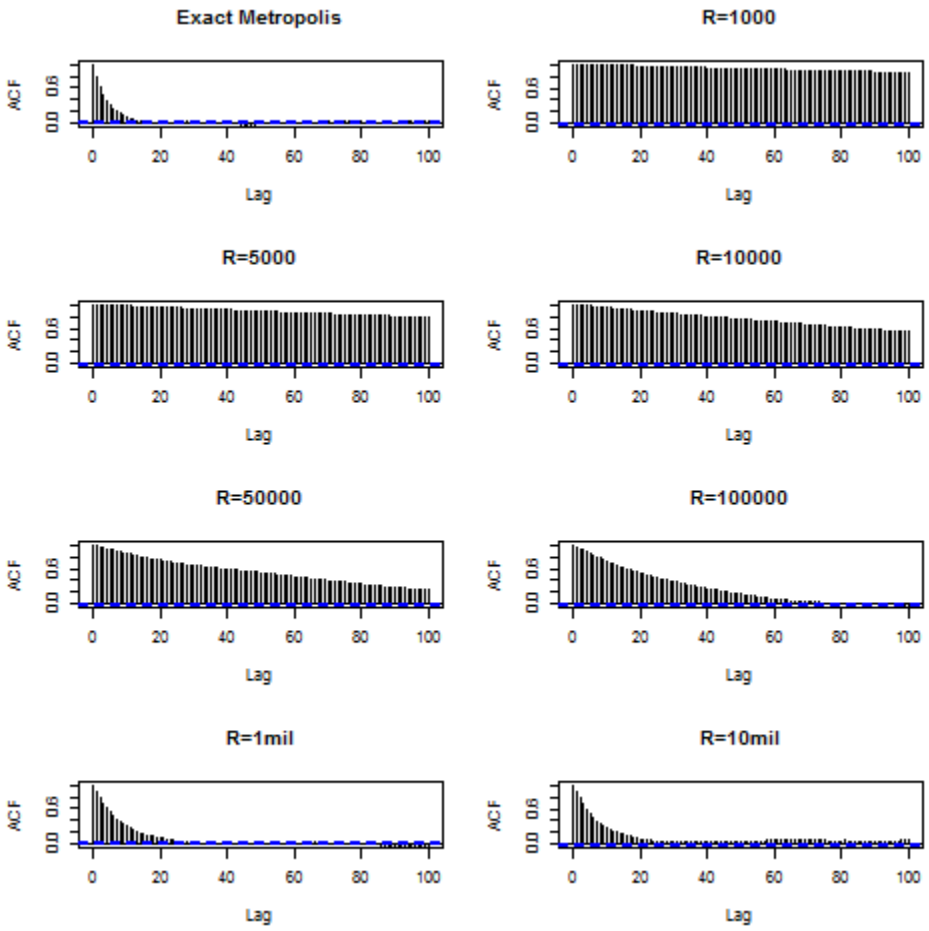
The graphs in Figure 7 show the convergence of the algorithm for overdispersed starting values. The simulations show fast convergence to the high density region of the posterior. The number R of network simulations per each proposed parameter does not seem to affect the convergence. Indeed the exact metropolis and the approximate exchange algorithm show very similar patterns of convergence. The latter result suggest that in practice it useful to run the sampler using a low R to find a plausible region for the starting values of the simulation.

The autocorrelation plots are reported in Figure 8. Notice that increasing the number of network simulations R contributes to the decrease in autocorrelation of the chains. We obtain good result for a moderate amount of simulations.

The bivariate posterior is shown in Figure 9. The exact metropolis sampler is indicated by the black solid line, with estimated posterior mean represented by the black triangle. The approximate algorithm is the red dashed line, with posterior mean shown as a red dot. While $R = 10000$ network simulations are not sufficient to obtain an accurate estimate of the posterior, we obtain a reasonable approximation with $R = 100000$ simulations, even if the approximate algorithm has fatter tails. The posterior mean is estimated precisely with as few as $R = 100000$ network simulations. However, to obtain an accurate posterior we need to use at least $R = 1000000$.

The same patterns are shown in Figure 10 for the marginal posteriors.

Figure 8: Autocorrelation functions, $n = 1000$



Autocorrelation plots for the model with utility from direct links and mutual links only, with $n = 1000$ players. Comparison between the exact metropolis-hastings algorithm and the output from the approximate exchange algorithm with R =10000, 100000, 1 million and 10 millions.

