Segregation in Social Networks: Theory, Estimation and Policy∗†

Angelo Mele‡

VERY PRELIMINARY - PLEASE DO NOT QUOTE

July 16, 2010

Abstract

How do desegregation policies affect the level of interracial interactions within schools? This paper develops and estimates a model of strategic network formation with heterogeneous agents. Individuals randomly meet over time and form links according to their preferences for socio-economic characteristics. The model has several desirable features. First, there is a unique stationary equilibrium. As a result, the estimation can be performed using only one observation of the network at a single point in time. Second, in equilibrium, exogenously identical individuals may have different linking strategies, due to their endogenous friends’ composition. The unique stationary equilibrium of the model characterizes the likelihood of observing a specific network in the data. However, the exact evaluation of the likelihood is computationally unfeasible. Therefore, to estimate the structural preference parameters, I propose a Bayesian Markov Chain Monte Carlo method that avoids direct evaluation of the likelihood. I use data from the restricted version of Add Health, a survey containing detailed information on the actual social network of each student in a representative sample of US high schools.

I find that students prefer to form friendship links to individuals of the same racial group, ceteris paribus. Furthermore, students prefer to link individuals of the same racial group that have a high fraction of friends of the same racial group, ceteris paribus. I simulate several desegregation programs in which minority students are taken from one school and reassigned to another school in the Add Health data. These simulations show that such policies may actually decrease racial integration in schools.

∗I am grateful to Roger Koenker for continuous encouragement and advice, generous financial support and for allowing me to use his computer cluster. I thank Ron Laschever for long and fruitful discussions about this research project. Dan Bernhardt and George Deltas provided useful inputs at crucial stages of this work. I thank Alberto Bisin, Ethan Cole, Aureo de Paula, Shweta Gaonkar, Darren Lubotsky, Antonio Mele, Luca Merlino, Dennis O’Dea, Micah Pollak, Sergey Popov, Sudipta Sarangi, Giorgio Topa, Antonella Tutino and participants to the UIUC Research Seminar, SED Meetings 2010, for helpful comments and suggestions. Financial support from the Robert Ferber Award, the Robert Willis Harbeson Memorial Dissertation Fellowship, and the Net Institute Summer Research Grant is gratefully acknowledged. 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 Dr., Urbana, IL 61801. Email: amele2@illinois.edu
1 Introduction

The debate on desegregation programs has been recently revived by a 2007 decision of the Supreme Court that prohibited the use of race in the school assignments. Since the Brown vs Board of Education decision in 1954, school districts were pushed towards desegregation and developed many busing programs, transporting minority students to homogeneously white schools. Some evidence suggests that these programs were successful in decreasing dropouts (Guryan, 2004) and improving academic achievement (Cooley, 2010; Angrist and Lang, 2004) of minorities.

This paper adds to the literature by studying how these programs may affect racial integration within schools. Busing programs are designed to increase diversity at the school level. However, social networks often display homophily: people are more likely to interact with similar individuals (Currarini et al., 2009). If students have strong preferences for interaction with friends of the same racial group, an increase in within-school diversity would not necessarily increase interracial friendships and interactions.

I develop a model of noncooperative social network formation with heterogeneous agents, that characterizes the level of interracial interaction as an equilibrium outcome. The model’s unique stationary equilibrium provides the likelihood of observing a network configuration in the data. I estimate the posterior distribution of the individual preference parameters driving the aggregate levels of interracial friendships, using Bayesian Markov Chain Monte Carlo methods. The estimation is based on data from the restricted version of the National Longitudinal Study of Adolescent Health (Add Health), containing detailed information on the actual friendship network of each student in a representative sample of US high schools. The simulation of several counterfactual policies allows me to predict their impact on the levels of racial segregation in the friendship network.

I find that racial homophily is pervasive: students prefer to form friendship links to individuals of the same racial group. Furthermore, I find strong homophily effects for friends of friends: students prefer to link individuals of the same racial group, that have a high fraction of friends of the same racial group, ceteris paribus. Both these effects translate in high levels of racial segregation in the social network.

The estimated model can be used to simulate alternative policies that promote integration. As an example, I simulate a re-assignment of minority students from one school of Add Health to another school. In the receiving school, Whites and African Americans have similar population shares (42% and 45.3%), while Hispanics are only 10.6% of the student body. When I re-assign 8 African Americans to this school, the model predicts that the level of segregation in the stationary equilibrium decreases for all racial groups. However, when I re-assign 16 Hispanic students to this school, segregation decreases for Whites and African Americans, but it increases for Hispanics. The intuition is that the increase in the Hispanic population is substantial and homophily pushes the hispanic students to form more links with hispanics, increasing aggregate segregation levels for this group. This does not happen for African Americans in the first policy, since the 8 additional students do not modify the population shares substantially. More generally, these effects may be highly nonlinear.

---

1 For example, consider a school with 100 kids, 98 whites and 2 blacks. Even if racial homophily is strong, the two black students could become friends with some of the white kids, unless they do not want to have more than one friend. Let’s consider a busing program that transports 30 more black students to this school. Now if the black kids want to have only black friends, they can, since the population is larger. In general these effects may be highly nonlinear and have nontrivial dynamics (Currarini et al., 2009).
The model incorporates ingredients from both strategic and random network formation literature (Jackson, 2008). The link formation is sequential: in each period only one agent is active and he only updates one link. At the beginning of the period, a random agent \(i\) is drawn from the population, he meets another agent \(j\) according to a random matching technology and decides whether to link \(j\). When updating the link, individual \(i\) receives a random shock to his preferences, which is unobserved by the econometrician. Agent \(i\) forms the link to \(j\) if his utility when the link exists is greater than the utility when the link does not exist. A crucial assumption is that individuals do not take into account how their current linking strategy affects the shape of future networks: they follow a stochastic best-response dynamics à la Blume (1993).²

The model has several desirable features. First, there are three levels of heterogeneity. Each individual is endowed with a set of \textit{exogenous} attributes. Furthermore, the dynamics of network formation generates \textit{endogenous} heterogeneity: each individual has a different set of friends and a different composition of friends’ attributes. Finally, there is behavioral heterogeneity: in equilibrium, two agents with the same exogenous attributes may have different linking strategies, because of their different endogenous positions in the network and the socioeconomic composition of their friends. Most models of strategic network formation incorporate the first two levels of heterogeneity but they are unable to generate different equilibrium behavior.³

Second, I can characterize the network formation game as a \textit{potential game}.⁴ All the players’ incentives in any state of the network are completely summarized by an aggregate function, \textit{the potential}, mapping networks and socioeconomic characteristics into potential levels. When an agent updates a link, the change in his utility is equal to the change in the potential. This simple characterization is very useful when studying a network with many agents, since the potential summarizes the incentives of all players with one real number: there is no need to keep track of the choices and utility levels of all \(n\) players. The existence of a potential simplifies 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 Nash networks of a model without any preference shock. Assuming a matching technology in which any pair of players has positive probability of meeting, I prove that in the long run, the network converges to one of the Nash equilibria with probability one. Assuming logistic preference shocks, I prove that the model is ergodic and it converges to a unique stationary distribution, which is a function of the potential.

Third, the model requires a minimal amount of data for estimation: Maximum Likelihood or Bayesian inference can be performed using only one observation of the network at a single point in time. The model’s unique stationary equilibrium is the likelihood of observing a specific network shape in the data. When data from multiple independent networks are available, the estimation strategy can be easily adapted.

