

Segregation in Social Networks:^{*,†} A Structural Approach

Angelo Mele[‡]

VERY PRELIMINARY AND INCOMPLETE

January 29, 2010

Abstract

I develop a dynamic heterogeneous agents model of strategic network formation, where a matching technology and preferences for same-type individuals drive the dynamics of friendship. The model converges to a unique stationary distribution that is structurally estimated using Markov Chain Monte Carlo methods and stochastic approximation techniques. I use data from the restricted version of Add Health, a representative survey of US high schools containing detailed data on the actual friendship network of each student.

Results show that the tendency to create links to individuals of the same racial group is pervasive. However, I find that a certain degree of heterogeneity in mutual and indirect friends increases utility of the agents.

I use the model to simulate policies whose goal is to increase the degree of interracial contact. I re-assign some minority students to another school where there are only few minority students. The simulations show that these policies decrease students' welfare in the new stationary equilibrium.

*I am extremely grateful to Roger Koenker, for continuous guidance and generous advice. I am also heavily indebted with Ron Laschever for several suggestions that improved the quality of the paper. I thank Dan Bernhardt, Alberto Bisin, Ethan Cole, George Deltas, Han Hong, Darren Lubotsky, Antonio Mele, Dennis O'Dea, Micah Pollak, Sergey Popov, Giorgio Topa, Harry Tsang, Antonella Tutino, Jungmo Yoon, David Wiczer, Junfu Zhang and participants to the Labor Lunch Seminar at UIUC, 2008 North American Summer Meeting of the Econometric Society, UIUC Research Seminar, EEA-ESEM 2009 and the Graduate Student Conference at Washington University St. Louis 2009 for helpful comments. All remaining errors are mine

[†]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.

[‡]*Address:* Dept. of Economics, University of Illinois at Urbana-Champaign, 419 David Kinley Hall, 1407 W. Gregory, Urbana, IL 61801. *Email:* amele2@uiuc.edu

1 Introduction

I develop a dynamic heterogeneous agents model of social network formation in which individuals randomly meet and form (or sever) links to other agents, according to their preferences for (observable) socioeconomic characteristics. I structurally estimate the model to recover the preference parameters determining the degree of interracial friendship, using Markov Chain Monte Carlo simulation and stochastic approximation methods. I rely on the restricted version of Add Health, that contains detailed data on the *actual* network of a representative sample of US high schools. To the best of my knowledge this is the first work performing a fully structural estimation of a network formation model.¹

The estimation results confirm that racial homophily, the tendency to form links with similar individuals, is a pervasive phenomenon: high school students prefer to be friends with individuals of their same racial group. I estimate a model in which individuals only care about the direct links they form with other people: such a model provides strong support for homophily. However, when I consider a model in which individuals care also about mutual links, the characteristics of friends of friends, and their popularity among other students, results change. Homophily is very important for the direct links, but *individuals prefer a more heterogeneous set of mutual and indirect links*. Previous empirical analysis failed to detect this effect since only the direct links were modeled explicitly.

I use the model to study counterfactual policy experiments and their effect on the students' welfare in the stationary equilibrium. I simulate a busing program that re-assigns the minority students enrolled in one of the Add Health schools (school 94) to another school in which most of the students are non-minority (school 25).² The results show that this reassignment can decrease welfare. I compute the average welfare of individuals without and with the policy and show that agents are better off without the policy. Alternative policies can be easily simulated using this framework.

The analysis of social network formation has become a topic of interest for economist in the last decade. Several studies have shown the importance of social networks in shaping education performance (De Giorgi, Pellizzari and Redaelli, 2009, Cooley, 2009), job market outcomes (Topa, 2001; Laschever, 2009; Conti et al, 2009), diffusion of innovations (Conley and Udry, 2009; Bandiera and Rasul, 2006), risky behavior (Nakajima, 2007; Krauth, 2006) and risk sharing (Fafchamps and Gubert, 2006; Comola, 2008).³

This literature assumes that an agent's decision is a function of his reference group's choices and the interactions occur in a network of social relationships. The latter network is endogenous, since people choose whom they want to interact with, and this imposes a major challenge to the estimation and identification of network effects. Therefore the study of how the social network is formed is important for at least two reasons: 1) its structure determines the way interactions among peers take place and the mechanism driving the diffusion of a certain behavior in the population; 2) a systematic analysis of social network formation may help in developing strategies to correct

¹Comola (2008) develops a strategy for estimating the connection model of Jackson and Wolinski (1996). Assuming normal unobservables she shows that the requirement of pairwise stability for the equilibria imposes a bivariate probit structure on the empirical model.

²See Angrist and Lang (2004) for the analysis of a busing program with this goal in mind.

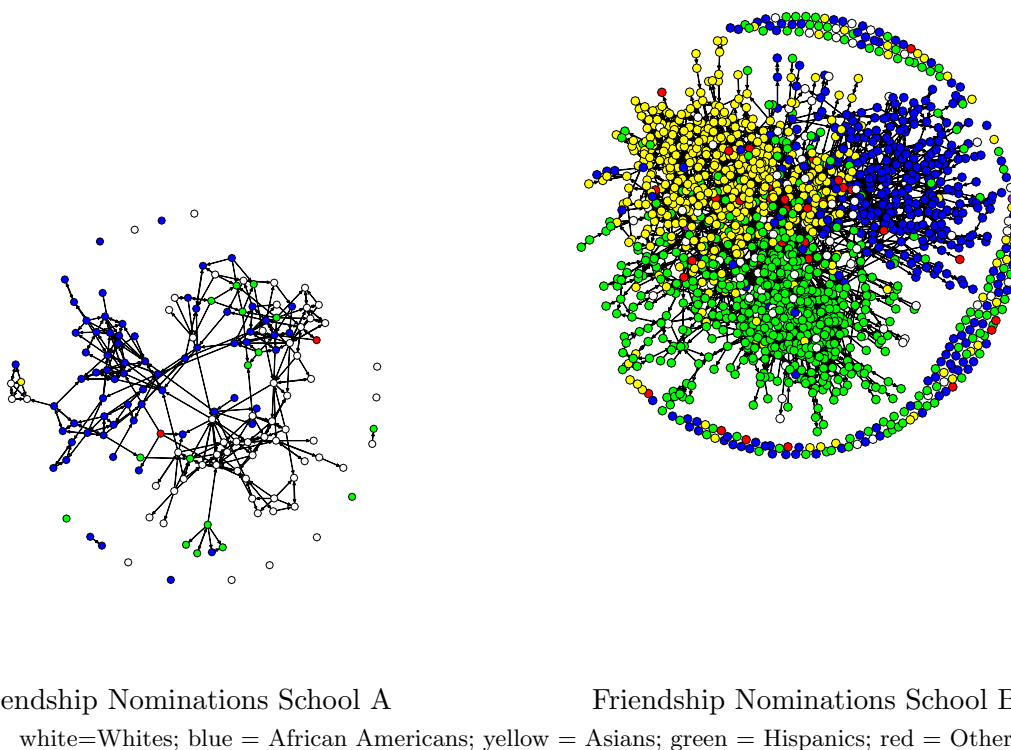
³For a survey see Jackson (2008) or Ioannides and Loury (2007)

the endogeneity when estimating peer effects models.⁴

Socially generated networks usually display segregation and homophily: individuals tend to link people with similar characteristics, creating clusters of highly connected agents homogeneous by race, gender, income or other socioeconomic characteristics.⁵

The graphs in Figure 1 show the friendship links among students of two schools, extracted

Figure 1: Two School Networks



Note: The graphs represent two different schools extracted from AddHealth. Each dot represents a student, each arrow is a friend nomination.

from the AddHealth data.⁶ Each dot represents a student, the color describes his racial group and an arrow from dot i to dot j indicates that i nominated j as friend. There are 150 students in Network A and 1664 in Network B. In Network A two separate communities emerge, one all

⁴See Bramoullé and Fortin (2009) for a review of the problems arising in estimating and identifying social networks effects.

⁵See Moody (2002) and Currarini, Jackson and Pin (2008)

⁶AddHealth is a nationally representative study of US high schools, containing detailed data on friendship and individual characteristics for each sampled school. In total there are data for 90118 students enrolled in 146 schools. More detail below.

white, the other all black. Several individuals have interracial friendships that constitute a bridge among these communities. In Network B we observe a similar network architecture, with several communities identified by their racial group.

The main goal of this paper is to study the determinants of segregation observed in many social networks, addressing two main questions: 1) how homophily affects the equilibrium structure of school friendship networks; 2) how public policies can modify this equilibrium outcome.

This paper provides several contributions to the economics literature on network formation. From a theoretical point of view, I develop a new model of network formation that allows structural estimation. The network formation is *sequential*: in each period only one agent is active and she can update only one link. There is a random matching technology that drives the meetings of agents in the economy. Agents follow a best-response dynamics⁷ and are heterogeneous with respect to observable exogenous characteristics.⁸ Individuals care about their links to other people, according to the socioeconomic characteristics of the pairs involved in each link. Agents care also about friends of friends and their own popularity, i.e. links to them created by other students. I assume individuals receive a preference shock when deciding to update a link.

The crucial assumption is a mild restriction of the preferences that imposes symmetry in the utility from mutual links and the same functional form for the utility from indirect links and popularity. This restriction allows me to characterize the network formation game as a potential game.⁹ All the incentive of the players in any state are summarized by an aggregate function, *the potential*, that maps networks configurations and socioeconomic characteristics into potential levels. When an agent updates a link, the change in her utility is equal to the change in the potential.

The potential has a crucial role in the characterization of the Nash equilibria, the long run dynamics, the stationary distribution, the estimation strategy and the simulation techniques. The local maxima of the potential function correspond to the Nash networks of a model without any preference shock. Assuming a matching technology in which any single match has positive probability, I prove that in the long run, the network converges to one of the Nash equilibria with probability one. Assuming logistic preference shocks, I show that the model is ergodic and has a *unique stationary distribution*, which is a function of the potential.

The second methodological contribution is the development of estimation techniques for this class of models. The dataset available to the econometrician is one observation of the network in a point in time. The model provides the structural dependence among links that allows estimation using a single observation.¹⁰

Assuming that the network observed in the data is drawn from the stationary distribution, the model can be estimated using Maximum Likelihood techniques, in principle.

However, the likelihood is known up to a normalizing constant that imposes serious computational problems. The normalizing constant is the sum of a function of the potential, where the sum is over all possible network configurations. To be concrete, for a network with 10 individuals, the

⁷Individuals behave myopically and maximize their current utility, as in Blume (1993). In principle it is possible to extend the model to forward-looking agents, but the computational complexity would make estimation impossible, as explained in the empirical section

⁸See Galeotti (2006), Currarini, Jackson and Pin (2008), De Marti and Zenou (2009) for models of network formation with heterogeneous agents.

⁹See Monderer and Shapley (1996) for a description of games with a potential.

¹⁰This estimation problem is common in spatial analysis, where the statistician observes a single realization of the spatial process and imposes the spatial dependence structure to allow estimation. See Cressie (1993) for an introduction to models and methods in spatial analysis.

number of possible configurations is $2^{90} \sim 10^{27}$. A very fast state-of-the art computer will take several months just to compute the sum for a given parameter value.

I tackle this estimation problem in three steps. First, assuming utility is linear in parameters, I show that the likelihood belongs to the exponential family with normalizing constant (See Lehman (1983)). The assumption of linearity in parameters is not necessary to estimate the model, but significantly simplifies the problem and reduces the computational time needed to reach convergence. Second, exploiting a property of the exponential family, I solve the system of first order conditions implied by the maximum likelihood problem using stochastic approximation methods. I simulate sample networks from the stationary distribution through a Metropolis-Hastings algorithm. Third, I improve the basic MCMC algorithm in order to take into account the extremely large state space of my theoretical model. I use Simulated Tempering and several variance reduction techniques to significantly reduce the length of the runs needed for estimation, and I exploit the sparsity of the network to increase computational speed.

One should notice that the sequential feature of the network formation model is not needed for estimation: it is just a framework through which we can interpret the estimated parameters.

1.1 Related Literature

The theoretical model presented here is closely related to Currarini, Jackson and Pin (2008) and De Marti and Zenou (2009). The former develop a model of network formation in which individuals enter the network and randomly meet other agents, paying a fixed cost for each match. Diminishing returns in the number of friends implies that this process ends after a finite number of meetings. A version of the model in which both preferences and matching process are biased towards same-type individuals is able to explain several patterns observed in the data. However in this model the structure of the network does not play an active role: two individuals of the same type will behave in the same way, no matter their position in the network.

De Marti and Zenou's (2009) model takes into account the network complexity. Individuals' costs to link different an agent depend on the agent's type, but also on how many other-type friends the agent has. This implies that the network structure has an effect on the benefits and costs perceived. In this model two identical individuals can make difference linking choices according to their positions in the network. However the model is static and there is no explicit meeting process; one could say that all the individuals in the network meet.

My model is based on a matching technology and biased preferences as in Currarini, Jackson and Pin (2008). Furthermore I explicitly model the effect of the network architecture in shaping the incentive of the players to form links as in De Marti and Zenou (2009): an agent benefits from a direct link but also from indirect links, mutual links and popularity. An agent with two African-American friends and one hispanic friend is a different "good" than an agent with one white friend and three asian friends.

In the model developed in this paper, two identical individuals in term of observable characteristics that link to each other, receive the same direct and mutual utility, but they can have different indirect utility and popularity according to their individual network structure.

The sequential network formation process has been used in previous works. Bala and Goyal (2000) analyze a game where myopic agents update all of their links in each period, assuming the strategies of other players are the same as previous period. They show convergence to the Nash

networks. Jackson and Watts (2001) model a sequential network formation game, but the link formation requires mutual consent.

