2017-10

# The Estimation of Network Formation Games with Positive Spillovers

Vincent Boucher

Septembre / September 2017

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

**www.crrep.ca**

# Abstract

I present a behavioural model of network formation with positive network externalities in which individuals have preferences for being part of a clique. The behavioural model leads to an associated supermodular (Topkis, 1979) normalform game. I show that the behavioural model converges to the greatest Nash equilibrium of the associated normal-form game. I propose an approximate Bayesian computation (ABC) framework, using original summary statistics, to make inferences about individuals' preferences, and provide an illustration using data on high school friendships.

JEL Codes: D85, C11, C15, C72
Keywords: Network Formation, Supermodular Games, Approximate Bayesian Computation

**Boucher:** Department of Economics, Universitée Laval, CRREP and CREATE,
vincent.boucher@ecn.ulaval.ca

**Acknowledgments**

This paper benefited from many discussions, comments and suggestions. I am particularly grateful to Luc Bissonnette and Arnaud Dufays for their many insightful comments and suggestions. Thank you also to Aureo de Paula for his many great comments! I would also like to thank Charles Bellemare, Bernard Fortin, Marion Goussé, Sergei Izmalkov, David Karp, Angelo Mele and participants of the "Séminaires du Jeudi Midi" at Université Laval and the annual workshop of the Coalition Theory Network (2016) for their helpful comments. I also gratefully acknowledge financial support from the Fonds de Recherche du Québec - Société et culture and from the Social Sciences and Humanities Research Council.

This research uses data from Add Health, a program directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health website (`http://www.cpc.unc.edu/addhealth`). No direct support was received from grant P01-HD31921 for this research.

## 1. Introduction

Friendships are often characterized by tightly connected groups. These are formally referred to as *cliques*, i.e. groups of individuals, each of whom is linked to all the others.[1] Indeed, some of the strongest friendships revolve not around two people but around groups of friends. An interesting feature of cliques is that it requires coordination among the individuals involved. Every member of a clique must want to be friends with every other member of the clique; otherwise, a clique will fall apart.[2]

In this paper, I present a model of network formation with positive network externalities in which individuals have preferences for being part of a clique. In particular, I present a behavioural model in which individuals have the opportunity to jointly revise their strategies. I characterize the equilibrium network and, although the structure of the equilibrium cannot be expressed analytically, I show that it can easily be simulated. Building on this key property, I propose an approximate Bayesian computation (ABC) framework to make inferences about individuals' preferences. I present an empirical illustration using friendship networks.

I present a behavioural model in which, at any point in time, a leader, whose identity may depend on the network structure, tries to implement a change in the network structure. The evolution the the network structure is shaped by the leader's strategy to implement his desired change. I show that the behavioural model converges to the greatest NE of an associated normal-form game.

Specifically, I show that the associated game is supermodular (Topkis, 1979) and focus on the greatest Nash equilibrium of that game.[3] If it is known that the greatest equilibrium of the game is coalition-proof (Milgrom & Roberts, 1994) and coalition-proof stochastically stable (Newton, 2012), I show that, in this

---

[1] For example, see Jackson (2008), chapter 2.2.3.

[2] See section 2 of this paper for details.

[3] Miyauchi (2016) uses a different approach and considers the set of all Nash equlibria. Hellmann (2013) also consider models with strategic complementarities, but focuses on pairwise stability. Harrison & Muñoz (2008) also study different equilibrium selection mechanisms for supermodular network formation games. Finally, Xu & Lee (2015) also focus on the greatest NE in a different context.

particular context, the greatest NE is also a strong Nash equilibrium (Aumann, 1959).

I develop a random utility model (McFadden, 1980) based on the microeconomic model of network formation. I propose an ABC framework that allows for the simulation of the posterior distribution of the unknown preference parameters. If the method is well suited to the study of the behavioural model presented, it is general enough to allow for the estimation of any network formation model for which the equilibrium network can be simulated rapidly (e.g. Miyauchi (2016) or Boucher (2017)).