In this paper I use Bayesian methods to estimate the posterior distribution of the parameters

---

²It is possible to relax the assumption of myopic agents, but the computational burden becomes challenging. The simple characterization of equilibrium behavior, long run dynamics and the estimation strategy depend on the best-response dynamics and may not extend to an economy with forward-looking agents.

³An exception is the model of De Marti and Zenou (2009).

⁴See Monderer and Shapley (1996) for a description of games with a potential. ⁷ investigates a model of network formation with a potential function. Their model only considers direct links utility, while mine includes indirect links, mutual links and popularity.
Bayesian inference is particularly useful in this context, since Maximum likelihood estimation is extremely hard: the likelihood evaluation is computationally unfeasible (see (Mele, 2010)). At the same time, Bayesian estimation of this model poses two challenges. First, the posterior distribution is proportional to a normalizing integral, that cannot be computed exactly. This is common to any Bayesian computation and it is solved using the Metropolis-Hastings algorithm to generate samples from the posterior distribution. However, there is a second challenge: the likelihood of my model is also proportional to a normalizing constant that cannot be computed exactly. As a consequence the Metropolis-Hastings ratio cannot be computed exactly. In similar contexts, some researchers have proposed approximations of the Metropolis-Hastings ratio, using approximated likelihoods. However, it is not clear that the chains generated using an approximated Metropolis-Hastings are samples from the right posterior distribution.

To solve this problem, I propose a variation of the exchange algorithm, first developed by Murray et al. (2006) for a similar family of distributions. My method is a Markov Chain Monte Carlo algorithm that removes the need to compute the normalizing constant both in the likelihood and the posterior. It therefore allows exact computation of the Metropolis-Hastings ratio. I prove that the proposed algorithm generates samples from the right posterior distribution. This comes with a cost: the chains of parameters are highly autocorrelated, therefore one has to run the simulations for many more iterations than a naive version of the Metropolis-Hastings algorithm, to estimate the posterior mean with precision.

The remainder of the paper is organized as follows. Section 2 provides a description of Add Health and summary statistics of the data. Section 3 illustrates the model setup, the stationary equilibrium and the dynamics. Section 4 presents the estimation strategy and the algorithms used for simulation. Section 5 discusses the estimation results and Section 6 concludes. Appendix A collects all the proofs for the theoretical model, while Appendix B provides all the details about the MCMC algorithm and convergence.

2 The Add Health Data

The National Longitudinal Study of Adolescent Health (AddHealth) collects data for a nationally representative sample of adolescent in United States. The participants were enrolled in grades 7-12 during the 1994-1995 school year and they have been followed into adulthood with four successive...
in-home interviews. In this paper I use only data from Wave I, containing information on 132 nationally representative US high schools. The survey collected information about each student through in-school and in-home interviews. School level information and parents background were collected in separate interview of school administrators and parents.

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. I use the friendship nominations as proxy for the social network in a school. The resulting network is directed: Paul may nominate Jim but this does not necessarily implies that Jim nominates Paul. The model developed in this paper takes this feature of the data into account.

A sub-sample of 20745 students was also given an in-home questionnaire, that collected most of the sensible data. I use data on racial group, grade and gender of individuals. A student with a missing value in any of these variables is dropped from the sample. Each students that declares to be of Hispanic origin is considered Hispanic. The remaining non-Hispanic 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 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. 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 African Americans. Others show high levels of diversity in their populations. The gender composition is more or less homogeneous across schools.

Table 2 contains 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 are reliable. If a student nominated a friend who is not enrolled in the same school, I drop that friendship from the dataset.

---

8 More 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

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

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

11 Some authors do not take into account this feature of the data and they recode the friendships as mutual: if a student nominates another one, the opposite nomination is also assumed. Echenique and Fryer (2007)

12 The 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

\[ FRAG = 1 - \sum_{k=1}^{K} (s_k)^2 \] (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.
Table 1: Summary Statistics of Racial Composition

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Mean</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>TotStudents</td>
<td>43</td>
<td>87</td>
<td>127</td>
<td>151.8</td>
<td>173</td>
<td>1664</td>
</tr>
<tr>
<td>Whites</td>
<td>0</td>
<td>0.3318</td>
<td>0.6909</td>
<td>0.6013</td>
<td>0.9061</td>
<td>1</td>
</tr>
<tr>
<td>Blacks</td>
<td>0</td>
<td>0.01114</td>
<td>0.0818</td>
<td>0.2147</td>
<td>0.3545</td>
<td>0.9866</td>
</tr>
<tr>
<td>Asians</td>
<td>0</td>
<td>0.01345</td>
<td>0.04652</td>
<td>0.1304</td>
<td>0.1704</td>
<td>0.901</td>
</tr>
<tr>
<td>Hispanics</td>
<td>0</td>
<td>0.01345</td>
<td>0.04652</td>
<td>0.1304</td>
<td>0.1704</td>
<td>0.901</td>
</tr>
<tr>
<td>Others</td>
<td>0</td>
<td>0</td>
<td>0.00678</td>
<td>0.01422</td>
<td>0.0191</td>
<td>0.3165</td>
</tr>
<tr>
<td>Females</td>
<td>0.01923</td>
<td>0.4896</td>
<td>0.5084</td>
<td>0.5085</td>
<td>0.5375</td>
<td>0.6538</td>
</tr>
<tr>
<td>FragnRace</td>
<td>0</td>
<td>0.122</td>
<td>0.3725</td>
<td>0.3471</td>
<td>0.5181</td>
<td>0.7641</td>
</tr>
</tbody>
</table>

Table 2: Summary Statistics for Grade Composition

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

3 A Model of Network Formation

3.1 Setup

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

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

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

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. $g_{ij} \neq g_{ji}$. This modeling choice reflects the structure of the Add Health data, where friendship nominations are not necessarily mutual. Some authors refer to this data as perceived networks.\(^1\)

The network represented by $G$ is the realization of the network at time $t$ be denoted as $g^t$ and the realization of the link between $i$ and $j$ at time $t$ be $g^t_{ij}$. The network including all the current links but $g^t_{ij}$, i.e. $g^t \setminus g^t_{ij}$, is denoted as $g^t_{-ij}$.

\(^1\)See Wasserman and Faust (1994) for some references.
The stochastic process described above generates a sequence of all agents. Since they can observe the entire network shape and the individual attributes complete information, each agent maximizes his current utility, taking the previous period network only one element of the random matrix $G$ that could influence the utility of a link, e.g. mood, gossips, etc. The shock models unobservable variables that the econometrician cannot observe. The shock models unobservable variables that the econometrician cannot observe. The matching process is a stochastic sequence $m = \{m^t\}_{t=1}^{\infty}$ with support $I \times I$. The realizations of the meeting process are ordered pairs $m^t = \{i, j\}$, indicating which agent $i$ should play and which link $g_{ij}$ can be updated at each $t$.

The probability that player $i$ meets agent $j$, conditional on the current network is

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

(2)

The matching probability depends on the current network (e.g. the existence of a common friend between $i$ and $j$) and the characteristics of the pair. This structure includes matching technologies with a bias for same-type individuals as in Currarini et al. (2009). The simplest example of (2) is an i.i.d. discrete uniform process with $\rho(g^{t-1}, X_i, X_j) = \frac{1}{n(n-1)}$. An example with bias for same-type agents is $\rho(g^{t-1}, X_i, X_j) \propto \exp[-d(X_i, X_j)]$, where $d(\cdot, \cdot)$ is a distance function.

Utility Maximization Conditional on the meeting $m^t = ij$, player $i$ updates the link $ij$ to maximize his current utility, taking the previous period network $g_{-ij}^{t-1}$ as given. The agents have complete information since they can observe the entire network shape and the individual attributes of all agents.