The crucial feature of this model is the existence of a potential function that characterizes the equilibria and provides a strategy for estimation. In the literature, Gilles and Sarangi (2004) study networks formation models with a potential in a context where agents do not receive utility from indirect links. Furthermore, in their model the link is mutual. Chakrabarti and Gilles (2007) show conditions under which a network formation game admits a potential. In this model consent is required for the existence of the link.¹¹ The contribution of my work is to recognize that the existence of a potential allows estimation through MCMC methods in a natural way.

This paper is also related to the statistical literature on exponential random graphs.¹² These are statistical models that assume markovian dependences among links resulting in a network whose distribution belongs to the exponential family with normalizing constant. In the present work I provide a micro-foundation of the exponential random graph model as stationary equilibrium of an evolutionary strategic game.

The present paper contributes also to the empirical literature on homophily and interracial contact. Moody (2001) shows that there is a positive relation between racial heterogeneity and friendship segregation in the AddHealth data. He suggests that concentrating minority students in big schools may create more segregation. Mouw and Entwisle (2006) argue that racial segregation in schools is also consistent with the effect of residential segregation. Residential segregation decreases the (average) physical distance among individuals of the same social group and increases it for individuals of different social groups. Therefore the cost of interacting with similar people is lower and the proportion of same group friends is higher in more spatially segregated communities. Using maximum pseudo-likelihood techniques and the AddHealth data, the authors estimate that about a third of the degree of racial segregation in school friendship can be attributed to residential segregation. Recent contributions from economists are Mayer and Puller (2007) and Comola (2008). Mayer and Puller (2007) study a model of network formation where individuals meet according to some exogenous probabilities and decide to link according to socioeconomic characteristics. Agents also meet friends of friends and may link to them. The latter mechanism generates the clustering necessary to match the data. They estimate the model using data from Facebook and find that racial homophily is pervasive, even after controlling for many socioeconomic characteristics. Policy experiments show that programs that modify the probability of meeting have limited effectiveness in promoting integration. Comola (2008) shows that the notion of pairwise stability for the equilibria provides enough restriction for the connection model (Jackson and Wolinski, 1996) to be estimated using a bivariate probit.

The remainder of the paper is organized as follows. Section 2 provides a description of AddHealth and summary statistics of the data. Section 3 develops the model and show the theoretical results on convergence to the stationary distribution. Section 4 discusses the estimation strategy and the solution method. Section 5 shows the estimation results and Section 6 concludes. Appendix A analyzes a model with homogeneous agents. Appendix B discusses several estimation details and Appendix C contains the proofs of propositions. Appendix D describes the sample selection and variable construction, Appendix E explain the details of the estimation and Markov Chain Monte Carlo methods for simulation.

¹¹In the context of coalitional games see also Slikker and van den Nouweland (2002)

¹²See Wasserman and Pattison (1996), Frank and Strauss (1986), Snijders (2002) for an introduction to these models.

2 The AddHealth Data

The *National Longitudinal Study of Adolescent Health* (AddHealth) is a study that collected data for a nationally representative sample of adolescent in United States. Participants were enrolled in grades 7-12 during the 1994-1995 school year and this cohort has been followed into adulthood with four successive in-home interviews.¹³

In this paper I use only data from Wave I, containing information on 132 nationally representative US high schools.¹⁴ The survey collected information about students through in-school and in-home interviews. School administrators and parents were also interviewed separately to collect school relevant information and parent backgrounds.

A total of 90118 students was given an *in-school* questionnaire, to collect individual characteristics. Each student was given a school roster and was asked to identify up to five male and five female friends.¹⁵ I use the friendship nominations as proxy for the social network in a school. Notice that the resulting network is *directed*: Paul may nominate Jim but this does not necessarily implies that Jim nominates Paul.¹⁶ The model developed in this paper takes this feature of the data into account.

A subsample of 20745 students was also given an *in-home* questionnaire, that collected most of the sensible data. This section of the dataset collects the spatial location of each housing unit which can be used to relate residential segregation and school segregation.

I use data on racial group, grade and gender of individuals. A student with a missing value in any of these variables is dropped from the sample. Each students that declares to be of hispanic origin is considered hispanic. The remaining non-hispanic 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.

Table 1 contains the descriptive statistics for racial composition. The smallest school has 43 students while the largest (the one in Figure 1(b)) has 1664 people. On average a school has 151.8 students, which means that the school in Figure 1(a) is a typical school.

The typical school has 60% whites, 21% blacks, 4% asians and 13% hispanics. On average the index of racial fragmentation is 0.3471.¹⁷ There is a lot of variation in the racial composition of schools: some schools are completely segregated, containing 100% of whites or almost 100% of

¹³More details about the sampling design and the representativeness are contained in Moody (2002) and the AddHealth website <http://www.cpc.unc.edu/projects/addhealth/projects/addhealth>

¹⁴A school is defined high school if it offers an 11th grade.

¹⁵One can think that this limit could bias the friendship data, but only 3% of the students nominated 10 friends (Moody (2002)).

¹⁶Some 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. Echenique and Fryer (200?)

¹⁷The index of racial fragmentation is the complement of the Herfindhal-Hirschman index. If there are K racial groups and the share of each race is s_k , the index is

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

An index of 0 indicates that there is only one racial group and therefore the population is perfectly homogeneous. Higher values of the index represents increasing levels of racial heterogeneity.

African Americans. Others show high levels of diversity in their populations. The gender composition is more or less homogeneous across schools.

	Min	Q1	Median	Mean	Q3	Max
TotStudents	43	87	127	151.8	173	1664
Whites	0	0.3318	0.6909	0.6013	0.9061	1
Blacks	0	0.01114	0.0818	0.2147	0.3545	0.9866
Asians	0	0	0.01156	0.03897	0.03739	0.6667
Hispanics	0	0.01345	0.04652	0.1304	0.1704	0.901
Others	0	0	0.00678	0.01422	0.0191	0.3165
Females	0.01923	0.4896	0.5084	0.5085	0.5375	0.6538
FragmRace	0	0.122	0.3725	0.3471	0.5181	0.7641

Table 1: Summary Statistics of Racial Composition

Table 2 contains the summary statistics for the grade composition. Not all the schools offer all grades 7-12. Some schools only offer grades 9-12 or 10-12 and they are in the same community of a school offering the remaining grades. When a school does not offer all the grades in the community, the other school is sampled too so that all the grades are covered. The typical school offers all the grades.

The final sample for the data analysis consists of 130 schools for which the network data

	Min	Q1	Median	Mean	Q3	Max
Seven	0	0.00437	0.1528	0.218	0.4846	0.6667
Eight	0	0.00871	0.1618	0.212	0.4578	0.5538
Nine	0	0.0226	0.1869	0.1554	0.259	0.3968
Ten	0	0	0.1816	0.147	0.2543	0.3663
Eleven	0	0	0.1776	0.1424	0.2491	0.345
Twelve	0	0	0.1562	0.1252	0.2137	0.3269

Table 2: Summary Statistics for Grade Composition

are reliable. If a student nominated a friend who is not enrolled in the same school, I drop that friendship from the dataset.

3 A Model of Network Formation

I develop a dynamic heterogeneous agents model of noncooperative network formation. The link formation is sequential: in each period only one agent is active and she only updates one link.¹⁸

An agent is characterized by her socioeconomic characteristics: gender, race, location, etc. In

¹⁸ Models of network formation with sequential link updates are not unusual in the literature. On one hand the theoretical literature. On the other hand it has been noted that the sequential structure is crucial for breaking one curse of dimensionality in the computation of equilibria and thus in estimation. See Dorajleksi and Judd (2008), Arcidiacono et al (2009).

each period a random agent i is drawn from the population, she meets another agent j according to a random matching technology and decides about linking j . When updating the link, the individual receives a random shock to her preferences. Agent i has complete information on the current network shape and the characteristics of each individual in the network. The decision of linking is made according to i 's preferences, given his information.

3.1 Setup

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

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

$$G_{ij} = \begin{cases} 1 & \text{if individual } i \text{ links individual } j \\ 0 & \text{otherwise} \end{cases}$$

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

The network represented by G is *directed*: the existence of a link from i to j does not imply the existence of the link from j to i , i.e. $g_{ij} \neq g_{ji}$.¹⁹ This modeling choice reflects the structure of the AddHealth data, in which friendship nominations are not necessarily mutual.

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_{ij}^t . The network including all the current links but g_{ij}^t , i.e. $g^t \setminus g_{ij}^t$, is denoted as g^t_{-ij} .

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

3.1.1 Network Formation Process

Individuals form links over time through a *best response dynamics*, generating a Markov chain of networks. The main ingredients of this process are random matching and utility maximization.

Matching Technology. In each period a random pair of individuals in the economy meet. The *meeting process* is a stochastic sequence $m = \{m^t\}_{t=1}^{\infty}$ with support $\mathcal{I} \times \mathcal{I}$. The realizations of the meeting process are ordered pairs $m^t = \{i, j\}$, indicating which player i should play and which link g_{ij} can be updated at each t .^{20,21}

¹⁹In the paper I rely on the sociomatrix representation of a network (see Wasserman and Faust (1994)), since it is more convenient for the empirical and computational strategy. Alternatively the network can be represented as a graph: individuals are the set of nodes, while their friendship relations are the edges of the graph. See Jackson and Wolinsky (1996) for example.

²⁰Several models incorporate a matching technology in the network formation process. Jackson and Watts (2002) assume individuals meet randomly according to a discrete uniform distribution. Currarini, Jackson and Pin (2008) introduce a matching process that is biased towards individuals of the same type. I am agnostic about the specification of the matching process.

²¹An equivalent formulation is as in Blume (1993), where agents have Poisson clocks that ring at random points in time to determine when the player can update her own strategy. Given the continuous-time formulation in each

In general, the probability that player i meets agent j , conditional on the current network is

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

and $\sum_{g \in \mathcal{G}} \sum_i \sum_{j \neq i} \rho(g^{t-1}, X_i, X_j) = 1$.

The matching probability depends on the current network (say the existence of a link between i and j , for example) and the characteristics of the pair. This formulation includes matching technologies with a bias for same-type individuals as in Currarini, Jacscon and Pin (2008), but it is more general since it is state-dependent. The simplest example of (2) is an i.i.d. discrete uniform process with $\rho_{ij} = \frac{1}{n(n-1)}$.

I make a specific assumption about the meeting process in next section.

Utility Maximization Conditional on the realization $m^t = ij$, player i updates the link ij to maximize her current utility, taking previous period network g_{-ij}^t as given and knowing the individual characteristics of the other agents in the population. The individual has *complete information* since she can observe the entire network shape and the individual attributes of all agents.

I assume that when the individual i has to play, she receives an idiosyncratic shock $\varepsilon \sim F(\varepsilon)$ to her preferences, that the econometrician cannot observe.

Agents behave strategically: when the realization of the meeting process is $m^t = ij$, player i chooses to link, $g_{ij}^t = 1$, if it is a best response to the current network configuration, i.e. if

$$U_i(g_{ij}^t = 1, g_{-ij}^{t-1}, X) + \varepsilon(g_{ij}^t = 1) \geq U_i(g_{ij}^t = 0, g_{-ij}^{t-1}, X) + \varepsilon(g_{ij}^t = 0)$$

This linking process generates endogenous interdependence among links (and agents choices).²²

The stochastic process described above generates a temporal sequence $[g^0, g^1, \dots, g^t]$ of networks. In each period only one element of the random matrix G is updated, conditioning on previous period network. Therefore the sequence is a Markov chain, with transition probabilities determined by the meeting process and agents' preferences.²³

instant there is at most one active player.

²² This is different from models in which the spatial dependence is assumed to arise because of the spatial correlation among errors. For example Conley and Udry (2004) and Fafchamps and Gubert (2006) estimate a model

$$\Pr(g_{ij} = 1) = f(X_i, X_j) + \epsilon_{ij} \quad (3)$$

where the interaction among links is obtained assuming spatial dependence among errors ϵ_{ij} 's. In the model developed here the spatial dependence is endogenous, as a consequence of interdependent individual decisions. In other words I am studying a model

$$\Pr(g_{ij} = 1 | g_{-ij}) = f(g_{-ij}, X_i, X_j) + \epsilon_{ij} \quad (4)$$

in which the spatial dependence is generated by the correlation between g_{ij} 's, even when the errors ϵ_{ij} 's are i.i.d.

²³ I do not go over the tedious derivation of the actual transition matrix for the chain. The set of all possible states is \mathcal{G} , the probability of transition from a network $g^t = g$ to next period network $g^{t+1} = (g'_{ij}, g_{-ij})$ is

$$\rho(g, X_i, X_j) I_{\{U_i(g'_{ij}, g_{-ij}, X) + \varepsilon(g'_{ij}) \geq U_i(g_{ij}, g_{-ij}, X) + \varepsilon(g_{ij})\}}$$

where $I_{\{\dots\}}$ is an indicator function. The transition probability is zero if the networks differ in more than one element.

3.1.2 Preferences

I define the following utility functions for all $i, j \in \mathcal{I}$: $u_{ij} = u(X_i, X_j)$, $m_{ij} = m(X_i, X_j)$, $v_{ij} = v(X_i, X_j)$, $w_{ij} = w(X_i, X_j)$. When the current network is g and the attributes are $X = (X_1, \dots, X_n)$, the utility of player i is

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

The agent receives utility from the network, according to her own socioeconomic attributes and those of the individuals she links and have linked her. The utility of the network is the sum of the utility received from each link, and links are considered perfect substitutes. The benefit from an *additional link* has four components, which can be positive or negative, as a function of the socioeconomic characteristics of the pair involved in the relationship.

First, when the agent links another individual, she receives an additional *direct* net benefit u_{ij} .²⁴ The direct utility includes both costs and benefits and it may possibly be negative.²⁵ In many models this component is parameterized 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} for the net benefit that does not require assumptions on the cost function.