I propose an original method to select the set of summary statistics. In the spirit of indirect inference (Gourieroux et al., 1993), I use the moments of an auxiliary model. The statistics are similar in spirit to the ones suggested by Fearnhead & Prangle (2012), although my approach is simpler to implement for network formation models. I also discuss the interpretation of insufficient summary statistics and show that they lead to identification issues, similar to those arising from General Method of Moments estimators (Gallant & Tauchen, 1996).

I argue that the Bayesian setting is well suited to models of network formation as they do not rely on identification and asymptotic conditions, which can rarely be expressed explicitly as a function of the model's primitive conditions. See the discussion in subsection 1.1.

I apply my method to the formation of friendships among high school students. The simulated nature of the estimation strategy allows for more flexibility than most existing models. In particular, it allows for non-symmetric preference shocks. Since friendships are relatively scarce (individuals have few friends relative to the size of the population), I estimate a model with non-symmetric errors, in the spirit of conditional log-log models.

The remainder of the paper is organized as follows. In the rest of this section, I review the recent literature. In section 2, I present the behavioural model, the associated normal-form game, as well as their properties. In section 3, I present the algorithm that allows the equilibrium network to be simulated. In section

4, I present the ABC methodology. In section 5, I present an illustration using the formation of friendship networks. I conclude in section 6.

### *1.1. Context*

This paper contributes to the empirical literature on strategic network formation.[4] Making inferences from network data is challenging due to two main issues: computational complexity and unknown asymptotic behaviour.

Computational complexity stems from the fact that the number of potential networks over a population of size $n$ is $2^{n(n-1)}$. Then, even for small populations, the number of possible network structures is intractable (e.g. $2^{90}$ for $n = 10$), leading to a *curse of dimensionality.* Non-standard asymptotic properties result from the strong dependence between links, e.g. the friend of my friend is more likely to be my friend than a random individual. In this section, I review some key contributions to the literature in light of those two challenges.

The first strand of the literature presents models based on asymptotic theory. Boucher & Mourifié (2012,2016) followed by Leung (2014) argue that *homophily*, the empirical fact that individuals with similar characteristics are more likely to be linked, implies a generalization of the $\alpha/\phi$-mixing properties found in the spatial econometrics literature.[5] Asymptotic independence is achieved as the distance (in terms of observed characteristics) between individuals goes to infinity. In Boucher & Mourifié (2012,2016), the computational challenge is solved by using an estimator based on the conditional distributions. In Leung (2014), it is done through the definition of feasible bounds, covering the identified set. Sheng (2012, 2014) also use tractable bounds to circumvent the computational complexity of the model.

Chandrasekhar & Jackson (2014, 2015) study the asymptotic property of exponential random graph models (ERGMs), they solve the computational challenge, by focussing on networks formed of a finite collection of subgraphs and

---

[4] See Chandrasekhar (2015) and De Paula (2015) for recent reviews.

[5] See Chandrasekhar (2015) for an intuitive presentation of the argument. Some papers also focus solely on homophilic preferences, such as Currarini et al. (2009) and Boucher (2015).

by exploiting the fact that networks with identical features have the same probability of being observed. If the network is relatively sparse, a simple estimator can be obtain by iteratively counting the subgraphs.

Leung (2015b) defines a two-step estimation procedure and limits the dependence between links by focusing on a game of incomplete information and by assuming an anonymity condition, while Leung (2015a) uses an underlying network of potential links in order to limit the dependence between links.

Finally, Graham (2016a) and Graham (2016b) present models with unobserved individual characteristics. In Graham (2016a) the focus is on a static model of network network formation, whereas in Graham (2016b), he presents a small-$T$ panel model, where the probability that link is formed can be a function of the network structure at the previous period.