Before updating his link to $j$, individual $i$ receives an idiosyncratic shock $\varepsilon \sim F(\varepsilon)$ to his preferences, that the econometrician cannot observe. The shock models unobservable variables that could influence the utility of a link, e.g. mood, gossips, etc.

Player $i$ links player $j$ at time $t$, i.e. $g_{ij}^t = 1$, if it is a best response to the current network configuration

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

(3)

The stochastic process described above generates a sequence $[g^0, g^1, \ldots, g^t]$ 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’ linking choices.

---

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

15 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 $I_{\{\ldots\}}$ is an indicator function. The transition probability is zero if the networks differ in more than one element.
### 3.1.2 Preferences

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

$$U_i(g, X) = \sum_{j=1}^{n} g_{ij} u_{ij} + \sum_{j=1}^{n} g_{ij} g_{ji} m_{ij} + \sum_{j=1}^{n} g_{ij} \sum_{k=1}^{n} g_{jk} v_{ik} + \sum_{j=1}^{n} g_{ij} \sum_{k=1}^{n} g_{ki} w_{kj}$$  \hspace{1cm} (4)$$

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

First, when the agent links another individual, she receives an additional direct net benefit $u_{ij}$. The direct utility includes both costs and benefits and it may possibly be negative: if there is homophily, the net utility $u_{ij}$ is positive if the two agents belong to the same group, while it is negative when they are of different types. 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 $u_{ij}$, since it does not require assumptions on the cost function.

Second, an individual receives additional utility $m_{ij}$ if the link is mutual. A friendship link is valued differently if the other agent reciprocates. The idea is that an agent may perceive another individual as being a friend, but that person may not perceive the relationship in the same way.

The players value the composition of friends of friends. When $i$ is deciding whether to link $j$, she observes $j$’s friends and their socioeconomic characteristics. Each of $j$’s friend provides additional utility $v(X_i, X_k)$ to the agent. In this model, a white student with three Hispanic friends is a different "good" than a white student with two white friends and one African American friend.\(^{16}\) 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. If individual $i$ links $j$, he automatically creates an indirect link for the all the agents that had a link to $i$. Thus $i$ generates an externality (positive or negative) for each $k$ that formed a link to him 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.\(^{17}\)

---

\(^{16}\) A similar assumption is used in De Marti and Zenou (2009) where the agents’ cost of linking depend on the racial composition of friends of friends. Their model is an extension of the connection model of Jackson and Wolinsky (1996), and the links are formed with mutual consent. The corresponding network is undirected.

\(^{17}\) An alternative interpretation of this last component could be that it introduces 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

I impose an additional assumption on the functional forms of the utility functions. While the assumption is not too restrictive, it provides 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.

I assume that the utility $m_{ij}$ obtained from mutual links is symmetric and that the utility of an indirect link $v_{ij}$ has the same functional form as the utility from the popularity effect $w_{ij}$.

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

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

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

therefore the utility function is

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

The symmetry in $m_{ij}$ does not imply that a mutual link between $i$ and $j$ 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 $X_i = X_j$ (say two males, whites, enrolled in eleventh grade) that form a mutual link receive the same $u_{ij}$ and $m_{ij}$, but they may have different utilities from the friendship because of the composition of their friends of friends and their popularity. Therefore, I argue that this restriction is not too strong.

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

Assumption 1 guarantees a closed form for the stationary equilibrium of the model. Without this assumption, the model still has a unique stationary equilibrium, but it is not possible to write the likelihood in closed form. In such model estimation could be performed using Approximate Bayesian Computations (see Marjoram et al. (2003) for example), but the computational burden is even more challenging.

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 deterministic incentives of any player in any state of the network are summarized by a potential function, $Q : G \times X \rightarrow \mathbb{R}$

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

and the network formation game is a Potential Game.
Proof. In Appendix A □

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

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

Consider two networks that differ only with respect to one link $g_{ij}$, chosen by individual $i$: the difference in utility that agent $i$ receives from the two networks 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 main advantage of characterizing network formation as a potential game, is that in order to compute the equilibria of the model, there is no need to keep track of the behavior of all the players: the potential function contains all the relevant information to compute the equilibria. This property is extremely useful for the analysis of networks with many players: the usual check for existence of profitable deviations from the Nash equilibrium can be performed using the potential instead of checking each player in sequence.

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

$$W(g, X) = \sum_{i=1}^{n} U_i(g, X)$$

$$= Q(g, X) + \sum_{i=1}^{n} \sum_{j=i}^{n} g_{ij}g_{ji}m_{ij} + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij}g_{ki}v_{kj}$$

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

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

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

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.\(^{18}\)

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.

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

\(^{18}\) 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 $d(\cdot, \cdot)$ is a distance function.
A Nash network is defined as a network in which any player has no incentive to modify his linking strategy, when randomly selected from the population. The following results characterize the set of the pure-strategy Nash equilibria and the long run behavior of the model 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 $\mathcal{NE}(G,X,U)$ of all pure-strategy Nash equilibria of the network formation game is completely characterized by the local maxima of the potential function.

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

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

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

**Proof.** In Appendix A □

Suppose that the network is a Nash network. As a consequence, if an agent deviates from the Nash network, he receives less utility. Since the change in utility for any agent is equivalent to the change in potential, any deviation from the Nash network must decrease the potential. It follows that the Nash network must be a local maximizer of the potential function.

Furthermore, the model will converge to one of the Nash Equilibria in the long run, independent of the initial network. 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.

In general the full model with preference shocks will have high probability of hitting a Nash network. However, the shocks allow the network to escape from such networks: this makes the model ergodic and eliminates absorbing states.

I make a parametric assumption on the shocks that provides the characterization of the stationary distribution and the transition probabilities.\(^{19}\)

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

\(^{19}\)An 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.
The probability of a link between \(i\) and \(j\), given a meeting \(m^t = ij\) and previous period network configuration \(g^{t-1}\)

\[
\Pr \left( g^t_{ij} = 1 \mid g^{t-1}_{-ij}, X \right) = \Pr \left[ \varepsilon^t_{ij}(0) - \varepsilon^t_{ij}(1) \leq U_i \left( 1, g^{t-1}_{-ij}, X \right) - U_i \left( 0, g^{t-1}_{-ij}, X \right) \right]
\]

\[
= \frac{\exp \left[ u_{ij} + g^{t-1}_{ji} m_{ij} + \sum_{k \neq i,j} g^{t-1}_{jk} v_{ik} + \sum_{k \neq i,j} g^{t-1}_{ki} v_{kj} \right]}{1 + \exp \left[ u_{ij} + g^{t-1}_{ji} m_{ij} + \sum_{k \neq i,j} g^{t-1}_{jk} v_{ik} + \sum_{k \neq i,j} g^{t-1}_{ki} v_{kj} \right]}
\] (10)

Under assumptions 1-3, the network evolves as a Markov chain with transition probability given by the conditional choice probabilities (10) and the probability law of the meeting process \(m^t\).

It can be shown that the sequence \([g^0, g^1, ..., g^t]\) is:

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

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

Intuitively, assuming \( \Pr \left( m^t = ij \right) > 0 \) for any \( ij \), implies that there is always a positive probability to reach a new network in which the link \( g_{ij} \) can 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.