Second, an individual receives additional utility m_{ij} if the link is mutual. If agent i links j , she will value that link differently if j reciprocates the link or not. Intuitively, an agent can think of another individual as being her friend, but that person may not perceive the relationship in the same way and therefore that link is less valuable for i .

Moreover the agents value *indirect* links as well. When i is deciding to link j , she observes j 's friends and their socioeconomic characteristics. Each of j 's friend provides additional utility v_{ik} to the agent, according to their socioeconomic characteristics.²⁶ I assume that only friends of friends are valuable and they are considered as perfect substitutes. Individuals do not receive utility from two-links-away friends.

The fourth component corresponds to a *popularity effect*. When individual i links j , she automatically creates an indirect link for the individuals k 's that are linked to i . Thus i generates an externality (positive or negative, according to the socioeconomic attributes) for each k that has linked her in previous periods. This externality makes i more or less popular. For example, Albert is deciding whether to link Bob. Albert has already three friends and they are all classic musicians and have strong negative opinions about heavy-metal. Bob instead has his own heavy-metal band and does not appreciate classic music. The indirect link that Albert creates for his friends will give them negative utility. In this sense Albert becomes less popular.²⁷

²⁴ The direct component of the utility is common in all strategic network formation models.

²⁵ For example, homophily will lead agents to form links with people which are similar to them. This may be captured with a net benefit which is positive when the agents belong to the same group/type and negative (or smaller) when they belong to different ones.

²⁶ This is also common in all the strategic network formation models.

²⁷ This last component can be interpreted as introducing some reduced form *forward-looking* behavior in the model, since the popularity is referred to how more/less likely are the other agents to maintain or create a link to individual i in future meetings.

3.2 Equilibrium Analysis

In order to characterize the equilibria of the network formation game I impose some symmetry in the utility function.

While the assumptions are not too restrictive, I will show in the remaining of the paper that they provide a simple strategy to characterize: 1) the Nash networks and their existence; 2) the long-run dynamics; 3) the stationary distribution and transition probabilities and 4) the likelihood function. Furthermore, the symmetry restrictions simplify the simulation of the network formation model and the estimation strategy, both with the classical and bayesian approach.

I assume that the utility m_{ij} obtained from mutual links is symmetric: the mutual link between i and j brings the same additional utility to i and j . I also assume that the utility individual i receives from the popularity effect, w_{kj} is exactly equal to the *indirect* utility of individual k from the indirect link kj through i , i.e. v_{kj} .

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

$$m(X_i, X_j) = m(X_j, X_i) \text{ for all } i, j \in \mathcal{I} \quad (6)$$

$$w(X_k, X_j) = v(X_k, X_j) \text{ for all } k, j \in \mathcal{I} \quad (7)$$

therefore the utility function is

$$U_i(g, X) = \sum_{j=1}^n g_{ij} u_{ij} + \sum_{j=1}^n g_{ij} g_{ji} m_{ij} + \sum_{j=1}^n g_{ij} \sum_{\substack{k=1 \\ k \neq i, j}}^n g_{jk} v_{ik} + \sum_{j=1}^n g_{ij} \sum_{\substack{k=1 \\ k \neq i, j}}^n g_{ki} v_{kj} \quad (8)$$

The symmetry in m_{ij} does not imply that a mutual link between i and j will give both the same utility. Indeed if i and j have a mutual link, they receive the same common utility component (m_{ij}) but they may perceive that particular friendship in a different way, as long as the utility from direct or indirect links are different for i and j . In other words, two individuals with the same exogenous characteristics $X_i = X_j$ that form a mutual link receive the same u_{ij} and m_{ij} but may have different utilities because of their indirect links and popularity are different.

The second restriction is more technical. Notice that when i forms a link to j , i creates an externality for all k 's that have linked her. Any such individual k has now an additional indirect friend, i.e. j , that agent k will value an amount $v(X_k, X_j)$. The assumption means that individual i values his popularity effect as much as k values the indirect link to j , i.e. i internalizes the externality she creates.

The following proposition shows the main implication of Assumption 1 and it is one of the crucial results of this paper.

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

$$Q(g, X) = \sum_{i=1}^n \sum_{j=1}^n g_{ij} u_{ij} + \sum_{i=1}^n \sum_{j>i}^n g_{ij} g_{ji} m_{ij} + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{\substack{k=1 \\ k \neq i, j}}^n g_{ij} g_{jk} v_{ik} \quad (9)$$

and the network formation game is a Potential Game.

Proof. In Appendix C ■

The explanation of the results is simple. Under the restrictions of Assumption 1, for any player i and any link $g_{ij} \neq g'_{ij}$ we have

$$Q(g_{ij}, g_{-ij}, X) - Q(g'_{ij}, g_{-ij}, X) = U_i(g_{ij}, g_{-ij}, X) - U_i(g'_{ij}, g_{-ij}, X)$$

Consider two networks that differ only with respect to one link ij , chosen by individual i : the difference in utility that agent i receives from the two networks is exactly equal to the change in the *potential* function.

When the player receives higher utility from changing her linking decision, this is reflected in the potential. Therefore the potential is an aggregate function that summarizes the state of the network and the incentives of the players in each state.

The potential function is unique up to a constant: for any $c \in \mathbb{R}$, also $Q' = Q - c$ is a potential for the same game. I will exploit this property when simulating and estimating the model.

The potential $Q(g, X)$ is not equivalent to the welfare function $W(g, X)$,

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

In order to analyze the long run behavior of the model I impose more structure on the matching technology.

ASSUMPTION 2 (Meeting Process) *The meeting process m does not depend on the current network structure and any meeting is possible, i.e. for any $ij \in \mathcal{I} \times \mathcal{I}$*

$$\rho(g^{t-1}, X_i, X_j) = \rho(X_i, X_j) > 0 \tag{10}$$

The meeting process is such that any individual can be chosen and any pair of agents can meet. This assumption is needed to guarantee that any Nash network can be reached with positive probability. For example, a discrete uniform distribution satisfies this assumption.²⁸ The meeting process does not depend on the current network. This assumption rules out the possibility that individuals with more links have higher probability of meeting people. The probability of encounter with an agent is then exogenous.²⁹

The reason of the latter part of the assumption is that a meeting process that depends on the current network cannot be identified in the data, since we observe the network only once and

²⁸ A symmetric distribution such that $\rho_{ij}(X_i, X_j) > 0$ for any ij satisfies the assumption. An example is

$$\rho(X_i, X_j) \propto \frac{1}{n} \exp[-d(X_i, X_j)]$$

where $d(\cdot)$ is a distance function.

²⁹ Jackson and Rogers (2007) analyze a model in which the probability of meeting other individuals is related to the the number of links. In their model one can meet friends of friends.

we cannot observe meetings among individuals. Assuming an exogenous ρ_{ij} makes the long run network distribution independent from the meeting process.

It is useful to consider a *special case* of the model, in which there are no idiosyncratic shocks: the characterization of equilibria and long run behavior for such model provides more intuition about the dynamic 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.

$$\mathcal{N}(g) \equiv \{g' : g' = (g'_{ij}, g_{-ij}), \text{ for all } g'_{ij} \neq g_{ij}, \text{ for all } i, j \in \mathcal{I}\} \quad (11)$$

The following results characterize the set of the pure-strategy Nash equilibria and the long run behavior of the model without shocks.

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

1. *There exist at least one pure-strategy Nash 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.*

$$\mathcal{NE}(\mathcal{G}, X, U) = \left\{ g^* : g^* = \arg \max_{g \in \mathcal{N}(g^*)} Q(g, X) \right\} \quad (12)$$

3. *Any pure-strategy Nash equilibrium is an **absorbing** state.*
4. *As $t \rightarrow \infty$, the network converges to one of the Nash networks with probability 1*

Proof. In Appendix C ■

Intuitively, in a model without preferences shocks a Nash network is a network in which no player is willing to change her linking strategy, when drawn from the meeting process. This means that an agent that changes a link will receive less utility from the network. 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. Therefore the Nash network must be a local maximizer of the potential function.

Furthermore, the model will converge to one of the Nash Equilibria in the long run, independent of the initial network. Intuitively, suppose an agent is drawn from the meeting process. Such agent will play a best response, therefore his utility in that period cannot decrease. This holds for any player, thus the potential is nondecreasing over time. There is a finite number of states, therefore in the long run the sequence of networks will reach a local maximum of the potential, i.e. a Nash equilibrium.

I make a parametric assumption on the shocks that provides the characterization of the stationary distribution and the transition probabilities.³⁰

³⁰An alternative assumption is that shocks follow a Gaussian distribution. The extreme value assumption implies a nice characterization of the stationary distribution and a simpler simulation and estimation strategy, which I need to overcome the computational complexity as explained in the empirical section.

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

The assumption has several implications. First, it delivers a simple conditional choice probability for the evolution of the network, since it reduces to a logistic model with endogenous regressors. Second, it makes the model ergodic with a unique stationary distribution. Third, it allows me to derive the stationary distribution in closed form.

Using the distributional assumption it is possible to compute the probability of a link between i and j , given $m^t = ij$ and previous period network configuration g^{t-1}

$$\begin{aligned} \Pr \left(g_{ij}^t = 1 \mid g_{-ij}^{t-1}, X \right) &= \Pr \left[\varepsilon_{ij}^t(0) - \varepsilon_{ij}^t(1) \leq U_i \left(1, g_{-ij}^{t-1}, X \right) - U_i \left(0, g_{-ij}^{t-1}, X \right) \right] \\ &= \frac{\exp \left[u_{ij} + g_{ji}^{t-1} m_{ij} + \sum_{k \neq i, j} g_{jk}^{t-1} v_{ik} + \sum_{k \neq i, j} g_{ki}^{t-1} v_{kj} \right]}{1 + \exp \left[u_{ij} + g_{ji}^{t-1} m_{ij} + \sum_{k \neq i, j} g_{jk}^{t-1} v_{ik} + \sum_{k \neq i, j} g_{ki}^{t-1} v_{kj} \right]} \end{aligned} \quad (13)$$

Under assumptions 1-3, the chain evolves like a mixing Gibbs sampler with transition probability given by the conditional choice probabilities and mixing probability given by the probability law of m^t .

It can be shown that $[g^0, g^1, \dots, g^t]$ is:

1. *irreducible*, i.e. every state of the network can be reached with positive probability in a finite number of steps
2. *aperiodic*, i.e. the chain does not get trapped in cycles, because the probability of moving from a state to another is always positive under the extreme value assumption

Intuitively, assuming $P(m^t = ij) > 0$ for any ij , implies that there is always a positive probability to reach a new network in which the link g_{ij} may be updated. The logistic assumption implies that there is always a positive probability of switching to another state of the network, thus eliminating absorbing states.

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

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

$$\lim_{t \rightarrow \infty} P(G^t = g \mid G^0 = g^0, X) = \pi(g, X) \quad (14)$$

2. *The stationary distribution $\pi(g, X)$ is*

$$\pi(g, X) = \frac{\exp [Q(g, X)]}{\sum_{\omega \in \mathcal{G}} \exp [Q(\omega, X)]} \quad (15)$$

where $Q(g, X)$ is the potential function (9).

Proof. In Appendix C ■

The first part of the proposition follows directly from the irreducibility and aperiodicity of the Markov chain of networks. The uniqueness of the stationary distribution is useful in estimation, since there is no concern about multiple equilibria.

The second part provides a closed form solution for the stationary distribution. The interpretation is that the system of interacting agents in the long run will visit more often those states/networks that have high potential. This means that we should observe networks that are close to the Nash equilibria more often.

Notice that $\pi(g, X)$ includes a normalizing constant

$$c(\mathcal{G}, X) \equiv \sum_{\omega \in \mathcal{G}} \exp [Q(\omega, X)] \quad (16)$$

The normalizing constant complicates the estimation, since standard maximum likelihood estimation is unfeasible, as explained in next section.

4 Empirical Strategy

4.1 General Idea and Computational Complexity

I use the data from the Wave I of AddHealth, therefore the sample available is a network realization g_{obs} and the individual characteristics X of all students.

I operationalize the estimation by assuming that the observed network is drawn from the stationary distribution.

ASSUMPTION 4 (*Observed Network Stationarity*) *The observed network g_{obs} is drawn from the stationary distribution $\pi(g, X)$ in (15).*

In principle one could use Maximum Likelihood techniques to estimate the parameters of the model, since the joint distribution $\pi(g, X; \theta)$ is known. However MLE is *practically unfeasible*, due to the normalizing constant $c(\mathcal{G}, X)$, which requires an enormous amount of computations for each iteration of the maximization routine, even when the network is very small.

To be concrete, suppose that there are only $n = 10$ agents. This translates in a total of $2^{90} \simeq 1.2379 \times 10^{27}$ possible networks configurations.³¹ The exact computation of the constant is unfeasible.

In the remaining of this section, I explain the methodology used to tackle this estimation problem. I refer the interested reader to Appendix E for more technical details.³²

In extreme synthesis the strategy is:

1. Assume the utility components u , m and v are linear functions of a vector of parameters. This implies that the stationary distribution belongs to the exponential family with normalizing constants.

³¹This is the number of all possible 10×10 binary matrices, since by assumption any state of the network can occur with positive probability. In general a network with n agents has $2^{n(n-1)}$ possible configurations (states).

³²In a companion paper, I develop methods for Bayesian estimation of the model, using an approximated version of the Metropolis-Hastings algorithm, similar to the one proposed by Koskinen (2008).

2. Solve the system of non-linear equations implied by the first order condition of the likelihood maximization problem.
3. Use Markov Chain Monte Carlo (MCMC) methods to sample from the stationary distribution
4. Use methods to accelerate the MCMC algorithm, increase the mixing of the chains and decrease the variance of the estimates.