The second strand of the literature focuses on Bayesian inference, or abstracts from the asymptotic properties of the models presented (e.g. Badev (2013), Mele (2016)).[6] Both models are based on an ERGM, leading to a likelihood which has a closed-form expression, but for which the denominator is intractable.[7] The likelihood can therefore be simulated using Markov chain Monte Carlo (MCMC) simulations.[8] A particular concern, however, is that the Markov chain only converges to the likelihood as time goes to infinity, so the estimation requires an extremely large number of simulations in order to convincingly approximate the likelihood. To circumvent this problem, Mele (2016) proposes an approximate exchange algorithm that only requires a finite number of draws in order to simulate the likelihood. Finally, Mele & Zhu (2016) circumvent the computational challenge by proposing a variational approximation method to approximate the intractable likelihood and show that their approximation is asymptotically exact.

---

[6] That being said, Mele (2016) presents a “large network” analysis and describes the properties of his inference strategy as the size of the network goes to infinity. Badev (2013) presents a maximum likelihood estimator (MLE) where the intractable likelihood is approximated using Markov chain Monte Carlo simulations.

[7] Formally, an ERGM is such that $P(G) = \frac{\exp\{Q(G)\}}{\sum_H \exp\{Q(H)\}}$ for some function $Q$.

[8] e.g. Metropolis-Hasting algorithms, see Mele (2016) for details.

In this paper, I also focus on Bayesian inference and present a micro-founded model for which the likelihood function has no closed-form expression, but can easily be simulated. Since the algorithm converges in finite time, the simulation process draws from the true density of the model with probability 1. The general ABC approach presented allows for much more flexibility to specify individual preferences and unobservable random shocks.

## 2. Model

The model consists of a population composed of $n \geq 3$ individuals. Individuals interact in a *directed network*, represented by a $n \times n$ binary matrix $G$ where $g_{ij} = 1$ means that $i$ is *linked* to $j$, and $g_{ij} = 0$ otherwise. Let $g_i$ represent the $i^{\text{th}}$ row in $G$, and $G_{-i}$ be the matrix where the $i$'$^{\text{th}}$ row is removed, so $(g_i, G_{-i}) = G$. Similarly, I also use the notation $G_{-ij}$, such that $G = (g_{ij}, G_{-ij})$.

Individuals have preferences over the whole network structure $G$. Preferences are represented by a utility function $U_i(g_i, G_{-i})$. I assume that there exists at least one $i \in N$ such that $U_i(g_i, G_{-i}) \neq U_i(h_i, G_{-i})$ whenever $g_i \neq h_i$.[9]

For example, preferences could be separable across links, as in $U_i(G) = \sum_{j \neq i} g_{ij} V_{ij}(G_{-i})$, where $V_{ij}$ represents the value for $i$ of a link with $j$, given $G_{-i}$. Such separable models can be found in Badev (2013), Leung (2015b) and Mele (2016), among others. Separable models imply that the linking decisions can be decomposed into link-specific decisions, which restricts the scope of analysis (see below, as well as Bramoullé & Fortin (2010) for a discussion).

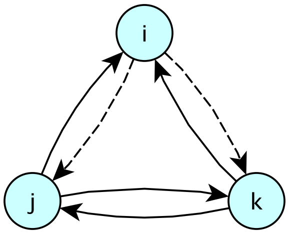
I am interested in non-separable models, where the best-response functions cannot always be decomposed into link-specific decisions. In particular, I am interested in capturing the strategic incentives inherent to the value of cliques. Figure 1 presents an example of the creation of a clique. In Figure 1, $j$ and $k$ are linked with each other, and also have links with $i$. If $i$ creates links with $j$ and $k$, there will be a clique between $i$, $j$ and $k$. This clique is assumed to generate

---

[9] This assumption plays the same role as a tie-breaking rule, ensures uniqueness and simplifies the analysis by removing uninteresting cases. In section 5, I present a random utility model for which this condition generically holds.

an additional benefit for $i$ (as well as for $j$ and $k$). An important consideration is that it is possible for $i$ to prefer to create a link with $j$ *and* $k$, even if he may not want to create either link independently. This motivates the study of non-separable models.

Figure 1: Benefit From Cliques



Since I'm interested in the prevalence of cliques, a natural assumption is therefore to consider positive network externalities.[10] I assume the following:

**Assumption 1 (Main Assumption).** *For all* $ij$, $U_i(1, G_{-ij}) - U_i(0, G_{-ij})$ *is non-decreasing in* $G_{-ij}$.