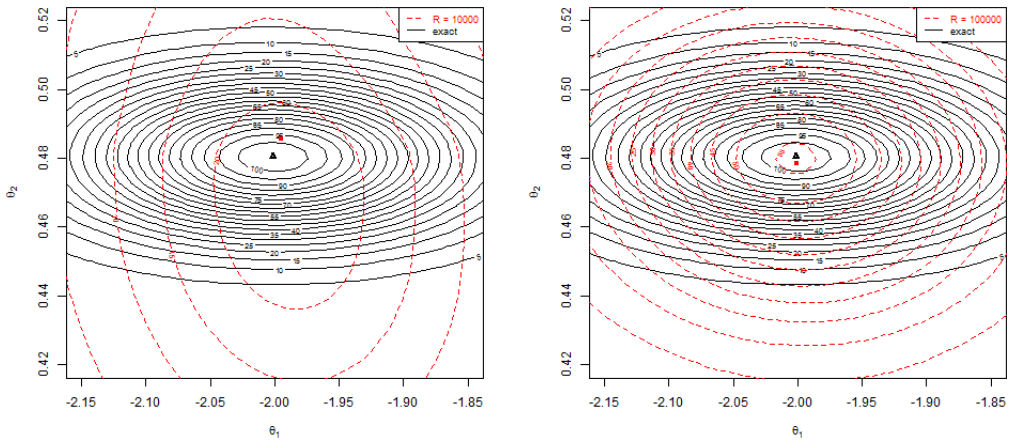
B.5 Parallel estimation with multiple networks

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

$$\pi\left(g_c, X_c, \theta\right) = \frac{\exp\left[Q\left(g_c, X_c, \theta\right)\right]}{\sum\limits_{\omega \in \mathcal{G}_c} \exp\left[Q\left(\omega_c, X_c, \theta\right)\right]} \tag{66}$$

Figure 9: Level curves for posterior, $n = 1000$



R=10000 R=100000

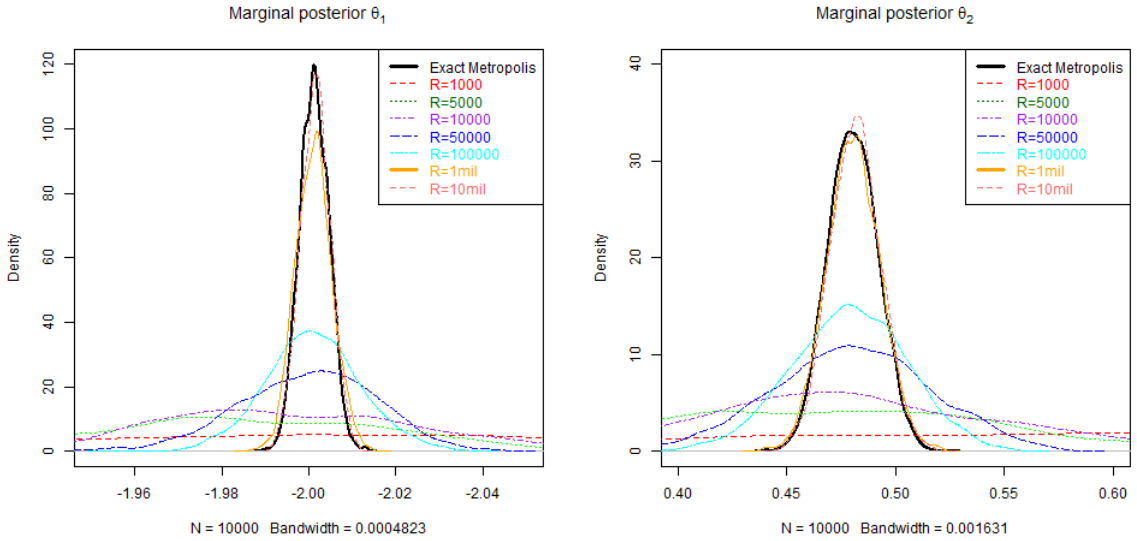
R=1 million R=10 millions

Level curves of estimated posteriors for different values of R (red dashed line), compared to the posterior computes using the exact metropolis-hastings algorithm (black solid). The black triangles is the posterior mean computed with the exact metropolis, and the red dot is the posterior computed with the approximate exchange algorithm .

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

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

Figure 10: Convergence of marginal posteriors, $n = 1000$



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 APPROXIMATE 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 $q_\theta(\cdot|\theta)$*

$$\theta' \sim q_\theta(\cdot|\theta) \tag{67}$$

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*

$$g'_c \sim \mathcal{P}^{(R)}_{\theta'}(g'_c|g_c) \tag{68}$$

3. *Update the parameter according to*

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

where

$$\alpha_{pex}(\theta,\theta') = \min\left\{1, \frac{\exp\left[\sum_{c=1}^{C} Q(g_c', X_c, \theta)\right]}{\exp\left[\sum_{c=1}^{C} Q(g_c, X_c, \theta)\right]} \frac{p(\theta')}{p(\theta)} \frac{q_\theta(\theta|\theta')}{q_\theta(\theta'|\theta)} \frac{\exp\left[\sum_{c=1}^{C} Q(g_c, X_c, \theta')\right]}{\exp\left[\sum_{c=1}^{C} Q(g_c', X_c, \theta')\right]}\right\} \quad (69)$$

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.6 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.[63] 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 careful use of prior distributions. This is standard practive in the DSGE estimation literature, where parameters are often ill-identified and prior distributions are used to produce more precise estimates (Fernandez-Villaverde et al. (2010)). 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 a specification of the parameters

$$\theta_p = \theta_{p0} + \sum_{c=1}^{C} \theta_{pc} Z_c \quad (70)$$

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 (see Appendix C).