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

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

\[
\lim_{t \to \infty} P \left( G^t = g \mid G^0 = g^0, X \right) = \pi \left( g, X \right)
\] (11)

2. Assume that the meeting probability of \(i\) and \(j\) does not depend on the existence of a link between them, i.e.

\[
\rho \left( g^{t-1}_{-ij}, X_i, X_j \right) = \rho \left( g^{t-1}_{-ij}, X_i, X_j \right)
\] (12)

Then the stationary distribution \( \pi(g, X) \) is

\[
\pi \left( g, X \right) = \frac{\exp \left[ Q \left( g, X \right) \right]}{\sum_{\omega \in \Omega} \exp \left[ Q \left( \omega, X \right) \right]}
\] (13)

where \( Q \left( g, X \right) \) is the potential function (6).

**Proof.** In Appendix A □

The first part of the proposition follows directly from the irreducibility and aperiodicity of the Markov process generated by the network formation game. The uniqueness of the stationary
distribution is useful in estimation, since there is no concern about multiple equilibria. Furthermore, the stationary equilibrium characterizes the likelihood of observing a specific network configuration in the data. As a consequence, I can estimate the structural parameters by observing only one network at a specific point in time.

The second part of the proposition provides a closed form solution for the stationary distribution. The intuition is straightforward: in the long run, the system of interacting agents will visit more often those states/networks that have high potential. Networks with high potential correspond to Nash equilibria described in Proposition 2. Therefore an high proportion of the possible networks generated by the network formation game, will correspond to the Nash networks.

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

\[
c(G, X) \equiv \sum_{\omega \in \Omega} \exp \left[ Q(\omega, X) \right]
\]

(14)
to guarantee that it is a proper probability distribution. However, the normalizing constant complicates the estimation, since it cannot be evaluated exactly or approximated with precision. This is explained in the estimation section.

4 Estimation Strategy

4.1 Estimation Problem

I bring the model to the data by assuming that the utility functions depend on a vector of parameters \( \theta = (\theta_u, \theta_m, \theta_v) \):

\[
\begin{align*}
    u_{ij} &= u(X_i, X_j, \theta_u) \\
    m_{ij} &= m(X_i, X_j, \theta_m) \\
    v_{ij} &= v(X_i, X_j, \theta_v)
\end{align*}
\]

The goal is to recover the parameters’ posterior distribution, given the data and the prior. Let \( p(\theta) \) be the prior distribution. Given the likelihood function \( \pi(g|X, \theta) \) of the observed data \((g, X)\), the posterior distribution of \( \theta \) can be written as

\[
p(\theta|g, X) = \frac{\pi(g|X, \theta) p(\theta)}{\int_{\Theta} \pi(g|X, \theta) p(\theta) d\theta}
\]

(15)
The estimation of the posterior imposes two computational challenges. First, the posterior depends on the normalizing integral \( \int_{\Theta} \pi(g|X, \theta) p(\theta) d\theta \). This problem is common to any Bayesian analysis, and it is usually solved using a Metropolis-Hastings algorithm that avoids direct computation of the integral. This algorithm generates a Markov chain of parameters whose unique invariant distribution is the posterior (15). The empirical distribution of the chain is used as estimate of the posterior.

At each iteration \( t \), with current parameter \( \theta_t = \theta \), a new parameter vector \( \theta' \) is proposed from a distribution \( q_\theta(\cdot|\theta) \). We compute the quantity

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

(16)
and at iteration $t + 1$ the new parameter $\theta_{t+1}$ is updated according to
\[
\theta_{t+1} = \begin{cases} 
\theta' & \text{with prob. } \alpha(\theta, \theta') \\
\theta & \text{with prob. } 1 - \alpha(\theta, \theta') 
\end{cases}
\] (17)

The appealing feature of this scheme is that in order to compute $\alpha(\theta, \theta')$ we don’t need to evaluate the integral, since the ratio of the posteriors is $p(\theta'|g, X)/p(\theta|g, X) = \frac{\pi(g|X, \theta') p(\theta')}{\pi(g|X, \theta) p(\theta)}$.

However, my model presents a second challenge: the likelihood function $\pi(g|X, \theta)$ is also known up to a normalizing constant that is computationally unfeasible.

\[
\alpha(\theta, \theta') = \min \left\{ 1, \frac{\exp[Q(g|X, \theta')]}{c(g, X, \theta')} p(\theta') q_0(\theta'|\theta) \right\} (18)
\]

\[
= \min \left\{ 1, \frac{\exp[Q(g|X, \theta')]}{c(g, X, \theta')} \frac{c(g, X, \theta)}{p(\theta')} \frac{q_0(\theta|\theta')}{q_\theta(\theta'|\theta)} \right\} (19)
\]

The Metropolis-Hastings ratio $\alpha(\theta, \theta')$ depends on the ratio $c(G, X, \theta)/c(G, X, \theta')$, that cannot be evaluated exactly. To be concrete, let me consider a small network with $n = 10$ agents. From (14) we know that $c(G, X, \theta) = \sum_{\omega \in G} \exp[Q(\omega, X, \theta)]$. In order to compute the constant at the current parameter $\theta$ we need to evaluate the exponential of the potential function for all $2^{200} \approx 10^{27}$ possible networks with 10 agents and compute their sum. This task would take a very long time even for a state-of-the-art computer. In general with a network containing $n$ players, we have to sum over $2^{n(n-1)}$ possible network configurations.\(^{20}\)

To solve this problem, I propose a variation of the exchange algorithm, first developed by Murray et al. (2006). This algorithm use a double Metropolis-Hastings step to avoid the computation of the normalizing constant in the likelihood.

### 4.2 Exchange Algorithm

The exchange algorithm allows sampling from the posterior distribution without computing the normalizing constant of the likelihood. This improvement comes with a cost: the algorithm can produce MCMC chains that have very poor mixing (Caimo and Friel, 2010) and high autocorrelation. Several authors have proposed similar algorithms in the related literature on Exponential Random Graphs Models (ERGM).\(^{21}\) However, the ERGM models estimated with this methodology have few parameters and use data from very small networks.

The idea of the algorithm is to sample from an augmented distribution using an auxiliary variable. At each iteration, the algorithm proposes a new parameter vector $\theta'$; in the second step, it samples a network from the likelihood $\pi(g', X, \theta')$; finally, the proposed parameter is accepted with a certain probability $\alpha_{ex}(\theta, \theta')$ which satisfies the detailed balance condition for the posterior

\(^{20}\)The school used in the empirical section has 150 enrolled students, i.e. $2^{22350}$ possible network configurations. The exact computation of the constant would take several months for a single iteration of the Metropolis-Hastings algorithm.

\(^{21}\)Caimo and Friel (2010) use the exchange algorithm to estimate ERGM. They improve the mixing of the algorithm using the snooker algorithm. Koskinen (2008) proposes the Linked Importance Sampler Auxiliary variable (LISA) algorithm, which uses importance sampling to provide an estimate of the acceptance probability. Another variations of the algorithm is used in Liang (2010).
distribution (15).

The original algorithm from Murray et al. (2006) is as follows

**ALGORITHM 1** Start at current parameter $\theta_t = \theta$ and network data $g$.

1. Propose a new parameter vector $\theta'$
   \[ \theta' \sim q_\theta(\cdot|\theta) \] (20)

2. Draw a sample network $g'$ with exact sampling from the likelihood
   \[ g' \sim \pi(\cdot|X,\theta') \] (21)

3. Compute the acceptance ratio
   \[
   \alpha_{ex}(\theta, \theta') = \min \left\{ 1, \frac{\exp[Q(g', X, \theta')] p(\theta') q_\theta(\theta'| \theta) \exp[Q(g, X, \theta)]}{\exp[Q(g, X, \theta)] p(\theta) q_\theta(\theta'| \theta) \exp[Q(g', X, \theta')]} \right\}
   \] (22)

4. Update the parameter according to
   \[ \theta_{t+1} = \begin{cases} 
   \theta' & \text{with prob. } \alpha_{ex}(\theta, \theta') \\
   \theta & \text{with prob. } 1 - \alpha_{ex}(\theta, \theta')
   \end{cases} \] (23)

The appeal of this algorithm is that all the quantities in the acceptance ratio (22) can be evaluated: there are no integrals or normalizing constants to compute. In Appendix B, I provide the algorithm details and the relative proofs of convergence and mixing. Below I provide some intuition.