This assumption implies that incentive to create links increase as links are added to the network. Assumption 1 allows to represent situations where linking decisions are costly, while indirect links have positive value (e.g. Bala & Goyal (2000), Galeotti et al. (2006), Fabrikant et al. (2003) and Miyauchi (2016), as well as the connection model from Jackson & Wolinsky (1996)). However, the assumption abstracts away from negative externalities (e.g. the coauthor model from Jackson & Wolinsky (1996)). See also the discussion in Miyauchi (2016).

I argue that Assumption 1 is likely to hold in situations where a high degree of clustering (e.g. cliques) is observed. I discuss the intuition in Example 1.

**Example 1.** *Assume that the preferences are such that:*

$$U_i(g_i, G_{-i}) = \sum_{j \neq i} g_{ij} V_{ij}(G_{-ij}).$$

---

[10] Even if homophily - the empirical fact that similar individuals have a higher probability of creating a link - can create a certain amount of clustering, the (observed) homophily may not be enough to replicate empirical facts. See for instance Graham (2016b) for a recent discussion.

*This model is intuitive as the function $V_{ij}$ can be interpreted as the value for i of a link with j (conditional of the set of other links). Note that this model is in general not separable since $V_{ij}(G_{-ij})$ is a function of $g_i$. Here, we have:*

$$\begin{aligned} U_i(1, G_{-ij}) - U_i(0, G_{-ij}) &= V_{ij}(G_{-ij}) + \\ &\sum_{k \neq i,j} g_{ik}[V_{ik}(g_{ij} = 1, G_{-ik-ij}) - V_{ik}(g_{ij} = 0, G_{-ik-ij})] \end{aligned}$$

*Here, assumption 1 implies that $V_{ij}$ does not decrease when links are created, and that the benefit of adding links does not decrease as links are added.* ■

The above discussion also shows that individuals may have an incentive to coordinate their actions in the presence of positive network externalities. In the next section, I present a behavioural model of dynamic network formation in which individuals are allowed to coordinate.

### *2.1. A Behavioral Model*

In this section, I present a behavioural model where the network structure evolves through time, and where the individuals are allowed to coordinate their strategies. Since network formation games are complex in nature, such a behavioural model might be judged more realistic than assuming that individuals are fully rational (Wallace & Young, 2014). Here, the coordination of joint deviations is helped by a leader, whose role is to propose modifications to the network structure. As it will be clear, I impose no assumption on the identity of such leader. His identify can be predetermined, random or deterministic, and potentially be defined as a function of the network structure.

I also follow the extensive literature on network formation and focus on the pair-based approach.[11]

The network evolves through time as follows. Let $t = 1, ..., \infty$, and start from any network $G^0$. Every period $t$ is characterized by the following phases:

1. **(Agenda)** A leader wants to change the status of a subset of links in the network $G^t$. His agenda (i.e. the proposed modification) is choosen as follows:

---

[11] A prime example is pairwise stability (Jackson & Wolinsky, 1996). See Vannetelbosch & Mauleon (2015) for a recent review of equilibrium concepts in network formation games.

   1. (a) The leader randomly selects a group of pairs $J \subseteq \{ij\}_{j \neq i}$ to be updated.[12]
      (b) The leader randomly selects his agenda $\hat{G}_J = \{g_{ij}\}_{ij \in J}$.[13]
2. **(Campaigning)** The leader campaigns to implement $\hat{G}_J$.
   (a) He simultaneously, but privately, ask all $ij \in J$ such that $g_{ij}^{t-1} \neq \hat{g}_{ij} = 0$ whether they would be willing to remove their link, but without sharing his agenda with them. A pair $ij$ therefore accepts iff

   $$U_i(0, G_{-ij}^{t-1}) \geq U_i(1, G_{-ij}^{t-1}).$$

   If at least one pair refuses, the campaigning fails.