4.2 Likelihood Function

I assume that the utility functions u , m and v depend *linearly* on a vector of parameters. Define $\theta_u = (\theta_{u1}, \theta_{u2}, \dots, \theta_{uP})'$, $\theta_m = (\theta_{m1}, \theta_{m2}, \dots, \theta_{mL})'$ and $\theta_v = (\theta_{v1}, \theta_{v2}, \dots, \theta_{vS})'$. Define the function $H : \mathbb{R}^A \times \mathbb{R}^A \rightarrow \mathbb{R}$.

ASSUMPTION 5 (*Linearity of Utility*) *The utility functions are linear in parameters*

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

This assumption leaves room for many interesting specifications. In particular the functions H do not rule out interactions among different characteristics, for example interactions of race and gender of both individuals.

There is no constraint in the specification of u , m and v but linearity in parameters. Therefore we can consider different specifications and include different set of variables for the direct, mutual and indirect links.

The assumption delivers a very useful result, which I will exploit in the structural estimation of the model.

PROPOSITION 4 (*Exponential Family Likelihood*) *Under assumptions 1-5, the stationary distribution $\pi(g, X)$ belongs to the **exponential family**, i.e. it can be written in the form*

$$\pi(g, X) = \frac{\exp[\theta' \mathbf{t}(g, X)]}{\sum_{\omega \in \mathcal{G}} \exp[\theta' \mathbf{t}(\omega, X)]} \quad (17)$$

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. In Appendix C ■

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

4.3 Estimation Strategy

The parameters are estimated using Maximum Likelihood methods. The log-likelihood is denoted by $\ell(g_{obs}; \theta)$ and the parameters are solution of the maximization problem

$$\hat{\theta}_{mle} = \arg \max_{\theta \in \Theta} \ell(g_{obs}; \theta) \quad (18)$$

In Appendix B, I show that solving the maximization problem (18) is equivalent to solving the system of nonlinear equations

$$t(g_{obs}, X, \theta) = \mathbb{E}[t(\omega, X, \theta)] \quad (19)$$

The system corresponds to the first order conditions of (18). The vector $t(g_{obs}, X, \theta)$ is the vector of canonical sufficient statistics *observed* in the data.

I solve the system of equations using the Robins-Monro *stochastic approximation* algorithm (Robins and Monro, 1951).³³ The idea is to generate a Markovian sequence of parameters whose mean converges to the solution of the equation. In this sense it is a sequential estimation method and it can be thought of as a Monte Carlo version of the Newton-Raphson algorithm.

The Robins-Monro algorithm sequentially updates the parameters using the recursion

$$\hat{\theta}^{(r+1)} = \hat{\theta}^{(r)} - a_r D_0^{-1} [\mathbf{t}(g^{(r)}, X, \theta) - \mathbf{t}(g_{obs}, X)] \quad (20)$$

where the *gain* a_r is a sequence of positive numbers converging to zero as r goes to infinity (usually $1/r$), $\mathbf{t}(g^{(r)}, X, \theta)$ is the vector of sufficient statistics of a network $g^{(r)}$ sampled from the stationary distribution at the current parameter value $\hat{\theta}^{(r)}$, and

$$D_0 = \text{diag} \left[\frac{\partial \mathbb{E} [\mathbf{t}(\omega, X, \theta^{(0)}) - \mathbf{t}(g_{obs}, X)]}{\partial \theta} \right] = \text{diag} [COV(\mathbf{t}(g, X, \theta^{(0)}))] \quad (21)$$

is a diagonal matrix containing the variances of the sufficient statistics computed at the initial guess $\theta^{(0)}$.

It can be shown that the sequence of $\hat{\theta}^{(r)}$ is Markov, by the ergodic theorem the average $\hat{\theta} = (1/R) \sum_{r=1}^R \theta^{(r)}$ converges almost surely to the true parameter value

$$\hat{\theta} = \frac{1}{R} \sum_{r=1}^R \theta^{(r)} \rightarrow \theta \quad (22)$$

The algorithm follows the 3 steps suggested in Snijders (2002)

³³Kushner and Yin (2003) provide an extensive review of the literature on stochastic approximation and applications of the algorithm. Snijders (2002) proposes to use Robins-Monro for the estimation of Exponential Random Graph Models (ERGM).

1. Guess a θ^0 and estimate D_0 in (21) through MCMC methods (see Appendix E).
2. Use the estimated \hat{D}_0 in (20). For a given $\hat{\theta}^{(r)}$ simulate a network and iterate one step ahead.
3. Once convergence is achieved, use the estimated $\hat{\theta} = (1/R) \sum_{r=1}^R \theta^{(r)}$ to run a long MCMC and estimate the standard errors of the parameters as

$$Asy\widehat{V}(\theta_{mle}) = \hat{D}^{-1} = \left[COV(\widehat{\mathbf{t}}(g, X, \theta)) \right]^{-1} \quad (23)$$

This algorithm works well as long as the initial θ^0 is not too far from the solution. In some cases this procedure needs to be iterated several times to reach reasonable convergence.

In order to initialize the parameters for the simulations, I follow the suggestions in Besag (1974), Snijders (2002) and Nakajima (2007). I use the Maximum Pseudo-Likelihood Estimator, which maximizes the product of the conditional probabilities (details of the derivation can be found in Appendix B)

$$\log PL(g; \theta, X) = \theta' \sum_i \sum_j g_{ij} \Delta \mathbf{t}(g_{-ij}, X) - \sum_i \sum_j \log [1 + \exp(\theta' \Delta \mathbf{t}(g_{-ij}, X))]$$

5 Empirical Results

5.1 Individual Schools

SCHOOL 25

I estimate several specifications for the network in Figure 1(a). This school has 150 students and a total of 355 friendship nominations. The school has 58.667% females, and a racial fragmentation of 0.606³⁴

Table 3 shows the estimated parameters for a model containing only the utility from direct links. This specification is in the spirit of dyadic regression (Fafchamps and Gubert, 2006; Bramoullé and Fortin, 2009) and the estimation is carried out using maximum likelihood.³⁵ It is equivalent to a logistic regression.

The estimated parameters indicate the marginal effect of the variable indicated on the utility function. For example, consider the gender dummies. The variable *male* \rightarrow *male* is a dummy that indicates if i and j are both males. The omitted category for the gender dummies is the indicator *female* \rightarrow *male* link. A positive coefficient for *male* \rightarrow *male* indicates that a male individual prefers to link another male instead of linking a female. Therefore the sign is useful in interpreting the direction of the effect.

Since this is a logistic regression the coefficients can also be interpreted in terms of odds-ratios.

Racial homophily is pervasive, as indicated by a positive and significant coefficient for same-race links. Students prefer to link individuals of same race and prefer to avoid links to other racial groups, as indicated by a negative coefficient for different race links. This result is robust in each of the columns 2-4.

The homophily effect is stronger for hispanics suggesting that a group may tend to over-segregate when it is minority. This is consistent with structural estimates from recent studies of

³⁴The racial composition is as follows: 42% whites, 45.3% blacks, 0.667% asians, 10.6% hispanics.

³⁵Notice that in this case maximum likelihood estimation and maximum pseudolikelihood are equivalent.

Table 3: Simple Logistic Model, School 25

	[1]	[2]	[3]	[4]
constant	-3.859** (0.080)	-4.382** (0.303)	-4.060** (0.314)	-4.000** (0.515)
male → male	-0.114 (0.144)		-0.125 (0.146)	-0.123 (0.147)
male → female	-0.657** (0.153)		-0.661** (0.155)	-0.658** (0.155)
female → female	-0.559** (0.148)		-0.564** (0.149)	-0.563** (0.150)
white → white		0.923** (0.317)	0.862** (0.319)	0.864** (0.319)
white → black		-1.204** (0.393)	-1.254** (0.394)	-1.251** (0.395)
white → hispanic		-0.736 (0.510)	-0.786 (0.510)	-0.783 (0.511)
black → white		-1.781** (0.451)	-1.825** (0.452)	-1.823** (0.452)
black → black		1.027** (0.314)	0.994** (0.315)	0.996** (0.315)
black → hispanic		-0.813 (0.510)	-0.849 (0.510)	-0.846 (0.510)
hispanic → white		-0.125 (0.429)	-0.169 (0.430)	-0.167 (0.430)
hispanic → black		-0.813 (0.510)	-0.849 (0.510)	-0.848 (0.510)
hispanic → hispanic		1.454** (0.416)	1.421** (0.417)	1.422** (0.417)
grade difference				-0.003 (0.023)
grade sum				-0.003 (0.023)

Note: Logistic regression for school 25 in AddHealth. Estimation method is Maximum Likelihood. The dependent variable is the link g_{ij} and variables indicate the exogenous characteristics of individuals. The variable *hispanic* → *black* indicates that i is hispanic while j is black. ** indicates significance at the 5% level. Standard errors in parentheses.

homophily that consider only direct links (Currarini, Jackson and Pin, 2009).

Gender homophily is not present, since the coefficients are all negative. Students prefer to link kids that are enrolled in a similar grade and older students form fewer links.

These results are in line with most of the empirical evidence on homophily. However when we assume that individual receive utility also from mutual, indirect and popularity links, results may change. This model allows me to test if there is homophily also in the mutual and indirect links. I present these results in Table 4. I compare the results of the full model with the ones in Table 3.

Table 4 reports the estimated parameters for the utility relative to direct links. The first important result is that, when controlling for the effect of mutual, indirect and popularity effects, *the magnitude of the direct utility effects is attenuated* for most of the variables and it becomes less significant.

Racial and gender homophily play still a major role in the direct links. However a link between a white and an Hispanic student is estimated to give positive utility, although not significant. This effect has the same magnitude as the utility of a link between two white individuals, which is

Table 4: Full Model, School 25

	Utility Component		
	Direct	Mutual	Indirect
constant	-3.932** (1.372)	-6.507** (2.699)	-0.415** (0.185)
male → male	0.132 (0.433)	0.697 (1.021)	-0.032 (0.061)
male → female	0.017 (0.338)		-0.039 (0.069)
female → female	-0.316 (0.327)	1.219 (0.977)	0.102 (0.065)
white → white	0.181 (0.514)	-0.975 (1.076)	0.276** (0.098)
white → black	-1.099** (0.473)		0.146 (0.108)
white → hispanic	0.248 (0.837)		-0.866** (0.459)
black → white	-1.630** (0.633)		0.078 (0.116)
black → black	0.440 (0.585)	-0.419 (0.936)	0.211 (0.101)
black → hispanic	-0.029 (0.977)		-0.139 (0.299)
hispanic → white	-0.742 (0.707)		0.342** (0.124)
hispanic → black	-0.291 (0.836)		-0.141 (0.153)
hispanic → hispanic	1.769 (1.201)	-2.096 (9.266)	0.036 (0.728)
grade difference	0.143** (0.066)	-0.697 (0.236)	-0.032** (0.007)
grade sum	-0.014 (0.065)	-0.148 (0.129)	-0.009 (0.009)

Note: Structural estimate of the network formation model for School 25 in AddHealth. Estimation through Method of Moments using Robins-Monro stochastic approximation algorithm. Each variable indicate the exogenous characteristics of individuals. For example, the variable *hispanic* → *black* indicates that *i* is hispanic while *j* is black. Each column represent a different part of the utility: direct, mutual and indirect/popularity. ** indicates significance at the 5% level. Standard errors in parentheses.

surprising. At the same time a link between a male student and a female provides positive utility. The effect of the grade is inverted: individual prefer to link with people enrolled in much higher or much lower grades, and older students have fewer links.

The second column of Table 4 reports the mutual utility parameters. While homophily is relevant for gender, we observe the opposite effect for race: a mutual link with same-race individuals actually decreases utility. This result is somewhat puzzling and the magnitudes of the coefficients seem too high. An interpretation is that individual value heterogeneity in mutual links, since they care about the total number of mutual links instead of their composition.

The third column of Table 4 shows the most interesting results for the indirect links. This part of the utility does not seem completely driven by homophily, but instead it appears that individuals value a certain heterogeneity in their indirect links. Therefore white students appreciate a set of indirect links that are blacks, african american students prefer white indirect friends, and hispanic

Table 5: Simple Logistic Model, School 94

	[1]	[2]	[3]	[4]
constant	-2.849** (0.169)	-3.586** (0.306)	-3.432** (0.375)	-9.650** (1.906)
male → male	-0.232 (0.248)		-0.045 (0.261)	-0.187 (0.266)
male → female	-0.630** (0.279)		-0.548 (0.285)	-0.658** (0.288)
female → female	-0.160 (0.246)		-0.053 (0.253)	-0.093 (0.255)
white → white		1.398** (0.333)	1.381** (0.345)	1.525** (0.349)
white → black		-1.707 (1.048)	-1.717 (1.053)	-1.568 (1.054)
white → hispanic		-0.306 (0.470)	-0.345 (0.476)	-0.170 (0.481)
black → white		-0.599 (0.657)	-0.636 (0.665)	-0.465 (0.667)
black → black		0.878 (0.600)	0.851 (0.610)	1.041 (0.615)
black → hispanic		0.574 (0.518)	0.522 (0.525)	0.736 (0.532)
hispanic → white		-0.306 (0.470)	-0.292 (0.476)	-0.092 (0.482)
hispanic → black		0.878 (0.476)	0.908 (0.484)	1.128** (0.492)
hispanic → hispanic		0.483 (0.434)	0.470 (0.438)	0.719 (0.448)
grade difference				0.154 (0.118)
grade sum				0.403 (0.120)

Note: Logistic regression for school 94 in AddHealth. Estimation method is Maximum Likelihood. The dependent variable is the link g_{ij} and variables indicate the exogenous characteristics of individuals. The variable *hispanic* → *black* indicates that i is hispanic while j is black. ** indicates significance at the 5% level. Standard errors in parentheses.

students value a set of white indirect friends.