The algorithm samples from the right distribution because it satisfies the detailed balance condition for the posterior distribution. The most difficult step for practical implementation is Step 2: we need to draw an exact sample from the stationary distribution of the model using the proposed parameter. Given the structure of the model this may require very long simulations that make the computational cost very high.

However, the requirement of an exact sample can be relaxed. If the network model is simulated using a Metropolis-Hastings algorithm that satisfies detailed balance for the stationary distribution (13), the number of iterations may be reduced drastically. Following the suggestion in Liang (2010), define $P^{(R)}_\theta(g'|g)$ as the transition probability of a Markov chain that generates $g'$ with $R$ Metropolis-Hastings updates, starting at the observed network $g$ and using the proposed parameter $\theta'$.

\[
P^{(R)}_\theta(g'|g) = P_\theta(g_1|g) P_\theta(g_2|g_1) \cdots P_\theta(g'_R|g_{R-1})
\] (24)

where $P_\theta(\cdot|\cdot)$ is the transition probability of the Metropolis-Hastings algorithm. Since the Metropolis-Hastings algorithm satisfies the detailed balance, we can prove the following

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

\[
\frac{P^{(R)}_\theta(g|g')}{P^{(R)}_\theta(g'|g)} = \frac{\exp[Q(g, X, \theta')]}{\exp[Q(g', X, \theta')]} 
\] (25)
Proof. In appendix B ■

The Lemma states that it is not necessary to get an exact sample from the stationary distribution \( \pi (\cdot , X, \theta') \) for the algorithm to work. The only requirement is that the simulations are started at the observed network and the updates follow a Metropolis-Hastings algorithm satisfying detailed balance.

The modified algorithm is as follows

**ALGORITHM 2** Start at current parameter \( \theta_t = \theta \) and network data \( g \).

1. Propose a new parameter vector \( \theta' \)
   \[
   \theta' \sim q_0(\cdot | \theta) \tag{26}
   \]

2. Draw a sample network \( g' \) from \( \pi (\cdot | X, \theta') \) using \( R \) Metropolis-Hastings updates, starting from the observed network \( g \).

3. Compute the acceptance ratio
   \[
   \alpha_{dnh}(\theta, \theta') = \min \left\{ 1, \frac{\exp \left[ Q(g', X, \theta) \right] p(\theta' | g) q_0(\theta' | \theta)}{\exp \left[ Q(g, X, \theta) \right] p(\theta | g) q_0(\theta | \theta') \frac{p^{(R)}(g'|g)}{p^{(R)}(g'|g')}} \right\} \tag{27}
   \]

4. Update the parameter according to
   \[
   \theta_{t+1} = \left\{ \begin{array}{ll}
   \theta' & \text{with prob. } \alpha_{dnh}(\theta, \theta') \\
   \theta & \text{with prob. } 1 - \alpha_{dnh}(\theta, \theta')
   \end{array} \right. \tag{28}
   \]

Intuitively, let’s assume that the current parameter \( \theta \) is a sample from the posterior distribution \( p(\cdot | g, X) \). If the proposed parameter \( \theta' \) is a good proposal, we will have \( \theta' \sim p(\cdot | g, X) \). It follows that \( g \) is a sample from the stationary distribution of the model with parameter \( \theta' \), i.e. \( g \sim \pi (g, X, \theta') \). Since \( g' \) was obtained by a Metropolis-Hastings simulation having as invariant distribution \( \pi (g, X, \theta') \), we have \( g' \sim \pi (g, X, \theta') \). Therefore the transition from \( \theta_t \) to \( \theta_{t+1} \) is not affected by the length of the simulation \( R \). However, if \( \theta' \) is a bad proposal, it will be unlikely drawn from the posterior distribution. This implies that the ratio \( \exp \left[ Q(g, X, \theta') \right] / \exp \left[ Q(g', X, \theta') \right] \) would be small, since \( g' \) is closer to the stationary distribution \( \pi (g, X, \theta') \) more than \( g \) is. At the same time also the ratio \( \exp \left[ Q(g', X, \theta) \right] / \exp \left[ Q(g, X, \theta) \right] \) is likely to be small, since \( g \sim \pi (\cdot, X, \theta) \) and \( g' \) is not and it has moved away from it. The acceptance ratio (27) is small and the proposal is likely to be rejected.

When the number of simulations is infinite \( R \to \infty \), this is equivalent to the original exchange algorithm. The improvement in computational efficiency depends on how small \( R \) can be chosen. This in general depends on the number of agents and the mixing of the algorithm used to simulate the network.

4.3 Network Simulations

A crucial step in the exchange algorithm is the simulation of samples from the stationary distribution of the model. The algorithm used in this paper is a variation of the Metropolis-Hastings algorithm proposed in Snijders (2002).
Let’s start at current network \( g \). At each iteration a random network \( g' \) is proposed from a proposal distribution \( q_g(g'|g) \) and the update is accepted with probability

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

(29)

Notice that the acceptance ratio (29) does not include the normalizing constant of the stationary distribution.

In the practical implementation of this algorithm, I use several moves and proposals. First, a move that updates only one link per iteration, proposing to swap the link value. At each iteration a random pair of agents \((i, j)\) is selected from a discrete uniform distribution, and it is proposed to swap the value of the link \( g_{ij} \) to \( 1 - g_{ij} \). Second, to improve the convergence I allow the sampler to propose bigger moves: instead of proposing to swap only one link, it proposes to swap the entire network matrix.\(^{22}\) With a small probability \( p_{inv} \), the sampler proposes a new network \( g' = 1 - g \), which is accepted with probability \( \alpha_{mh}(g, g') \).\(^{23}\)

### 4.4 Likelihood Function

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

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

\[ u_{ij} = u(X_i, X_j, \theta_u) = \sum_{p=1}^{P} \theta_{up} H_{up} (X_i, X_j) = \theta_u' H_u (X_i, X_j) \]

\[ m_{ij} = m(X_i, X_j, \theta_m) = \sum_{l=1}^{L} \theta_{ml} H_{ml} (X_i, X_j) = \theta_m' H_m (X_i, X_j) \]

\[ v_{ij} = v(X_i, X_j, \theta_v) = \sum_{s=1}^{S} \theta_{vs} H_{vs} (X_i, X_j) = \theta_v' H_v (X_i, X_j) \]

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. We can consider different specifications and include different sets 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.

---

\(^{22}\)This move is suggested in Geyer (1992) and Snijders (2002). Snijders (2002) argues that this is particularly useful in case of a bimodal distribution.