   (b) He jointly asks all pairs $ij \in J$ such that $g_{ij}^{t-1} \neq \hat{g}_{ij} = 1$ whether they would be willing to create a new link, sharing his agenda with them. A pair $ij$ therefore accepts iff

   $$U_i(1, \hat{G}_{J-ij}, G_{-J}^{t-1}) \geq U_i(0, \hat{G}_{J-ij}, G_{-J}^{t-1}).$$

   If at least one pair refuses, the campaigning fails.
3. **(Outcome)** If the campaigning succeeds (i.e. if does not fail), the agenda is implemented: $G^t = (\hat{G}_J, G_{-J}^{t-1})$. Otherwise, the network stays unchanged: $G^t = G^{t-1}$.

Here, the leader's campaigning strategy is crucial for the evolution dynamics as it helps to discipline number of possible deviations. However, note that the model is explicitly agnostic about the identity of the leader(s). This behavioural dynamics is therefore likely to represent reasonably well situations where changes in the network structure is affected by a leader, echoing the large literature on key-players in social networks (e.g. Zenou (2015)).

The leader's campaigning strategy can be rationalized as follows. Since the removal of a link does not require coordination, the first step of the campaigning

---

[12] The only assumption on the distribution $\lambda(\cdot|G^{t-1})$ is that $\lambda(J|G^{t-1}) > 0$ for all $J \neq \emptyset$.

[13] The only assumption on the distribution $\hat{\lambda}(\cdot|J, G_{t-1})$ is that $\hat{\lambda}(\hat{G}_J|J, G_{t-1}) > 0$ for all $\hat{G}_J$.

simply allows unprofitable links to be removed.[14] At this point, the leader's role is somewhat artificial and, indeed, could be replaced by simply allowing a subset of pairs in $J$ to sever their links.

The crucial role played by the leader, however, is to coordinate strategies for the creation of links. Since the addition of links is more profitable as links are added (under Assumption 1), in the second stage of the campaigning (i.e. step 2(b)), the leader discloses his agenda as a mean to stimulate cooperation.

This random network dynamic extends Mele (2016) link-based meeting process by allowing joint deviations.[15] In the next section, I study the properties of this dynamic process and show that it converges to a well-defined network structure as $t \to \infty$.

### *2.2. The Game*

In this section, I introduce a game which allows to study the properties of the behavioural model presented in section 2.1. If the game is interesting in itself, it can also simply be viewed as a technical tool that allows to derive the properties of the behavioural model. The game is defined as follows:

**Definition 1.** *The game of interest is* $\Gamma^P = \langle g_{ij}, U_{ij} \rangle_{ij:i\neq j}$*, where preferences are defined as:* $U_{ij}(g_{ij}, G_{-ij}) = g_{ij}[U_i([1, g_{i-j}], G_{-i}) - U_i([0, g_{i-j}], G_{-i})]$.

The game $\Gamma^P$ is an induced game where each link acts as a player, partially internalizing the externalities generated on the other links of the same individual.[16] Crucially, the payoffs in $\Gamma^P$ are standardized so that $U(0, G_{-ij}) = 0$ for all $ij$.

Also, remark that, rewritten in terms of the game's notation, Assumption 1 is equivalent to: $U_{ij}(g_{ij}, G_{-ij})$ being non-decreasing in $G_{-ij}$. This formally implies that $\Gamma^P$ is supermodular (Milgrom & Roberts, 1990; Topkis, 1979, 1998).

---

[14] This is similar to pairwise stability Jackson & Wolinsky (1996) where a network is pairwise stable only if all links are pairwise rational.

[15] Note however that here, the meeting process does not include temporary shocks.

[16] As discussed in section 2.1, if the pair-based approach is widely adopted in the literature, it comes with a cost. Specifically, for non-separable models, even a strong NE (Aumann, 1959) of $\Gamma^P$ may not be individually rational. However, if $U_i$ is separable across links, all NE of $\Gamma^P$ are individually rational.