In summary, for school 25, homophily plays a prominent role in driving the network structure only for direct links. Students prefer to form direct friendship relationships with similar individuals. On the other hand, a set of heterogeneous mutual and indirect friends gives more utility than an homogeneous group.

SCHOOL 94

School 94 is smaller than previous school: it has only 53 students enrolled and they form a total of 123 links. The racial composition is similar to school 25 and the fragmentation index is 0.661.³⁶

The results of the simple logistic regressions are reported in Table 5. Racial homophily is present in this school too, as indicated by the positive coefficients for same race links. However there seem to be a tendency of blacks and hispanic students to link each other.

The full model results are shown in Table 6. In this school, students have positive utility from

³⁶The racial composition is as follows: 47% whites, 15.1% blacks, 3.77% asians, 30.6% hispanics.

Table 6: Full Model, School 94

	Utility Component		
	Direct	Mutual	Indirect
constant	-4.954 (3.008)	11.506 (9.036)	-1.678** (0.438)
male → male	-0.458 (0.476)	0.928 (1.136)	0.121 (0.134)
male → female	-0.601 (0.438)		0.257 (0.199)
female → female	-0.248 (0.423)	0.201 (1.347)	0.293 (0.166)
white → white	0.529 (0.481)	0.98 (1.008)	0.193 (0.109)
white → black	-1.082 (0.811)		-0.303 (0.239)
white → hispanic	0.02 (0.618)		-0.035 (0.195)
black → white	0.73 (0.734)		-0.81** (0.352)
black → black	0.284 (0.885)		-0.08 (0.512)
black → hispanic	0.281 (0.878)		0.512 (0.288)
hispanic → white	0.83 (0.686)		-0.425** (0.215)
hispanic → black	0.698 (0.768)		0.42 (0.319)
hispanic → hispanic	0.441 (0.763)	-0.357 (2.267)	-0.042 (0.258)
grade difference	0.385 (0.221)	-0.045 (0.72)	-0.046 (4.381)
grade sum	0.122 (0.188)	-0.619 (0.588)	0.09 (3.024)

Note: Structural estimate of the network formation model for School 94 in AddHealth. Estimation through Method of Moments using Robins-Monro stochastic approximation algorithm. Each variable indicate the exogenous characteristics of individuals. For example, the variable *hispanic* → *black* indicates that *i* is hispanic while *j* is black. Each column represent a different part of the utility: direct, mutual and indirect/popularity. ** indicates significance at the 5% level. Standard errors in parentheses.

direct links to any racial group: only a link from a white to a black student provides negative utility. There is homophily in some of the mutual links, both for gender and race, but mutual links among hispanic reduce utility. Indirect links show more heterophily.

It must be noted that the estimates are not very precise, since we obtain large standard errors. This is probably due to the small sample size, since this school contains only 53 people and 123 links.

5.2 Full Sample

SCHOOLS 25,32 and 94

I show estimates for a network composed of 3 schools. I chose the schools because of their racial composition similarity. I use school 25, 32 and 94 in estimation, assuming that each school network

is independent. Having more than a school allows me to control for school-specific characteristics that cannot be included in the single school specifications.

Table 7 shows a specification with individual controls but no school-specific effects. In this

Table 7: Full Model, School 25,32 and 94

	Utility Component		
	Direct	Mutual	Indirect
constant	0.165 (1.201)	3.171 (5.431)	-0.659 (0.942)
male → male	-2.46 (0.413)	-2.994 (6.486)	0.117 (0.307)
male → female	-3.695** (0.706)		0.091 (0.338)
female → female	-2.059** (0.444)	1.047 (2.048)	0.204 (0.267)
white → white	0.647 (0.71)	-0.724 (2.564)	0.146 (0.445)
white → black	-1.012 (0.966)		-0.008 (0.428)
white → hispanic	-1.216 (0.873)		-0.221 (0.6)
black → white	-0.732 (0.817)		-0.173 (0.494)
black → black	0.401 (0.674)	-0.239 (2.567)	0.071 (0.36)
black → hispanic	-4.418** (1.219)		0.034 (0.498)
hispanic → white	-1.55 (0.82)		-0.036 (0.68)
hispanic → black	-7.466** (1.528)		-0.078 (0.651)
hispanic → hispanic	-4.714** (1.329)	-20.235 (3633.072)	-0.055 (0.806)
grade difference	0.139 (0.081)	-0.726 (0.948)	-0.027 (0.057)
grade sum	-0.343** (0.06)	-0.011 (0.28)	0.03 (0.042)

Note: Structural estimate of the network formation model for School 25, 32 and 94 in AddHealth. Estimation through Method of Moments using Robins-Monro stochastic approximation algorithm.

Each variable indicate the exogenous characteristics of individuals.

For example, the variable *hispanic* → *black* indicates that *i* is hispanic while *j* is black.

Each column represent a different part of the utility: direct, mutual and indirect/popularity.

** indicates significance at the 5% level. Standard errors in parentheses.

specification there is evidence of homophily in direct links for whites and blacks, but not for hispanics. For indirect links, there is mixed evidence: the estimates for whites seem to point to homophily, while this is not the case for blacks and hispanics.

The mutual link between hispanics gives very large negative utility in this specification and this could point to some heterogeneity among schools.

In Table 8 I use the same specification and add controls for the racial composition in each school in the direct utility component. In particular I use the shares of whites and blacks as controls for specific school characteristics.

An alternative specification could introduce a school fixed effect and interactions between the

Table 8: Full Model with school shares, School 25,32 and 94

	Utility Component		
	Direct	Mutual	Indirect
constant	-11.516 (10.753)	3.967 (7.651)	-0.223 (1.686)
male → male	-0.214 (1.381)	0.682 (1.778)	0.113 (0.622)
male → female	-0.28 (1.222)		0.059 (0.591)
female → female	-0.364 (1.262)	1.222 (1.915)	0.154 (0.557)
white → white	0.42 (1.834)	-0.615 (2.441)	0.209 (0.568)
white → black	-0.919 (3.356)		0.062 (1.015)
white → hispanic	0.159 (2.473)		-0.351 (1.031)
black → white	-0.655 (3.204)		-0.15 (1.231)
black → black	0.645 (2.637)	-0.002 (3.289)	0.12 (0.817)
black → hispanic	-0.108 (2.427)		-0.025 (1.397)
hispanic → white	0.161 (2.285)		-0.05 (0.856)
hispanic → black	0.1 (3.142)		-0.112 (1.049)
hispanic → hispanic	1.193 (1.891)	-0.886 (3.922)	-0.109 (1.37)
grade difference	0.144 (0.39)	-0.505 (1.086)	-0.031 (0.144)
grade sum	-0.006 (0.324)	-0.039 (0.439)	0.002 (0.104)
share whites	0.176 (0.127)		
share blacks	0.001 (0.052)		

Note: Structural estimate of the network formation model for School 25, 32 and 94 in AddHealth. Estimation through Method of Moments using Robins-Monro stochastic approximation algorithm.

Each variable indicate the exogenous characteristics of individuals.

For example, the variable *hispanic* → *black* indicates that *i* is hispanic while *j* is black.

Each column represent a different part of the utility: direct, mutual and indirect/popularity.

** indicates significance at the 5% level. Standard errors in parentheses.

individual controls and school characteristics. A model with school varying coefficients could also be estimated.

Considering the direct utility estimates, one notice that some of the coefficients change sign with respect to results in Table 7. In particular the model suggests now that hispanic students like to link other hispanic students, but they have positive utility also from links to other racial groups. These estimates are reversed with respect to the model without school controls.

Indirect links display homophily effects, with the exception of links between white and blacks or between hispanics.

Future research will incorporate more schools and school fixed effect, both in levels and inter-

acted with individual controls.

5.3 Policy Experiments

I use the model to study the effects of different counterfactual policies whose goal is to promote interracial contact. Several of these policies are in place in metropolitan areas of the US (see Angrist and Lang, 2004 for an example).

I consider three different policies

1. Policy 1: assign the African American students of school 94 to school 25
2. Policy 2: assign the Hispanic students of school 94 to school 25
3. Policy 3: assign the African American and Hispanic students of school 94 to school 25

Table 9: Policy Experiments

	Average per capita welfare	Variance per capita welfare	Difference
No Busing	-165.1719	16.40598	
Policy 1	-173.2276	27.93296	-8.0557**
Policy 2	-176.8941	27.39433	-11.7222**
Policy 3	-177.3898	25.89589	-12.2179**

I compare the average welfare of all students in the four different scenarios and the results of my simulations are summarized in Table 9.³⁷

The outcome clearly points to a significant decrease of welfare after the implementation of the policies. Notice that I am comparing stationary distributions and therefore abstracting from the dynamics effect of the policy during the transition to the new stationary distribution.

An alternative visualization is shown in Figure 2 where I show the distribution of welfare in the four different stationary distribution. The decrease in welfare is very clear from these pictures.

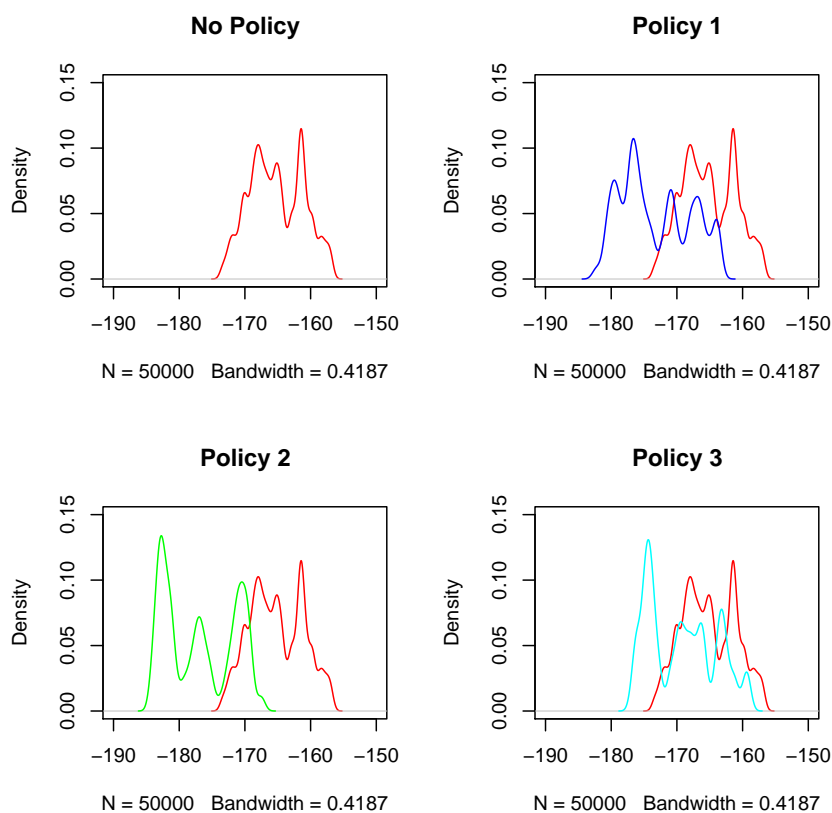
6 Conclusions

This paper studies the determinants of segregation in social networks, focusing on preferences for same race individuals. I develop a dynamic model of social network formation in which heterogeneous agents form links and receive utility from each friend according to their socioeconomic characteristics. I estimate the structural preference parameters driving the process of link formation among agents, using Markov Chain Monte Carlo methods and stochastic approximation.

I use data from Add Health, a unique dataset containing the *actual* social network of a representative sample of US high schools.

³⁷I start the simulation at the observed network and run the Simulated Tempering for 100000 iterations so that the chains reach the new stationary distribution. Then I draw 50000 networks from the Simulated Tempering sampler.

Figure 2: Policy Experiments



Distribution of Welfare in each of the policies. The red distribution refers to the scenario without any policy. The blue is relative to a re-assignment of African-American students, the green to a re-assignment of Hispanic students, the light blue to a re-assignment of both minority groups. The distribution of welfare is computed in the stationary distribution.

My results confirm that racial homophily is a pervasive phenomenon: students prefer to be friends with individuals of their same racial group. I estimate a model in which individuals only care about the direct links they form with other people: such a model provides strong support for homophily.

However, when I consider a model in which individuals care also about mutual links, the characteristics of friends of friends, and their popularity among other students, results change. Homophily is very important for the direct links, but *individuals prefer a more heterogeneous set of mutual and indirect links*. Previous empirical analysis failed to detect this effect since only the direct links were modeled explicitly.

I use the model to study counterfactual policy experiments and their effect on the welfare of students in the stationary equilibrium. I simulate a busing program that transport minority students to a school in which most of the students are non-minority.³⁸ The simulations of the model

³⁸See Angrist and Lang (2004) for the analysis of one of this programs.

suggest that such policies may decrease welfare of all the students.