\(^{23}\)I also experimented with the Simulated Tempering algorithm proposed in Mele (2010). This algorithm is especially useful when the distribution has more than one mode and it improves the mixing of the chain. However it does so by increasing the time needed to collect a sample. Since in this context there is virtually no difference between the Simulated Tempering results and the simpler Metropolis-Hastings updates, I use the latter in the rest of the paper.
PROPOSITION 3. (Exponential Family Likelihood) Under assumptions 1-3, the stationary distribution $\pi(g, X)$ belongs to the exponential family, i.e. it can be written in the form

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

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

Proof. In Appendix A ■

The vector $t(g, X) = (t_1(g, X), ..., t_K(g, X))$ is a vector of statistics for the network formation model. This vector can contain the number of links, the number of whites-to-whites links, the number of male-to-female links and so on. Interactions 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. Furthermore, the statistics are sufficient.

The likelihood is very similar to the one of an exponential random graph model (Snijders, 2002). My model can be thought of as a microeconomic foundation of the exponential random graph models. In this sense, we can interpret the ERGM as the stationary equilibrium of a strategic game of network formation with myopic agents following a stochastic best response dynamics, when utility are linear functions of the parameters.

4.5 Practical Implementation

I choose the priors that have high variability

$$
p(\theta) = N(0, 3I_P)
$$

where $P$ is the number of parameters. I use a random walk Metropolis-Hastings algorithm to sample from the posterior, with proposal distribution

$$
q_\theta(\cdot|\theta) = N(0, \Sigma)
$$

I use an adaptive procedure to determine a suitable $\Sigma$. I start the iterations with $\Sigma = I_P$; I run the chain and monitor convergence using standard methods. I then estimate the covariance matrix of the chains and use it as an approximate $\Sigma$.

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

The posterior distributions shown in the graphs are obtained with a simulation of 50000 Metropolis-Hastings updates of the parameters. These simulations start from values found after long experimentation with different starting values and burn-in periods, monitoring convergence using standard methods. For each parameter update, I simulate the network for 3000 iterations to collect a sample from the stationary distribution.
Figure 1: A School Network

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

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

5 Empirical Results

5.1 Parameter Estimates

In this section I analyze School 25 of the Add Health dataset. The school has 150 students and 58.7% females, with a total of 355 friend nominations. The clustering coefficient is 0.2906 and the racial fragmentation is 0.606. The racial composition is as follows: 42% whites, 45.3% blacks, 0.667% asians, 10.6% hispanics. Figure 1 shows the friendship network of this school: each dot represents a student, the color represents his racial group and an arrow is a friend nomination.

The results for three alternative specifications of the model are presented in Table 3. I report the posterior mean and standard deviation. Each estimate measures the marginal effect of the variable: for example, the parameter associated with the direct utility of white-white measures the marginal utility of a white individual forming a link to another white, other things being equal.

The first column contains posterior mean and standard deviation of a specification in which the direct utility is a function of total number of links (constant), total number of links in which both are Whites, Blacks or Hispanic. This specification tests for the presence of differential homophily:
Table 3: Three Specifications, School 25

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

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

Each racial group may have different homophily levels. The other controls are for the number of reciprocated links (mutual constant) and for the number of indirect friends (friends of friends).

The output of the simulations is shown in Figure 4, while in Figure 5 I present the estimated marginal posterior distribution of each parameter. These results points to strong racial homophily effects for each racial group. Each additional link is costly as indicated by the negative coefficient of the constant. However, an additional link is less costly if the pair belongs to the same racial group: all the homophily coefficients are positive. A mutual link increases utility as expected, while linking to an individual with many friends decreases it. The latter effect can be due to congestion: an individual with many links has less time to devote to each of his friends.\(^{24}\)

Figure 6 shows the autocorrelations of the simulated parameters: these are very high and they disappear after 350 lags. This is a problem of the exchange algorithm: it produces samples from the correct posterior distribution, but it does so with a slow mixing chain and high autocorrelated

---

\(^{24}\) At the same time one should notice that the homophily effect for Hispanics is estimated with higher variability: this is because there are very few Hispanics in the dataset, and they form few links. A partial solution is to run more simulations. Alternatively one could estimate a model with multiple schools and exploit the variability among schools as a source of identification.
output. Furthermore, the acceptance of proposed parameters is very low, in the order of 10%. The problem is more serious when we increase the number of parameters.

The marginal posterior distributions for Model 2 are shown in the third and fourth column of Table 3 and in Figure 7. Model 2 includes controls for the racial composition of mutual friends and friends of friends. This model confirms the existence of homophily in direct links, but also in mutual and indirect links. The only exception is for links that involve Hispanics: mutual and indirect links decrease utility.

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

5.2 Policy Experiments

The model can be used to assess the effectiveness of busing programs in promoting interracial integration. In this section I simulate two policies. The first is a reassignment of 8 African-Americans from school 94 to school 25 of Add Health. The second is a reassignment of 16 Hispanic students from school 94 to school 25 of Add Health. In both cases I compute the segregation levels in the stationary equilibrium before and after the implementation of the policy. I use the Freeman’s Segregation Index (see Freeman (1972)) to measure segregation for the three relevant groups: Whites, African-Americans and Hispanics.

The results are reported in Figure 3. The first row shows the distribution of the predicted segregation without policy (blue) and after the implementation of the policy (red) when we reassign 8 African-Americans from school 94 to school 25. For all the racial groups the average segregation goes down (the vertical line) and there is an increase in variance. The second row shows that when Policy 2 is implemented, the results change. While for both White and African American students the segregation decreases, segregation for Hispanics increases.

The intuition is that the increase of 16 Hispanics makes it easier for an Hispanic student to meet another Hispanic student to form a friendship with. Since there is strong homophily, these type of links will be formed with high probability, resulting in higher equilibrium segregation for this group.

6 Conclusions

In this paper, I propose a structural model of strategic network formation that generates the clustering and segregation observed in the data as an equilibrium outcome. I develop Bayesian methods of estimation to recover the posterior distribution of structural preference parameters, determining the aggregate levels of segregation. I use data from Add Health, containing detailed information on the actual social network of all the students enrolled in a representative sample of US high schools. My results show that individuals have strong preference for interaction with the students of the same racial group. I use the model to simulate alternative Busing programs, showing that sometimes these policies may have an adverse effect on the levels of racial segregation within the school.

My model mixes ingredients from both the strategic and the random network formation literature, that guarantee the existence of a unique stationary equilibrium. This is in contrast with
Estimated marginal posterior distributions for Model 3. The blue solid line is the estimated posterior, obtained by kernel smoothing of the simulation output. The red line indicates the estimated posterior mean. The dotted black line is the prior distribution $\mathcal{N}(0, 3 \times I_{18})$. The posterior is estimated with a simulated sample of 50000 parameters.

most strategic models of network formation where multiplicity of equilibria is the rule, a feature that makes estimation and identification extremely challenging. Here the uniqueness of equilibrium allows estimation using only one network observation at a single point in time. When multiple
The graphs show the distribution and average of Freeman’s Segregation Index for the 3 racial groups before the policy (BLUE) and after the policy is implemented (RED). The graphs in the upper row show Policy 1, a reassignment of 8 African-American students from school 94 to school 25 of Add Health. The graphs in the lower row refer to Policy 2, a reassignment of 16 Hispanic students from school 94 to school 25 of Add Health.

networks are observed independently the estimation strategy is easily adapted.

The possibility of running counterfactual policy experiments makes this model extremely useful for policy purposes. One limitation is that, when considering the effect of a policy, I am abstracting from the effect of the social network on individual outcomes. Future research should model the co-evolution of the social network and individual actions, in order to obtain more accurate predictions of the policy impacts.