Assumption 1 also implies that $\Gamma^P$ has positive spillovers (e.g. Milgrom & Roberts (1990, 1994)), since indirect links have a positive impact on payoffs. In any supermodular game, the set of Nash equilibria is non-empty, and there exists a greatest and a least NE.[17]

Importantly, Assumption 1 implies that the greatest NE of $\Gamma^P$ has robustness properties. Indeed, a standard argument for equilibrium refinements is that equilibria should be robust so as to allow individuals to communicate and coordinate while they choose their strategies. Aumann (1959) introduces the concept of a *strong Nash equilibrium*, which requires that the strategy profile be robust to deviations from any possible coalition. Bernheim et al. (1987) argues that this refinement is too demanding (in terms of existence and strategic sophistication) and proposes a weaker equilibrium refinement, called a *coalition-proof Nash equilibrium.* Coalition-proof equilibria are defined inductively and require that the strategy profiles be (Pareto) efficient, among the set of self-enforcing agreements.

Milgrom & Roberts (1994) showed that the greatest NE of a supermodular game with positive spillovers is coalition-proof and Newton (2012) showed that the greatest NE can be reached by a simple evolutionary process: *coalitional stochastic stability.* This particular evolutionary dynamic is such that 1) profitable unilateral deviations are more likely than coalitional profitable deviations and 2) coalitional profitable deviations from smaller coalitions are more likely.

I show below that under Assumption 1, the specific nature $\Gamma^P$ implies an equivalence between coalition-proof NE and strong NE. Perhaps more importantly, I show that the behavioural model presented in section 2.1 converges to the greatest NE of $\Gamma^P$. Formally:

**Proposition 1 (Main Result).** *Under Assumption 1, the greatest NE of* $\Gamma^P$ *is the unique Strong NE of* $\Gamma^P$*. Moreover, the behavioural model presented in section 2.1 converges to the greatest NE of* $\Gamma^P$ *with probability* $1$ *as* $t \to \infty$*.*

Proposition 1 shows that the behavioural model presented in section 2.1

---

[17] Formally, there exists $\bar{g}$ and $\underline{g}$ such that $\bar{g} \geq g^* \geq \underline{g}$ for all NE $g^*$, including $\bar{g}$ and $\underline{g}$, where $\geq$ holds element-wise (Topkis, 1979).

is well defined and that the NE reached has robustness properties. However, one drawback of the behavioural model is that it only reaches the greatest NE asymptotically, which makes simulations impracticable. In the next section, I present an alternative procedure to simulate the greatest NE of $\Gamma^P$.

## 3. Simulating the Greatest NE

I now discuss the simulation of the greatest NE of $\Gamma^P$. Topkis (1979) provides an algorithm that allows for the computation of the greatest NE of a game. The algorithm proceeds by best-response iterations, starting with the complete network.[18] In the context of $\Gamma^P$, the equilibrium converges to the greatest NE in at most $T = m(m-1)$ steps, where $m = n(n-1)$ is the total number of pairs. I show that the following (slightly modified algorithm) is faster.

**Algorithm 1 (Topkis (1979)).** *Define* $G^0$ *as the complete network. For every step* $t = 1, ..., T$*, and for all* $i, j$ *such that* $g_{ij}^{t-1} = 1$*, generate*

$$g_{ij}^t = \sup \arg\max_{g_{ij}} U_{ij}(g_{ij}, G_{-ij}^{t-1})$$

*until* $G^t = G^{t-1}$.

The algorithm is modified from the typical Topkis' algorithm since, at every step, one needs only to check for linked pairs. This reduces the convergence time, as shown in the following proposition.

**Proposition 2.** *Algorithm 1 converges to the greatest NE of* $\Gamma^P$ *in at most* $T = m(m+1)/2$ *steps.*

Proposition 2 implies that the greatest NE can be simulated in a finite number of steps. Also, since $T = m(m+1)/2$ is an upper-bound, convergence to the greatest NE may be much faster. This allows for the design of a simulation-based inference procedure. In the next section, I describe an approximate Bayesian computation (ABC) procedure that allows for the simulation of the posterior distribution of a set of preference parameters.

---

[18] Similar algorithms also exist for the simulation of the least NE; Echenique (2007) provides an algorithm that identifies all the NE.