References

- [1] Bala, Venkatesh and Goyal, Sanjeev (2000), "A Noncooperative Model of Network Formation", *Econometrica* 68 (5), 1181-1229
- [2] Besag, Julian (1974), "Spatial Interaction and the Statisticsn Analysis od Lattice Systems", *Journal of the Royal Statistical Society. Series B (Methodological)* 36(2):192-236
- [3] Breuckner, Jan (2006), "Friendship Networks," *Journal of Regional Science* 46, 847-865
- [4] Conley, Timothy and Udry, Christopher (2004), "Social Network in Ghana", draft, Yale University
- [5] De Giorgi, Giacomo, Pellizzari, Michele and Redaelli, Silvia (2007), "Be as Careful of the Books You Read as of the Company You Keep. Evidence on Peer Effects in Educational Choices", IZA Discussion Paper No. 2833
- [6] Fafchamps, Marcel and Gubert, Flore (2006), "The Formation of Risk Sharing Networks", draft, University of Oxford
- [7] Frank, Ove, Strauss, David, (1986), "Markov graphs", *Journal of the American Statistical Association* 81:832-842.
- [8] Geyer, Charles J. and Elizabeth A. Thompson (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data", *Journal of the Royal Statistical Society. Series B (Methodological)* 54(3):657-699.
- [9] Goldstein, Markus and Udry, Chrsitopher (1999), "Agricultural Innovation and Resource Management in Ghana", draft, Yale University
- [10] Hoxby, Caroline M. (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation", NBER Working Paper 7867
- [11] Jackson, Matthew and Wolinski, Asher (1996), "A Strategic Model of Social and Economic Networks" *Journal of Economic Theory* 71(1):44-74
- [12] Jackson, Matthew and Watts, Allison (2002), "The Evolution of Social And Economic Networks", *Journal of Economic Theory* 106(2):265-295
- [13] La Ferrara, Eliana (2003), "Kin groups and Reciprocity: A Model of Credit Transactions in Ghana", *American Economic Review*, 93(5), 1730-1751
- [14] Manski, Charles (1993), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies* 60(3):531-42
- [15] Monderer, Dov and Shapley, Lloyd S. (1996), "Potential Games", *Games and Economic Behavior* 14:124-143

- [16] Munshi, Kaivan and Rosenzweig, Mark, (2006), "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *American Economic Review* 96(4):1225-1252.
- [17] Munshi, Kaivan (2003), "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." 2003. *Quarterly Journal of Economics* . 118(2): 549-597.
- [18] Nakajima, Ryo (2007), "Measuring Peer Effects on Youth Smoking Behavior" *forthcoming* in *Review of Economic Studies*.
- [19] Robert, Christian and Casella, George (1999), "Monte Carlo Statistical Methods", Springer-Verlag
- [20] Robins, Garry, Elliott, Peter and Pattison, Philippa (2001), "Network Models for Social Selection Processes", *Social Networks* 23:1-30
- [21] Sacerdote, Bruce (2001), Peer Effects with Random Assignment: Results for Dartmouth Roommates, *Quarterly Journal of Economics*, 116, 681-704
- [22] Snijders, Tom A.B, (2002) "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models", *Journal of Social Structure* 3(2)
- [23] Wasserman, Stanley and Pattison, Philippa (1996) "Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*", *Psychometrika* 61(3):401-425

A Equilibrium Analysis for the homogenous model

Let's consider a simplified version of the model, in which all the agents are of the same type. Assume $u_{ij} = u$, $m_{ij} = m$, $v_{ij} = v$ and $w_{ij} = w$ for all i, j . The utility is then

$$\begin{aligned}
U_i(g) &= u \sum_{j=1}^n g_{ij} + m \sum_{j=1}^n g_{ij} g_{ij} + v \sum_{j=1}^n g_{ij} \sum_{\substack{k=1 \\ k \neq i, j}}^n g_{jk} + w \sum_{j=1}^n g_{ij} \sum_{\substack{k=1 \\ k \neq i, j}}^n g_{ki} \\
&= u d_i^{out} + m d_i^{mut} + v \sum_{j=1}^n g_{ij} (d_j^{out} - g_{ji}) + w \sum_{j=1}^n g_{ij} (d_i^{in} - g_{ji}) \\
&= u d_i^{out} + m d_i^{mut} + v \sum_{j=1}^n g_{ij} d_j^{out} - v \sum_{j=1}^n g_{ij} g_{ji} + w \sum_{j=1}^n g_{ij} d_i^{in} - w \sum_{j=1}^n g_{ij} g_{ji} \\
&= u d_i^{out} + (m - v - w) d_i^{mut} + v \sum_{j=1}^n g_{ij} d_j^{out} + w d_i^{in} d_i^{out}
\end{aligned}$$

where $d_i^{out} = \sum_{j=1}^n g_{ij}$ is the out-degree of agent i , $d_i^{mut} = \sum_{j=1}^n g_{ij} g_{ji}$ is the number of mutual links of agent i , $d_i^{in} = \sum_{k=1}^n g_{ki}$ is the in-degree of agent i .

Let's consider a specification of the model in which $u = b - c$ so that each direct link has a cost c and a benefit of b , and $v = w$ as in Assumption 1.

The Nash networks are characterized in the next proposition.

PROPOSITION 5 *Assume $u = b - c$ and $w = v$.*

1. *If $c < u + m + 2(n - 2)v$ then the complete network is a Nash network.*
2. *If $c > b$ then the empty network is a Nash network.*

Proof. The complete network is a Nash equilibrium if any player is not willing to sever any link. The utility of a player when the network is complete is

$$U_i(g) = (u - c)(n - 1) + (m - 2v)(n - 1) + 2v(n - 1)(n - 1)$$

and when the player cuts one link it becomes

$$U_i(g) = (u - c)(n - 2) + (m - 2v)(n - 2) + 2v(n - 1)(n - 2)$$

Therefore the player will not sever the link if $(u - c) + (m - 2v) + 2v(n - 1) > 0$ or if

$$c < u + m + 2v(n - 2)$$

■

In the next proposition I show sufficient conditions for the empty network or the complete network to be unique equilibria. In essence the uniqueness results are based on the relative importance of mutual links and indirect links for the players.

PROPOSITION 6 *Assume $u = b - c$ and $w = v$.*

1. *The empty network is the unique equilibrium if*
 - *$m > 2v$ and $c > b + m + 2v(n - 2)$*
 - *$m < 2v$ and $c > b + v(2n - 3)$*
2. *The complete network is the unique equilibrium if*
 - *$m < v$ and $c < b + m - v$*
 - *$m > v$ and $c < b$*

Proof. PROOF TO BE COMPLETED HERE ■

TO BE COMPLETED

B Computational and Statistical Properties used in Estimation

I have data for multiple schools. I assume that each school is an independent network. Therefore each school is a draw from the stationary distribution

$$\pi(g_c, X_c) = \frac{\exp[\theta' \mathbf{t}(g_c, X_c)]}{\sum_{\omega \in \mathcal{G}} \exp[\theta' \mathbf{t}(\omega_c, X_c)]} \quad (24)$$

The likelihood for the entire network is therefore

$$\pi(g, X) = \prod_{c=1}^C \pi(g_c, X_c) \quad (25)$$

and the log-likelihood is

$$\log \pi(g, X) = \sum_{c=1}^C \log \pi(g_c, X_c) \quad (26)$$

$$= \theta' \sum_{c=1}^C \mathbf{t}(g_c, X_c) - \sum_{c=1}^C \log \mathcal{K}(g_c, X_c, \theta) \quad (27)$$

$$= \theta' \sum_{c=1}^C \mathbf{t}(g_c, X_c) - \log \mathcal{K}(g, X, \theta) \quad (28)$$

Notice that the stationary distribution can be rewritten this way

$$\pi(g, X) = \frac{\exp[Q(g, X)]}{\sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X)]} \quad (29)$$

$$= \frac{\exp[Q(g, X)] \exp[Q(g_{obs}, X)]}{\sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X)] \exp[Q(g_{obs}, X)]} = \quad (30)$$

$$= \frac{\exp[Q(g, X) - Q(g_{obs}, X)]}{\sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X) - Q(g_{obs}, X)]} \quad (31)$$

This allows me to have a probability of observed network equal to

$$\pi(g_{obs}, X) = \left\{ \sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X) - Q(g_{obs}, X)] \right\}^{-1}$$

and the observed canonical statistics are normalized as

$$\mathbf{t}(g_{obs}, X) = \mathbf{0}$$

B.1 Simulation of the Model

The dynamic of the model is the same as the *Gibbs Sampler*. In order to sample networks from the stationary distribution one can simulate the model ahead until convergence to the stationary distribution and then collect several realizations of the network.

This is clearly the simplest way to sample from the stationary distribution. However, notice that any markov chain that has stationary distribution π can be used for the simulations. One can use a *Metropolis-Hasting* algorithm for example. In alternative one can use algorithms that update more than one link per period, as long as the detailed balance condition is satisfied. Inversion methods can be used (see Snijders (2002) for details) to decrease the variability of the estimates.

It is convenient to simulate odds ratios instead of conditional probabilities, given the linearity assumed in assumption 4. Indeed the odds ratio is

$$\frac{\Pr(g_{ij} = 1 | g_{-ij}, X)}{\Pr(g_{ij} = 0 | g_{-ij}, X)} = \frac{\frac{\Pr(g_{ij} = 1, g_{-ij}, X)}{\Pr(g_{-ij}, X)}}{\frac{\Pr(g_{ij} = 0, g_{-ij}, X)}{\Pr(g_{-ij}, X)}} = \frac{\Pr(g_{ij} = 1, g_{-ij}, X)}{\Pr(g_{ij} = 0, g_{-ij}, X)} \quad (32)$$

$$= \frac{\pi(g_{ij} = 1, g_{-ij}, X)}{\pi(g_{ij} = 0, g_{-ij}, X)} = \frac{\exp[Q(g_{ij} = 1, g_{-ij}, X)]}{\exp[Q(g_{ij} = 0, g_{-ij}, X)]} \quad (33)$$

$$= \exp[Q(g_{ij} = 1, g_{-ij}, X) - Q(g_{ij} = 0, g_{-ij}, X)] \quad (34)$$

$$= \exp\{\theta' [\mathbf{t}(g_{ij} = 1, g_{-ij}, X) - \mathbf{t}(g_{ij} = 0, g_{-ij}, X)]\} \quad (35)$$

$$= \exp\{\theta' \Delta \mathbf{t}(g_{-ij}, X)\} \quad (36)$$

where $\Delta \mathbf{t}(g_{-ij}, X)$ indicates the change in the canonical statistics from a network in which $g_{ij} = 1$ to a network in which $g_{ij} = 0$.

It follows that

$$\log \left[\frac{\Pr(g_{ij} = 1 | g_{-ij}, X)}{\Pr(g_{ij} = 0 | g_{-ij}, X)} \right] = \theta' \Delta \mathbf{t}(g_{-ij}, X) \quad (37)$$

which is linear and therefore simpler to compute in each step of the simulation.

B.2 First order Conditions for Maximum Likelihood

I start from the likelihood function computed at the observed network g_{obs}

$$\pi(g_{obs}, X; \theta) = \frac{\exp[\sum_k \theta_k t_k(g_{obs}, X)]}{\sum_{\omega \in \mathcal{G}} \exp[\sum_k \theta_k t_k(\omega, X)]}$$

Therefore the first order conditions are

$$\begin{aligned} t_k(g_{obs}, X) \sum_{\omega \in \mathcal{G}} \exp \left[\sum_k \theta_k t_k(\omega, X) \right] &= \sum_{\omega \in \mathcal{G}} t_k(\omega, X) \exp \left[\sum_k \theta_k t_k(\omega, X) \right] \\ t_k(g_{obs}, X) &= \sum_{\omega \in \mathcal{G}} t_k(\omega, X) \frac{\exp[\sum_k \theta_k t_k(\omega, X)]}{\sum_{\omega \in \mathcal{G}} \exp[\sum_k \theta_k t_k(\omega, X)]} \\ t_k(g_{obs}, X) &= E[t_k(\omega, X; \theta)] \end{aligned}$$

Notice that from the log-likelihood equation

$$\log \pi(g_{obs}, X) = \sum_k \theta_k t_k(g_{obs}, X) - \log c(\mathcal{G}, X, \theta)$$

we obtain first order conditions

$$t_k(g_{obs}, X) - \frac{1}{c(\mathcal{G}, X, \theta)} \frac{\partial c(\mathcal{G}, X, \theta)}{\partial \theta_k} = 0$$

and so it follows that

$$E [t_k (\omega, X; \theta)] = \frac{1}{c(\mathcal{G}, X, \theta)} \frac{\partial c(\mathcal{G}, X, \theta)}{\partial \theta_k}$$

Suppose the researcher has an i.i.d. sample of networks, $g^{(1)}, \dots, g^{(R)}$, drawn from the stationary distribution. The likelihood of the sample is

$$L = \prod_{r=1}^R \frac{\exp [\sum_k \theta_k t_k (g^{(r)}, X)]}{\sum_{\omega \in \mathcal{G}} \exp [\sum_k \theta_k t_k (\omega, X)]} = \frac{\exp [\sum_k \theta_k \sum_r t_k (g^{(r)}, X)]}{c(\mathcal{G}, X, \theta)^R}$$

$$L = \prod_{r=1}^R \frac{\exp \left[\sum_{k=1}^K \theta_k t_k (g^{(r)}, X) \right]}{\sum_{\omega \in \mathcal{G}} \exp \left[\sum_{k=1}^K \theta_k t_k (\omega, X) \right]} = \frac{\exp \left[\sum_{k=1}^K \theta_k \sum_{r=1}^R t_k (g^{(r)}, X) \right]}{c(\mathcal{G}, X, \theta)^R}$$

taking logs

$$\log L = \sum_k \theta_k \sum_r t_k (g^{(r)}, X) - R \log c(\mathcal{G}, X, \theta)$$

$$\log L = \sum_{k=1}^K \theta_k \sum_{r=1}^R t_k (g^{(r)}, X) - R \log c(\mathcal{G}, X, \theta)$$

and taking the first order conditions we get

$$\sum_{r=1}^R t_k (g^{(r)}, X) - R \frac{1}{c(\mathcal{G}, X, \theta)} \frac{\partial c(\mathcal{G}, X, \theta)}{\partial \theta_k} = 0$$

$$\frac{1}{R} \sum_{r=1}^R t_k (g^{(r)}, X) = \frac{1}{c(\mathcal{G}, X, \theta)} \frac{\partial c(\mathcal{G}, X, \theta)}{\partial \theta_k}$$

and therefore we conclude that

$$E [t_k (\omega, X; \theta)] = \frac{1}{R} \sum_{r=1}^R t_k (g^{(r)}, X; \theta)$$

In order to find the MLE parameters we can solve the system

$$t(g_{obs}, X) = E [t(\omega, X; \theta)] = \frac{1}{R} \sum_{r=1}^R t(g^{(r)}, X)$$

The sample of networks is generated using a Markov Chain Monte Carlo method.