References


Haakon Austad and Nial Friel. Deterministic bayesian inference for the p* model. Journal of


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


A Proofs

Proof of Proposition 1

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

$$Q (g_{ij} = 1, g_{-ij}, X) - Q (g_{ij} = 0, g_{-ij}, X) = u_{ij} + g_{ji}m_{ij} + \sum_{k=1}^{n} g_{jk}v_{ik} + \sum_{k=1}^{n} g_{ki}v_{kj}$$

$$= U_i (g_{ij} = 1, g_{-ij}, X) - U_i (g_{ij} = 0, g_{-ij}, X)$$

\[25\]

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

$$W(g, X) = \sum_{i=1}^{n} U_i(g, X)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}u_{ij} + \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ji}g_{ij}m_{ij} + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1, k \neq i,j}^{n} g_{ij}g_{jk}v_{ik} + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1, k \neq i,j}^{n} g_{ij}g_{ki}v_{kj}$$

$$= Q(g, X) + \sum_{i=1}^{n} \sum_{j=i}^{n} g_{ij}g_{ji}m_{ij} + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=1, k \neq i,j}^{n} g_{ij}g_{ki}v_{kj}$$

**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

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

$$= \sum_i \sum_j Pr(m^{t+1} = i j) Pr[U_i\left(g_{ij}^{t+1}, g_{-ij}^t, X\right) \geq U_i\left(g_{ij}^t, g_{-ij}^t, X\right) | m^{t+1} = i j] = 1$$

because agents play Best Response, conditioning on $m^{t+1}$

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

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

\[
\lim_{t \to \infty} P_r \left[ g^t \in NE \mid g^0 \right] = 1
\]

**Proof of Theorem 1**

1. The sequence of networks \([g^0, g^1, \ldots]\) 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 form the ergodic theorem.

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

\[
P_{gg'} \pi_g = P_{g'g} \pi_{g'}
\]

where

\[
P_{gg'} = \Pr \left( g^{t+1} = g' \mid g^t = g \right)
\]

\[
\pi_g = \pi \left( g^t = g \right)
\]

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

\[
P_{gg'} \pi_g = \Pr \left( m^t = ij \right) \Pr \left( g_{ij} = 0 \mid g_{-ij} \right) \frac{\exp \left[ Q (1, g_{-ij}, X) \right]}{\sum_{\omega \in \mathcal{G}} \exp \left[ Q (\omega, X) \right]}
\]

\[
= \rho (g_{-ij}, X_i, X_j) \times \frac{1}{1 + \exp \left[ \Delta Q \right]} \times \frac{\exp \left[ Q (1, g_{-ij}, X) \right]}{\sum_{\omega \in \mathcal{G}} \exp \left[ Q (\omega, X) \right]}
\]

\[
= \rho (g_{-ij}, X_i, X_j) \times \frac{1}{1 + \exp \left[ \Delta Q \right]} \times \frac{\exp \left[ Q (1, g_{-ij}, X) \right]}{\sum_{\omega \in \mathcal{G}} \exp \left[ Q (\omega, X) \right]}
\]

\[
= \rho (g_{-ij}, X_i, X_j) \frac{\exp \left[ \Delta Q \right]}{1 + \exp \left[ \Delta Q \right]} \frac{\exp \left[ Q (0, g_{-ij}, X) \right]}{\sum_{\omega \in \mathcal{G}} \exp \left[ Q (\omega, X) \right]}
\]

\[
= \Pr \left( m^t = ij \right) \Pr \left( g_{ij} = 1 \mid g_{-ij} \right) \frac{\exp \left[ Q (0, g_{-ij}, X) \right]}{\sum_{\omega \in \mathcal{G}} \exp \left[ Q (\omega, X) \right]}
\]

\[
= P_{g'g} \pi_{g'}
\]

So the distribution (13) satisfies the detailed balance condition. Therefore it is a stationary distribution for the network formation model. From part 1) of the proposition, we know that the process is ergodic and it has a unique stationary distribution. Therefore \( \pi (g, X) \) is also the unique stationary distribution.
Proof of Proposition 3

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

\[
\sum_i \sum_j g_{ij} u_{ij} = \sum_i \sum_j g_{ij} \sum_{p=1}^P \theta_{up} H_{up}(X_i, X_j) = \sum_{p=1}^P \theta_{up} \sum_i \sum_j g_{ij} H_{up}(X_i, X_j)
\]

define \( t_{up}(g, X) \equiv \sum_i \sum_j g_{ij} H_{up}(X_i, X_j) \), therefore

\[
= \sum_{p=1}^P \theta_{up} t_{up}(g, X) = \theta'_{u} t_{u}(g, X)
\]

where \( \theta_u = (\theta_{u1}, \theta_{u2}, ..., \theta_{uP})' \) and \( t_{u}(g, X) = (t_{u1}(g, X), t_{u2}(g, X), ..., t_{uP}(g, X))' \). Analogously define \( \theta_m = (\theta_{m1}, \theta_{m2}, ..., \theta_{mL})' \) and \( t_{m}(g, X) = (t_{m1}(g, X), t_{m2}(g, X), ..., t_{mL}(g, X))' \) and \( \theta_v = (\theta_{v1}, \theta_{v2}, ..., \theta_{vS})' \) and \( t_{v}(g, X) = (t_{v1}(g, X), t_{v2}(g, X), ..., t_{vS}(g, X))' \). It follows that

\[
\sum_i \sum_{j>i} g_{ij} g_{ji} m_{ij} = \sum_i \sum_{j>i} g_{ij} g_{ji} \sum_{l=1}^L \theta_{ml} H_{ml}(X_i, X_j) = \sum_{l=1}^L \theta_{ml} \sum_i \sum_{j>i} g_{ij} g_{ji} H_{ml}(X_i, X_j) = \sum_{l=1}^L \theta_{ml} t_{ml}(g, X) = \theta'_{m} t_{m}(g, X)
\]

and

\[
\sum_i \sum_j g_{ij} \sum_{k \neq i,j} g_{jk} v_{ij} = \sum_i \sum_j g_{ij} \sum_{k \neq i,j} g_{jk} \sum_{s=1}^S \theta_{vs} H_{vs}(X_i, X_k) = \sum_{s=1}^S \theta_{vs} \sum_i \sum_{j \neq i,j} g_{ij} \sum_{k \neq i,j} g_{jk} H_{vs}(X_i, X_k) = \sum_{s=1}^S \theta_{vs} t_{vs}(g, X) = \theta'_{v} t_{v}(g, X)
\]
Therefore \( Q(g, X) \) can be written in the form \( \theta' t(g, X) \), where \( \theta = (\theta_u, \theta_m, \theta_v)' \) and \( t(g, X) = [t_u(g, X), t_m(g, X), t_v(g, X)]' \):

\[
Q(g, X) = \theta'_u t_u(g, X) + \theta'_m t_m(g, X) + \theta'_v t_v(g, X) = \theta' t(g, X)
\]

and the stationary distribution is then

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

## B Computational Details

In this section I provides the technical details for the algorithm proposed in the empirical part of the paper. The first set of results show that the exchange algorithm provides indeed samples from the posterior distribution (15).

The exchange algorithm works because it satisfies detailed balance condition for the posterior distribution, i.e. for any couple of parameters \((\theta_i, \theta_j) \in \Theta\) we have

\[
Pr[\theta_j|\theta_i, g, X] p(\theta_i|g, X) = Pr[\theta_i|\theta_j, g, X] p(\theta_j|g, X)
\]

The detailed balance condition is sufficient condition for the Markov chain generated by the algorithm to have stationary distribution the posterior (15) (for details see Robert and Casella (2005) or Gelman et al. (2003)).