## 4. Simulating the Posterior Distribution

In the remainder of this paper, I enrich the model by assuming the following random utility model (McFadden, 1980):

$$U_i(G) = U_i(G; X, \mathcal{E}, \theta) \tag{1}$$

where $X$ is a $n \times k$ matrix of observed characteristics, $\mathcal{E}$ is a vector of size $m \geq 1$ of unobserved characteristics with known distribution $\upsilon$, and $\theta$ is a vector of size $r \geq 1$, representing the parameter of interest.

At this point, it is worth noting that the procedure developed this section does not depend on the model presented in section 2 and can be used for alternative models, whenever the equilibrium network can be easily simulated (e.g. Miyauchi (2016) and Boucher (2017)).

That being said, I assume here that the data is generated by the greatest NE. Therefore, the likelihood function is given by:

$$P(G|X;\theta) = \int \mathbb{1}\{A(\mathcal{E}, X; \theta) = G\}\upsilon(d\mathcal{E}) \tag{2}$$

where $A(\mathcal{E}, X; \theta)$ is the network produced by Algorithm 1, given $\mathcal{E}$, $X$ and $\theta$. Since this likelihood has no closed-form expression, it has to be simulated.

As discussed in the introduction, my approach contrasts with the literature (e.g. Mele (2016)), where the likelihood *can* be written explicitly, but where the computation of the denominator is infeasible. Here, since (2) has no explicit form, it is not possible to adapt the approximate exchange algorithm of Mele (2016) to simulate the model. Fortunately, from Proposition 2, Algorithm 1 allows us to easily draw samples from the $P(G|X;\theta)$ since the algorithm terminates in finite time.

I use approximate Bayesian computation (ABC) to simulate the posterior distribution of $\theta$. ABC extends standard Bayesian inference for simulated likelihood models. Recall that standard Bayesian inference requires us to compute $p(\theta|X, G)$, the posterior distribution of $\theta$, given the data. From Bayes' rule, we have:

$$p(\theta|X,G) \propto p(\theta)P(G|X;\theta)$$

where $p(\theta)$ is the prior distribution on $\theta$ and $P(G|X;\theta)$ is the likelihood function. Since $p(\theta|X,G)$ usually cannot be written explicitly, most Bayesian analysis uses MCMC algorithms in order to draw samples from the posterior distribution $p(\theta|X,G)$. Specifically, the objective is to generate a random sequence $\{\theta^t\}_t$, such that $\theta^t$ is drawn from the posterior distribution as $t \to \infty$.[19]

However, these MCMC methods rely on researchers' abilities to evaluate the likelihood $P(G|X;\theta)$. ABC circumvents this requirement by using data augmentation. That is, MCMC methods for ABC generate a random sequence $\{H^t,\theta^t\}_t$, where the $H^t$ are pseudo-observations, i.e. networks simulated from the model, given $\theta^t$. In other words, the $H^t$ represent the greatest NE of a game where the individuals' utilities are given by (1), for different draws of $\mathcal{E}^t$.

If the sequence $\{H^t,\theta^t\}_t$ is such that $(H^t,\theta^t)$ is drawn from $p(\theta,H|X,G)$ as $t \to \infty$, the posterior distribution can be obtained by simply integrating over $H$.[20]

The literature on ABC is large and rapidly expanding. Marin et al. (2012) offer a review of the main concepts, as well as the most common MCMC algorithms used for ABC (see also Beaumont (2010)). Most MCMC methods for ABC (including the one I use here) are such that:

$$p(\theta,H|X,G) \propto p(\theta)P(H|X;\theta)\mathbb{1}\{\rho(S(H),S(G)) \leq \nu\}$$

where $\rho$ is a distance function, $S$ is a set of network summary statistics and $\nu$ is the tolerance level. The relevant posterior distribution is therefore:

$$p(\theta|X,G) \propto p(\theta)\sum_H P(H|X;\theta)\mathbb{1}\{\rho(S(H),S(G)) \leq \nu\} \qquad (3)$$

If the set of summary statistics is *Bayes sufficient*, (3) correctly approximates

---

[19] See for instance Greenberg (2012) for a good introduction to Bayesian inference.