B.3 Derivation of the PseudoLikelihood

In order to perform the estimation procedure I need to initialize the vector of parameters θ . The choice of the initial value is not trivial, but I follow the suggestions of Geyer and Thompson (1992) and Snijders (2002), using the Pseudo-Likelihood estimate.³⁹

I derive here the Pseudo-Likelihood for the model.

$$\begin{aligned}
PL(g; \theta, X) &= \prod_{ij} \Pr(G_{ij} = g_{ij} | G_{-ij} = g_{-ij}, X) \\
&= \prod_{ij} \frac{\exp \left[g_{ij} \left(u_{ij} + g_{ji} m_{ij} + \sum_{k \neq i, j} g_{jk} v_{ik} + \sum_{k \neq i, j} g_{ki} v_{kj} \right) \right]}{1 + \exp \left[u_{ij} + g_{ji} m_{ij} + \sum_{k \neq i, j} g_{jk} v_{ik} + \sum_{k \neq i, j} g_{ki} v_{kj} \right]} \\
&= \prod_{ij} \frac{\exp [\theta' g_{ij} \Delta \mathbf{t} (g_{-ij}, X)]}{1 + \exp (\theta' \Delta \mathbf{t} (g_{-ij}, X))}
\end{aligned} \tag{38}$$

Taking logs we get

$$\begin{aligned}
\log PL(g; \theta, X) &= \sum_i \sum_j g_{ij} \left(u_{ij} + g_{ji} m_{ij} + \sum_{k \neq i, j} g_{jk} v_{ik} + \sum_{k \neq i, j} g_{ki} v_{kj} \right) \\
&\quad - \sum_i \sum_j \log \left[1 + \exp \left(u_{ij} + g_{ji} m_{ij} + \sum_{k \neq i, j} g_{jk} v_{ik} + \sum_{k \neq i, j} g_{ki} v_{kj} \right) \right] \\
&= \theta' \sum_i \sum_j g_{ij} \Delta \mathbf{t} (g_{-ij}, X) - \sum_i \sum_j \log [1 + \exp (\theta' \Delta \mathbf{t} (g_{-ij}, X))]
\end{aligned}$$

Let's rewrite the equation as

$$\begin{aligned}
\ell(\theta) &= \log PL(g; \theta, X) \\
&= \sum_i \sum_j g_{ij} \sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X) - \sum_i \sum_j \log \left[1 + \exp \left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X) \right) \right]
\end{aligned}$$

Therefore the First order conditions are

$$\begin{aligned}
\frac{\partial \ell(\theta)}{\partial \theta_k} &= \sum_i \sum_j g_{ij} \Delta t_k (g_{-ij}, X) - \sum_i \sum_j \Delta t_k (g_{-ij}, X) \frac{\exp \left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X) \right)}{1 + \exp \left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X) \right)} \\
&= \sum_i \sum_j \Delta t_k (g_{-ij}, X) [g_{ij} - p_{ij}]
\end{aligned}$$

³⁹In a similar setting, Nakajima (2007) uses the same idea.

where I defined

$$p_{ij} \equiv \frac{\exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)}{1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)}$$

The Hessian is therefore

$$\begin{aligned} \frac{\partial^2 \ell(\theta)}{\partial \theta_k \partial \theta_s} &= -\left\{ \sum_i \sum_j \Delta t_k (g_{-ij}, X) [\Delta t_s (g_{-ij}, X) \frac{\exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)}{\left[1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)\right]^2} \left[1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)\right] \right. \\ &\quad \left. - \frac{\exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)}{\left[1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right)\right]^2} \Delta t_s (g_{-ij}, X) \left[\exp\left(\sum_{k=1}^K \theta_k \Delta t_k (g_{-ij}, X)\right) \right] \right\} \\ &= -\sum_i \sum_j \Delta t_k \Delta t_s \frac{\exp\left(\sum_{k=1}^K \theta_k \Delta t_k\right)}{\left[1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k\right)\right]^2} \left[1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k\right)\right] \\ &\quad - \sum_i \sum_j \Delta t_k \frac{\exp\left(\sum_{k=1}^K \theta_k \Delta t_k\right)}{\left[1 + \exp\left(\sum_{k=1}^K \theta_k \Delta t_k\right)\right]^2} \Delta t_s \left[\exp\left(\sum_{k=1}^K \theta_k \Delta t_k\right) \right] \\ &= -\sum_i \sum_j \Delta t_k \Delta t_s p_{ij} [1 - p_{ij}] \end{aligned}$$

The maximizer can be found with a Newton-Raphson algorithm

$$\theta^{(n+1)} = \theta^{(n)} - \left[\frac{\partial^2}{\partial \theta \partial \theta'} \ell(\theta^{(n)}) \right]^{-1} \frac{\partial}{\partial \theta} \ell(\theta^{(n)}) \quad (39)$$

C Proofs

Proof of Proposition 1

The potential is a function Q from the space of actions to the real line such that $Q(g_{ij}, g_{-ij}, X) - Q(g'_{ij}, g_{-ij}, X) = U_i(g_{ij}, g_{-ij}, X) - U_i(g'_{ij}, g_{-ij}, X)$, for any ij .⁴⁰ A back-of-the-envelope computation shows that, for any ij

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

⁴⁰ For more details and definitions see Monderer and Shapley (1996).

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

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

Proof of Proposition 2

- 1) This 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 $g_{ij} \neq g_{ij}^*$

$$U_i(g_{ij}^*, g_{-ij}^*, X) \geq U_i(g_{ij}, g_{-ij}^*, X)$$

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

$$Q(g_{ij}^*, g_{-ij}^*, X) \geq Q(g_{ij}, g_{-ij}^*, X)$$

Therefore g^* is a maximizer of Q . The converse is easily checked using the opposite 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 will not escape from that state.

- 4) The probability that the potential will increase from t to $t + 1$ is

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

By part 3) of the proposition, a Nash network is an absorbing state of the chain. Therefore any probability distribution that puts probability 1 on a Nash network is a stationary distribution. For

any initial network, the chain will converge to one of the stationary distributions. It follows that in the long run the model will be in a Nash network, i.e. for any $g^0 \in \mathcal{G}$

$$\lim_{t \rightarrow \infty} \Pr [g^t \in NE | g^0] = 1$$

Proof of Proposition 3

1. The sequence of networks $[g^0, g^1, \dots]$ 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 result then follows from the ergodic theorem.

2. I show the proof for the case of a *symmetric* meeting process ρ_{ij} . The case for a generic ρ_{ij} is slightly more complicated, but it follows closely Besag (1974).⁴¹ A sufficient condition for stationarity is the *detailed balance*. In our case this means to prove

$$P_{gg'}\pi_g = P_{g'g}\pi_{g'} \quad (40)$$

where

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

Notice that the transition from g to g' is possible if they 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 $\Delta Q = Q(1, g_{-ij}, X) - Q(0, g_{-ij}, X)$

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

⁴¹The realization of the network is an example of Markov Random Field (MRF). In the network there is spatial dependence among the different links, since the probability of a link depends on the state of the rest of the network. Besag (1974) provides the characterization of the probability distribution of a MRF for any spatial dependence.

From part 1) of the proposition, we know that the process is ergodic. Therefore $\pi(g, X)$ is also the unique stationary distribution.

Proof of Proposition 4

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

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

where $\theta_u = (\theta_{u1}, \theta_{u2}, \dots, \theta_{uP})'$ and $\mathbf{t}_u(g, X) = (t_{u1}(g, X), t_{u2}(g, X), \dots, t_{uP}(g, X))'$. Analogously define $\theta_m = (\theta_{m1}, \theta_{m2}, \dots, \theta_{mL})'$ and $\mathbf{t}_m(g, X) = (t_{m1}(g, X), t_{m2}(g, X), \dots, t_{mL}(g, X))'$ and $\theta_v = (\theta_{v1}, \theta_{v2}, \dots, \theta_{vS})'$ and $\mathbf{t}_v(g, X) = (t_{v1}(g, X), t_{v2}(g, X), \dots, t_{vS}(g, X))'$. It follows that

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

and

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

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

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

and the stationary distribution is then

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

D Variables Definition and Sample Selection

I consider only schools with at least 30 students and with reliable network data, i.e. there is more than 1 link. These criteria imply the deletion of 16 schools, with a resulting sample of 130 schools.⁴²

The network population is defined as all the students in the school. The friendship nominations are obtained from the *in-school* survey. I consider only the links between pairs enrolled in the same school.⁴³ Nominations of students that are also romantic partners are deleted from the list of friends, since it is not clear how a romantic involvement affects friendship.⁴⁴

The control for racial groups is constructed from the *in-home* questionnaire, as in Moody (2002). An individual is considered *hispanic* if he/she self identified himself as such. All non-hispanic students are assigned to four racial groups (white, black, asian and other race) as identified by the interviewer.⁴⁵

⁴²This is the same sample considered in Moody (2002).

⁴³This is standard in the literature using the AddHealth data.

⁴⁴This is standard when using these data. See Moody (2002) and Mihaly (2007).

⁴⁵There are several ways to assign individuals to racial groups. A question asks the student to identify which racial group(s) he belongs to. If a survey participant identifies in more than one racial group, he/she is asked to determine which one is the most important. Also the interviewer is asked to report the racial group of the student. I use the latter variable since it seems more reliable.

Since I focus my analysis on segregation by race and gender, and since most individuals seem to nominate people in the same grade, any student with a missing value for gender, racial group or grade is deleted from the school network.

E Estimation and MCMC details

This appendix reports in more detail the techniques used in estimation and simulation. In particular I focus on the properties of the likelihood, the stochastic approximation algorithm and the details of the Markov Chain Monte Carlo method for simulations.

E.1 Details for Estimation

The parameters are estimated using Maximum Likelihood methods. The log-likelihood is denoted by $\ell(g_{obs}; \theta)$ and the parameters are solution of the maximization problem

$$\hat{\theta}_{mle} = \arg \max_{\theta \in \Theta} \ell(g_{obs}; \theta) \quad (41)$$

This maximization problem can be solved using MCMC-MLE techniques as proposed by Geyer and Thompson (1992) and Geyer (1994). Unfortunately in this particular model, the MCMC-MLE methodology is numerically problematic. The problem is the computation of exponential functions for very big numbers which may occasionally cause overflow.

I therefore exploit a property of the exponential family to obtain an estimation method that avoids the use of exponentials. In Appendix B, I show that solving the maximization problem (41) is equivalent to solving the system of nonlinear equations

$$t(g_{obs}, X, \theta) = \mathbb{E}[t(\omega, X, \theta)] \quad (42)$$

The system corresponds to the first order conditions of (41). The vector $t(g_{obs}, X, \theta)$ is the vector of canonical sufficient statistics observed in the data.

The formulation of the problem as in (42) suggests a simple Monte Carlo technique for estimation. Given a specific value of θ we can draw sample networks from the stationary distribution and compute the sample mean of the sufficient statistics. The problem is to find the $\hat{\theta}$ that equates the sample mean to the observed sufficient statistics.

This simple strategy is computationally inefficient since for each iteration at a given parameter value one has to draw a new sample, simulating from the stationary distribution. Even assuming that the simulation is very fast (which is not the case here) one ends up discarding a lot of generated samples.⁴⁶

I solve the system of equations using the Robins-Monro *stochastic approximation* algorithm (Robins and Monro, 1951).⁴⁷ The idea behind this method is to generate a Markovian sequence of

⁴⁶In this sense the MCMC-MLE method is the most computationally efficient, since it requires to draw a single sample to perform the maximization.

⁴⁷Kushner and Yin (2003) provide an extensive review of the literature on stochastic approximation and applications of the algorithm. Snijders (2002) proposes to use Robins-Monro for the estimation of Exponential Random Graph Models (ERGM).

parameters whose mean converges to the solution of the equation. In this sense it is a sequential estimation method and it can be thought of as a Monte Carlo version of the Newton-Raphson algorithm.

I use several useful properties of the exponential family of distributions. The Hessian of the maximum likelihood problem is equivalent to the variance-covariance matrix of the sufficient statistics, i.e.

$$COV(t_k(g, X, \theta), t_l(g, X, \theta)) = \frac{\partial^2 \pi}{\partial \theta_k \partial \theta_l} \quad (43)$$

for all k, l . The asymptotic variance-covariance matrix of the maximum likelihood estimates $AsyV(\theta_{mle})$ is the inverse of (43).

$$AsyV(\theta_{mle}) = [COV(\mathbf{t}(g, X, \theta))]^{-1} \quad (44)$$

Since none of these quantities can be computed analytically, I estimate them using Markov Chain Monte Carlo methods.

The Robins-Monro algorithm sequentially updates the parameters using the recursion

$$\hat{\theta}^{(r+1)} = \hat{\theta}^{(r)} - a_r D_r^{-1} [\mathbf{t}(g^{(r)}, X, \theta) - \mathbf{t}(g_{obs}, X)] \quad (45)$$

where the *gain* a_r is a sequence of positive numbers that converges to zero as r goes to infinity (usually $1/r$) and

$$D_r = \frac{\partial \mathbb{E} [\mathbf{t}(\omega, X, \theta^{(r)}) - \mathbf{t}(g_{obs}, X)]}{\partial \theta} = COV(\mathbf{t}(g, X, \theta)) \quad (46)$$

The second equality follows from the properties of the exponential family. It can be shown that the sequence $\hat{\theta}^{(r)}$ is Markov and it is asymptotically normal (see Snijders (2002) and Kushner and Yin (2003)).

$$\hat{\theta}^{(r)} \rightarrow N \left(\theta, \frac{1}{n} COV(\mathbf{t}(g, X, \theta))^{-1} \right) \quad (47)$$

Since for each iteration the matrix D_r must be re-estimated through simulation from the current parameter $\hat{\theta}^{(r)}$, this procedure is also inefficient. However it can be shown that if D_r is *positive definite* then a constant matrix D_0 can be used as long as the average of the sequence $\hat{\theta}^{(r)}$ is used as estimate for θ and the gain sequence is slightly modified. This is the case here since D_r is the variance-covariance matrix of the sufficient statistics.