**Lemma 2** The exchange algorithm produces a Markov chain with invariant distribution (15).

**Proof.** Define \( Z \equiv \int_\Theta \pi(g|X, \theta) \rho(\theta) \ d\theta \). In the algorithm the probability \( Pr[\theta_j|\theta_i, g, X] \) of transition to \( \theta_j \), given the current parameter \( \theta_i \) and the observed data \((g, X)\), can be computed as

\[
Pr[\theta_j|\theta_i, g, X] = q_\theta(\theta_j|\theta_i) \frac{\exp[Q(g', X, \theta_j)]}{c(G, X, \theta_j)} \alpha_{ex}(\theta_i, \theta_j)
\]

This is the probability of proposing \( \theta_j \), \( q_\theta(\theta_j|\theta_i) \), times the probability of generating the new network \( g' \) from the model's stationary distribution, \( \frac{\exp[Q(g', X, \theta_j)]}{c(G, X, \theta_j)} \) and accepting the proposed parameter.
\( \alpha_{ex}(\theta_i, \theta_j) \). Therefore the left-hand side of (34) can be written as

\[
\Pr \left[ \theta_j | \theta_i, g, X \right] p(\theta_i | g, X) = \frac{q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_j) \right]}{c(G, X, \theta_j)} \alpha_{ex}(\theta_i, \theta_j) p(\theta_i | g, X)
\]

\[
= \frac{q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_j) \right]}{c(G, X, \theta_j)} \alpha_{ex}(\theta_i, \theta_j) \frac{\exp \left[ Q(g, X, \theta_j) \right] p(\theta_i)}{Z}
\]

\[
= \frac{q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_j) \right]}{c(G, X, \theta_j)} \alpha_{ex}(\theta_i, \theta_j) \frac{\exp \left[ Q(g, X, \theta_j) \right] p(\theta_i)}{Z}
\]

\[
= \min \left\{ q_0(\theta_j | \theta_i) \frac{\exp \left[ Q(g', X, \theta_j) \right] \exp \left[ Q(g, X, \theta_j) \right] p(\theta_i)}{c(G, X, \theta_j) Z}, q_0(\theta_i | \theta_j) \frac{\exp \left[ Q(g', X, \theta_i) \right] \exp \left[ Q(g, X, \theta_i) \right] p(\theta_j)}{c(G, X, \theta_i) Z} \right\}
\]

\[
= \min \left\{ \frac{q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_i) \right] \exp \left[ Q(g, X, \theta_i) \right] p(\theta_j)}{c(G, X, \theta_i) Z}, \frac{q_0(\theta_i | \theta_j) \exp \left[ Q(g', X, \theta_i) \right] \exp \left[ Q(g, X, \theta_i) \right] p(\theta_j)}{c(G, X, \theta_i) Z} \right\}
\]

\[
= \frac{q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_j) \right] \exp \left[ Q(g, X, \theta_i) \right] p(\theta_i)}{c(G, X, \theta_j) Z} \times \min \left\{ 1, \frac{\exp \left[ Q(g', X, \theta_j) \right] p(\theta_j)}{\exp \left[ Q(g, X, \theta_i) \right] p(\theta_i)} \right\}
\]

\[
= q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_i) \right] \exp \left[ Q(g, X, \theta_i) \right] \frac{p(\theta_j)}{c(G, X, \theta_j) Z} \times \min \left\{ 1, \frac{\exp \left[ Q(g', X, \theta_j) \right] p(\theta_j)}{\exp \left[ Q(g, X, \theta_i) \right] p(\theta_i)} \right\}
\]

\[
= q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_i) \right] \exp \left[ Q(g, X, \theta_i) \right] \frac{\alpha(\theta_j, \theta_i) \exp \left[ Q(g, X, \theta_j) \right] p(\theta_j)}{c(G, X, \theta_j) Z}
\]

\[
= q_0(\theta_j | \theta_i) \exp \left[ Q(g', X, \theta_i) \right] \exp \left[ Q(g, X, \theta_i) \right] \frac{\alpha(\theta_j, \theta_i) p(\theta_j | g, X)}{c(G, X, \theta_i)}
\]

\[
= \Pr \left[ \theta_j | \theta_i, g, X \right] p(\theta_j | g, X)
\]

The latter step proves the detailed balance for a generic network \( g' \). Since the condition is satisfied for any network, the detailed balance follows. ■

**Proof of Lemma 1**

The main computational difficulty in the algorithm is the second step, that requires an exact sample from the distribution \( \pi(g, X, \theta') \). Exact sampling may be very costly in terms of computational time, since the model’s convergence to the stationary distribution is very slow.

Following a suggestion in Liang (2010), we can show that for this model it is sufficient to run a simulation of moderate size, starting at the observed network. Lemma 1 shows that if we sample from the stationary distribution of the model using a Metropolis-Hastings algorithm satisfying detailed balance for \( \pi(g, X, \theta') \), we only need a finite number of network updates.

Let \( \mathcal{P}^{(R)}_{\theta'}(g'|g) \) be defined as in (24). This is the transition probability of the chain that generates \( g' \) with \( R \) Metropolis-Hastings updates, starting at the observed network \( g \) and using the proposed parameter \( \theta' \). Notice that the Metropolis-Hastings algorithm satisfies the detailed balance for \( \pi(g, X, \theta') \), therefore we have

\[
\mathcal{P}^{(R)}_{\theta'}(g'|g) \pi(g', X, \theta') = \mathcal{P}^{(R)}_{\theta'}(g'|g) \mathcal{P}^{(R)}_{\theta'}(g'|g) \cdots \mathcal{P}^{(R)}_{\theta'}(g'|g) \pi(g', X, \theta')
\]

\[
= \mathcal{P}^{(R)}_{\theta'}(g|g) \mathcal{P}^{(R)}_{\theta'}(g|g) \cdots \mathcal{P}^{(R)}_{\theta'}(g|g) \pi(g, X, \theta')
\]

\[
= \mathcal{P}^{(R)}_{\theta'}(g'|g) \pi(g, X, \theta')
\]
It follows that

\[
\frac{\mathcal{P}_{\theta'}^{(R)}(g'|g)}{\mathcal{P}_{\theta'}^{(R)}(g'|g)} = \frac{\pi(g, X, \theta')}{\pi(g', X, \theta')}
= \frac{\exp[Q(g, X, \theta')] c(\mathcal{G}, X, \theta')}{\exp[Q(g', X, \theta')] c(\mathcal{G}, X, \theta')}
= \frac{\exp[Q(g, X, \theta')]}{\exp[Q(g', X, \theta')]} \cdot \frac{c(\mathcal{G}, X, \theta')}{c(\mathcal{G}, X, \theta')}
\]

C Additional Figures

Figure 4: Output of Exchange algorithm

Markov chains of parameters obtained with the exchange algorithm. The red line indicates the estimated posterior mean.
Figure 5: Marginal Posterior distributions, Model 1

Estimated posterior distributions of Model 1. The blue solid line is the estimated posterior, obtained by kernel smoothing of the simulations. The red line indicates the estimated posterior mean. The dotted black line is the prior distribution $\mathcal{N}(0, 3 \times I_6)$. 
Autocorrelation functions of the chain produced by the exchange algorithm. The autocorrelation of the series goes to zero very slowly, especially for the parameters of the mutual and indirect part of utility.
Estimated posterior distributions of Model 2. The blue solid line is the estimated posterior, obtained by kernel smoothing of the simulations. The red line indicates the estimated posterior mean. The dotted black line is the prior distribution $\mathcal{N}(0, 3 \times I_{12})$. The posterior is estimated with a simulated sample of 50000 parameters.