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

$$p(\theta) = \mathcal{N}(\mathbf{0}, 3\mathbf{I}_P), \quad (71)$$

[63] 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

$$q_\theta(\cdot|\theta) = \mathcal{N}(\mathbf{0}, \delta\boldsymbol{\Sigma}), \tag{72}$$

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

B.7 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.[64] 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, and the researchers made a significant effort to avoid any missing information on the students.[65]

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.[66] 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 necessarily imply that Jim nominates Paul.[67] 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 student that declares to be of Hispanic origin is considered Hispanic. The remaining non-Hispanic

[64] 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

[65] While this sample contains no missing covariate information for the students, there are several missing values for the parental variables.

[66] One can think that this limit could bias the friendship data, but only 3% of the students nominated 10 friends (Moody, 2001). Moreover, the estimation routine could be easily extended to deal with missing links, as reported in Appendix.

[67] 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.

In a previous version of the paper, I also control for homophily in income. I construct the income of the family using a question from the parent questionnaire.[68] In the estimated models I control for income difference between the students and income levels.

There may be some unobservable variables that affect network formation. For example some students may be "cool" and receive more friendship links than others. To partially control for such effects, I use information from the interviewer remarks about the physical attractiveness and personality of the student interviewed. I define a dummy variable "beauty", which is equal to 1 if the interviewer told that the students was very attractive. Analogously, the dummy "personality" is equal to 1 if the interviewer responded that the personality of the student was very attractive.

Descriptive statistics for the saturated sample are in Table 5. 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. Many schools are almost 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.[69] 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

[68] There are several cases in which the family income is missing. For those observations, I imputed values drawn from the unconditional income distribution of the community. An alternative but computationally very costly alternative is to introduce an additional step in the simulation, in which the imputation of missing incomes is done at each iteration.

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

$$FRAG = 1 - \sum_{k=1}^{K} (s_k)^2 \tag{73}$$

Table 5: **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

Descriptive statistics of the saturated sample for Wave 1 of Add Health. Segregation is measured with the Freeman (1972) segregation index. The index varies from 0 (perfect integration) to 1 (perfect segregation). The racial fragmentation index measures the probability that two randomly chosen students belong to different racial groups. Higher values of the index indicate a more racially heterogeneous population.

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.

B.8 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

$$FSI = \frac{\mathbb{E}[n_{AB}] + \mathbb{E}[n_{BA}] - (n_{AB} + n_{BA})}{\mathbb{E}[n_{AB}] + \mathbb{E}[n_{BA}]} \tag{74}$$

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[70]

$$SEG = \max\{0, FSI\} \tag{75}$$

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

[70] The index (74) 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

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

$$U_i\left(g_{ij}=1,g_{-ij},X\right)+\eta_i+\eta_j+\nu_{ij1}\geq U_i\left(g_{ij}=0,g_{-ij},X\right)+\eta_i+\nu_{ij0} \tag{76}$$

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

$$\mathcal{Q}\left(g,X,\theta;\eta\right)=Q\left(g,X,\theta\right)+\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ij}\eta_j \tag{77}$$

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

$$\pi\left(g,X,\theta\right)=\int_{\Upsilon}\frac{\exp\left[\mathcal{Q}\left(g,X,\theta;\eta\right)\right]}{\sum_{\omega\in\mathcal{G}}\exp\left[\mathcal{Q}\left(\omega,X,\theta;\eta\right)\right]}dW\left(\eta\right) \tag{78}$$

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 (Rossi et al. (1996), Athey and Imbens (2007)).

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

$$\Pr\left(\eta|g,X,\theta\right)=\frac{W\left(\eta\right)\pi\left(g,X,\theta;\eta\right)}{\pi\left(g,X,\theta\right)} \tag{79}$$

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

$$\alpha_\eta\left(\eta,\eta',g,\theta_s\right)=\left\{1,\frac{W\left(\eta'\right)\pi\left(g,X,\theta;\eta'\right)q_\eta\left(\eta|\eta'\right)}{W\left(\eta\right)\pi\left(g,X,\theta;\eta\right)q_\eta\left(\eta'|\eta\right)}\right\} \tag{80}$$

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

$$\theta_p=\theta_{p0}+\sum_{c=1}^{C}\theta_{pc}Z_c+\xi_c \tag{81}$$

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

The estimation method is flexible enough to allow for estimation when there are missing links. The Add Health dataset could raise some concern about missing links, since the original questionnaire asks 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 vector. Then the algorithm proceeds with the exchange algorithm as before using the augmented data $g = \{g_{mis}, g_{obs}\}$.

The main cost of these extensions is the increased computational burden, which may be substantial.