Therefore the algorithm follows the 3 steps suggested in Snijders (2002)

1. Guess a θ^0 and estimate D_0 through MCMC methods.
2. Use the estimated \hat{D}_0 in (45). For a given $\hat{\theta}^{(r)}$ simulate a network and iterate one step ahead.
3. Once convergence is achieved, use the estimated $\hat{\theta} = (1/n) \sum_{r=1}^R \theta^{(r)}$ to run a long MCMC and estimate the standard errors of the parameters as \hat{D}^{-1} .

This algorithm works well as long as the initial θ^0 is not too far from the solution. In some cases this procedure needs to be iterated several times to reach reasonable convergence.

In order to initialize the parameters for the simulations, I follow the suggestions in Besag

(1974), Snijders (2002) and Nakajima (2007). I use the Maximum Pseudo-Likelihood Estimator, which maximizes the product of the conditional probabilities (details of the derivation can be found in Appendix B)

$$\log PL(g; \theta, X) = \theta' \sum_i \sum_j g_{ij} \Delta \mathbf{t}(g_{-ij}, X) - \sum_i \sum_j \log [1 + \exp(\theta' \Delta \mathbf{t}(g_{-ij}, X))]$$

E.2 Simulation Methods

E.2.1 The Main Building Blocks

The dynamic of model replicates the *Gibbs Sampler* dynamics, as noted above. Therefore the simulation through Gibbs Sampling appears the natural candidate in this setting. For each step of the sampler one randomly chooses an element of the matrix G to update, and updates with probability given by (13).

However, *any* MCMC algorithm that converges to the stationary distribution of interest is of practical interest. If a sampler satisfies the *detailed balance* condition is suitable for the simulation of the network model⁴⁸

$$\Pr [g'|g] \pi(g) = \Pr [g|g'] \pi(g') \quad (48)$$

In particular I use the *Metropolis-Hastings* algorithm in order to sample networks from the stationary distribution for estimation. For each iteration one element g_{ij} of the matrix is randomly selected and it is proposed to swap its value to $1 - g_{ij}$.

ALGORITHM 1 *Metropolis-Hastings Algorithm* Start at some g^0 . One simple choice is the observed network g_{obs} .

1. At the current state g select an element g_{ij} of the matrix with probability $\rho_{ij} = 1/(n(n-1))$
2. Update the network from g to $g' = (1 - g_{ij}, g_{-ij})$ with probability

$$r(g, g') = \min \{1, \exp [Q(g', X, \theta) - Q(g, X, \theta)]\} \quad (49)$$

There are both a theoretical and a practical reason for using the Metropolis-Hastings algorithm instead of the Gibbs sampler. First, it is known that in the context of exponential family the Gibbs sampler produces a lot more rejections of the proposed states than the Metropolis-Hastings algorithm.⁴⁹ Therefore the Gibbs sampler remains in the same state for long times and the mixing is slow (see more detail below). Second, most of the techniques I use to accelerate the mixing and the speed of computations are easier to apply in the Metropolis-Hasting context.

⁴⁸ I have proven that the Gibbs sampler satisfies this condition in the proof of Proposition 4, when the meeting process is symmetric. The proof for the Metropolis-Hastings algorithm is trivial and it is not given here.

⁴⁹See Snijders (2002) for a discussion

E.2.2 Slow Mixing

In theory *any* MCMC scheme satisfying the detailed balance is successful in producing samples from the stationary distribution of the model, provided that the researcher is willing to run the chain for a *long enough* amount of time (Geyer (1991)). However in models with large state spaces or when the distribution is multimodal, the MCMC may require unfeasible long runs to reach a reasonable mixing rate. In this context the problem is serious enough to require drastic changes in the MCMC scheme to speed up convergence.

I tackle this problem using several improvements of the basic Metropolis-Hasting scheme, with three main goals: 1) increase mixing; 2) decrease the sample size needed for simulations; 3) decrease the variance of the estimates.

E.2.3 Large Steps

In many application it is convenient to propose larger steps.⁵⁰ Geyer (1991) suggests a swap step, where it is proposed to update the current state of the chain with his complement. Snijders (2001 and 2002) suggests a similar step in the context of the exponential random graphs models: the proposal is to swap the current network g to the complement $\mathbf{1} - g$, i.e. the network matrix is inverted so that all the entries with 1 become 0, and viceversa.

Snijders (2002) argues that this step is extremely useful when the stationary distribution is bimodal, since it allows the chain to jump from one mode to the other more easily.⁵¹

I use the swap step in my simulation. In each iteration of the sampler with probability p_{inv} a swap is proposed and it is accepted with probability

$$p_{swap}(\mathbf{1} - g, g) = \min \{1, \exp [Q(\mathbf{1} - g, X, \theta) - Q(g, X, \theta)]\} \quad (50)$$

which satisfies the detailed balance condition.

However, ideally one is looking for a scheme that allows to explore the entire state space without knowing too much about the shape of the distribution. Therefore I combine this technique with other MCMC schemes.

E.2.4 Tempering

Simulated Tempering (Marinari and Parisi(1992), Geyer and Thompson (1995)) is an MCMC scheme that involves simulation from different MCMC samplers. This produces a long chain that contains variable length runs from each sampler and switches among samplers.

The intuition of the procedure is simple and resembles closely the *simulated annealing* algorithm used in optimization. Assume that the stationary distribution of the network model is the black distribution of Figure 2(a). The distribution is bimodal and the modes are separated by a very

⁵⁰Most MCMC algorithms suggest *small steps*: a typical Gibbs sampler cycles through each element/variable of the state in a systematic order, updating only a single element per iteration.

As noted in Geyer (1991) there is no reason to restrict the practice of MCMC to only Gibbs or Metropolis-Hasting alone algorithms. *As long as the chain generated by the algorithm has the right stationary distribution, the ergodic theorem insures convergence of the estimated averages.*

⁵¹ The example provided by Geyer (1991) is the Ising model, which has a bimodal stationary distribution for some parameter values. The stationary distribution of the Ising Model is a two-parameter exponential family with normalizing constant, similar to the model used here

steep valley. Starting from a network g^0 a proposed network g^1 will be accepted by the Metropolis-Hastings algorithm. But in order to reach the alternative mode, the algorithm should accept a sequence of downhill moves until it reaches g^2 . This sequence has very low probability to occur. Therefore the sampler remains trapped in one mode for a very long period: it eventually reaches the alternative mode, but it may take unfeasible long runs of the chain to do so.

Consider instead the blue and red distributions in Figure 2(b). These are obtained from the black distribution $\pi(g)$ by tempering, i.e. the blue has stationary distribution $\pi(g)^{1/\tau_1}$ and the red $\pi(g)^{1/\tau_2}$, with $1 < \tau_1 < \tau_2$. A Metropolis-Hastings algorithm that has the blue or red distribution as stationary distribution is much more likely to switch among modes, since the hills are much flatter than the original distribution.

The simulated tempering algorithm samples from the different chains in sequence, switching from the black distribution to the blue, from the blue to the red and viceversa. Only the samples obtained from the black distribution are used in estimation of the average sufficient statistics: the *tempered* distributions are used only for being able to explore the state space faster and to avoid that the sampler get trapped in one of the modes.

More formally, the state of the chain is described by the pair $(\tau^{(r)}, g^{(r)})$, where $\tau^{(r)}$ is the indicator of the current sampler (the temperature) and $g^{(r)}$ is the state of the chain at iteration r . Let $h(g) \equiv \exp[Q(g, X, \theta)]$ be the unnormalized density associated with the stationary distribution $\pi(g, X, \theta)$.

There are m unnormalized densities $h_\tau(g)$, where $\tau = \tau_1, \dots, \tau_m$, defined over the same state space \mathcal{G} . The stationary distribution of the chain obtained from simulated tempering is

$$h_{st}(g, \tau) \propto h_\tau(g) \mathcal{K}_\tau \quad (51)$$

where the constants \mathcal{K}_τ , called *pseudopriors*, are chosen by the researcher so that the chain spends roughly the same time in each temperature. Geyer and Thompson (1995) show that the optimal value for the constants is $\mathcal{K}_\tau = 1 / \sum_{\omega \in \mathcal{G}} h_\tau(\omega)$. The probabilities for switching among temperatures are governed by a random walk with transition probabilities $q_{\tau, \tau'}$, such that $q_{1,2} = q_{m,m-1} = 1$ and $q_{\tau, \tau+1} = q_{\tau, \tau-1} = 1/2$.

ALGORITHM 2 (*Simulated Tempering*) Start the iteration at current state $(\tau^{(r)}, g^{(r)})$

1. Update $g^{(r+1)}$ with a Metropolis-Hastings step using $h_{\tau^{(r)}}(g)$ (propose the swap step with probability p_{inv} and the one-link update with probability $1 - p_{inv}$)
2. Pick $\tau' = \tau \pm 1$ according to probabilities $q_{\tau, \tau'}$
3. Accept the transition from temperature $\tau^{(r)} = \tau$ to $\tau^{(r+1)} = \tau'$ with probability

$$\alpha_{ST} = \min \left\{ 1, \frac{h_{\tau'}(g^{(r)}) \mathcal{K}_{\tau'} q_{\tau', \tau}}{h_\tau(g^{(r)}) \mathcal{K}_\tau q_{\tau, \tau'}} \right\} \quad (52)$$

Only the samples from the cold distribution $h_{\tau_1}(g)$ are used to compute the average sufficient statistics. The Simulated Tempering sampler mixes better than the original Metropolis-Hastings algorithm, since the heated temperatures allow the chain to visit all states.

In order to compute and update the pseudopriors, I use the stochastic approximation algorithm proposed in Geyer and Thompson (1995).

E.2.5 Improving the Simulated Tempering Algorithm

The most unappealing feature of the simulated tempering is that it is computationally inefficient, since it discards all the samples generated by the hot temperatures. I borrow some suggestions from Li, Protopopescu and Gorin (2004), to speed up the simulated tempering algorithm.

The idea is to switch among temperatures using the probabilities $q_{i,i+1} < q_{i,i-1}$ so that lower temperatures are visited more often.⁵²

The second idea is to use larger steps for higher temperatures than for the lower ones. Since the only relevant samples are the ones from the cold distribution, there is no need to get too much detail at higher temperatures. Therefore, I use larger steps for higher temperatures, by modifying the probability of a swap update of the network. The probability is now $p_{inv}(\tau)$ and I set $p_{inv}(1) = .1$ and $p_{inv}(\tau + 1) = p_{inv}(\tau) + .02$.

These improvements modify the algorithm directly, but we can increase the speed of simulation by modifying the storage of the data. The idea is to use the sparsity of the network to speed up the computations of the sufficient statistics and the Metropolis-Hastings updates. I store the network matrix in Compressed Sparse Row (CSR) form and use sparse matrix algebra routines to compute the network statistics.⁵³ I adapt various codes from the package SPARSEKIT, available online at <http://www-users.cs.umn.edu/saad/software/SPARSKIT/sparskit.html>.

I also employ variance reduction techniques. In the context of this model one can use the swap steps to reduce the variance of estimated averages. This is used when estimating the average and the variance-covariance matrix of the sufficient statistics in the first and third step of the Robins-Monro algorithm. As suggested in Snijders (2002), it can be proven that

$$\mathbb{E}[t(g)] = \mathbb{E}[pr(g)t(\mathbf{1} - g) + (1 - pr(g))t(g)]$$

where

$$pr(g) = \frac{\exp[\theta t(\mathbf{1} - g)]}{\exp[\theta t(\mathbf{1} - g)] + \exp[\theta t(g)]}$$

It can be shown that the right hand side has lower variance.

TO BE COMPLETED: Insert proof here as from notes

F A Note on the log sum of exponentials

Consider the log sum of exponentials, which is needed to compute the approximated log likelihood in MCMC-MLE and in the approximated Metropolis-Hastings algorithm used in the Bayesian estimation

$$\log \left\{ \frac{1}{R} \sum_{r=1}^R \exp \left[Q \left(g^{(r)}, X, \theta^{(it)} \right) - Q \left(g^{(r)}, X, \theta^* \right) \right] \right\} \quad (53)$$

This is problematic because of the floating point arithmetic of the computer. Sometimes the differences $Q(g^{(r)}, X, \theta^{(it)}) - Q(g^{(r)}, X, \theta^*) \equiv \Delta Q(g^{(r)})$ inside the exponential are too big and this

⁵²This is the same random walk across temperatures, but it is asymmetric.

⁵³A description of the different storage modes is contained in the vignette pdf for the package SparseM, an R package designed to implement sparse matrix algebra in the R software environment. The vignette is available at <http://www.econ.uiuc.edu/roger/research/sparse/SparseM.pdf>.

causes overflow, since the exponential is evaluated as infinity. The trick is to make the overflow less likely. Define

$$a = \max_i \sum_{r=1}^R \exp \left[\Delta Q \left(g^{(r)} \right) \right] \quad (54)$$

$$\sum_{r=1}^R \exp \left[\Delta Q \left(g^{(r)} \right) \right] = \sum_{r=1}^R \exp \left[\Delta Q \left(g^{(r)} \right) - a \right] \exp(a) \quad (55)$$

and therefore

$$\log \left\{ \frac{1}{R} \sum_{r=1}^R \exp \left[\Delta Q \left(g^{(r)} \right) \right] \right\} = \log \left\{ \frac{1}{R} \sum_{r=1}^R \exp \left[\Delta Q \left(g^{(r)} \right) - a \right] \right\} + a \quad (56)$$