[20] That is, $p(\theta|X,G) = \sum_H p(\theta,H|X,G)$.

the true posterior distribution as $\nu \to 0$.[21] The parameter $\nu$ therefore controls the precision of the approximation.[22]

A typical MCMC method for ABC goes therefore as follows:

**Algorithm 2 (MCMC).** *Start with $M$ parallel chains $\{H^0_m, \theta^0_m\}_m$, where $H^0_m$ is a pseudo-observation, simulated given $\theta^0_m$. For every step of the algorithm $t = 1, ..., \infty$, do the following:*

```
For $m = 1, ..., M$:
    1. Generate $\theta^*_m = \theta^{t-1}_m + jump(\{\theta^{t-1}_m\})$
    2. Simulate (potentially many) pseudo-observations $\{H^*_m\}$ given $\theta^*_m$
    3. Accept $\theta^*_m$ with probability $p_{acc}$
```

I use the Metropolis-Hasting acceptance probability:[23]

$$p_{acc} = \min\left(1, \frac{p(\theta^*_m)}{p(\theta^{t-1}_m)}\right) \cdot \mathbb{1}\{\rho(S(H^*_m), S(G)) \leq \nu\}$$

The (random) *jump* function gives the updating process of $\theta^t$. For example, one could simply choose a random-walk updating where $jump(\{\theta^{t-1}_m\}) \sim N(0, \Sigma)$. However, as the number of dimensions of $\theta$ grows, such an approach is likely to lead to bad acceptance rates and slow convergence.

Among the many possibilities, I use the jump function proposed by Sadegh & Vrugt (2014), formally described in the Appendix. This particular jump function has the advantage of using the information from the other parallel chains, while only updating a subset of dimensions of $\theta$ at each step.[24]

Finally, note that the algorithm is defined conditional on a given set of summary statistics; I discuss the selection of summary statistics in the next subsection.

---

[21] A set of summary statistics $S(G)$ is Bayes sufficient if for any prior $p(\theta)$, $p(\theta|S(G)) = p(\theta|G)$, a.s. See section 4.1 for a detailed discussion.

[22] Note that ABC can also be seen as an exact inference procedure, in the presence of model uncertainty (Wilkinson, 2013).

[23] See the Appendix for details about the acceptance probability during the burn-in phase.

[24] There are many variant of ABC-MCMC algorithms and the one I use here is by no mean the only possibility. The recent interest in ABC can also be the existence of off-the-shelf software packages. See for instance the "EasyABC" R package.

### *4.1. Selecting the Summary Statistics*

There is no consensus on how to determine optimal summary statistics in practice. As discussed by Robert et al. (2011), finding sufficient statistics is often impossible and most ABC is performed using insufficient statistics. Indeed, useful sufficient statistics can only be found for distributions of the exponential family.[25] Since most distributions of interest are not of the exponential family, most authors (e.g. Fearnhead & Prangle (2012) and Creel & Kristensen (2015)) focus on weaker conditions such as linear sufficiency.[26]

Fearnhead & Prangle (2012) show that best linearly sufficient summary statistic, in the sense of minimizing the expected quadratic error loss, is given by the true posterior mean. Since the mean of the posterior distribution is not known in practice, the authors propose approximating it using an auxiliary model.

Indeed, as noted by many authors (e.g. Forneron & Ng (2015)), the selection of summary statistics is analogous to the challenge of determining the binding function for (simulated) minimum distance estimators such as indirect inference (Gourieroux et al., 1993). I will use a similar approach and define my summary statistics using an auxiliary model.

Let $G$ be the observed network, and $H$ represent any network. Consider an auxiliary empirical model, represented by the set of moments $M(H;X,\beta)$.[27] I define my summary statistic as the moments of the auxiliary model:

$$S(H) = M(H;X,\hat{\beta})$$

for some $\hat{\beta}$. For example, $\hat{\beta}$ can be such that $M(G;X,\hat{\beta}) = 0$, i.e. the generalized method of moments (GMM) estimator for the auxiliary model.

---

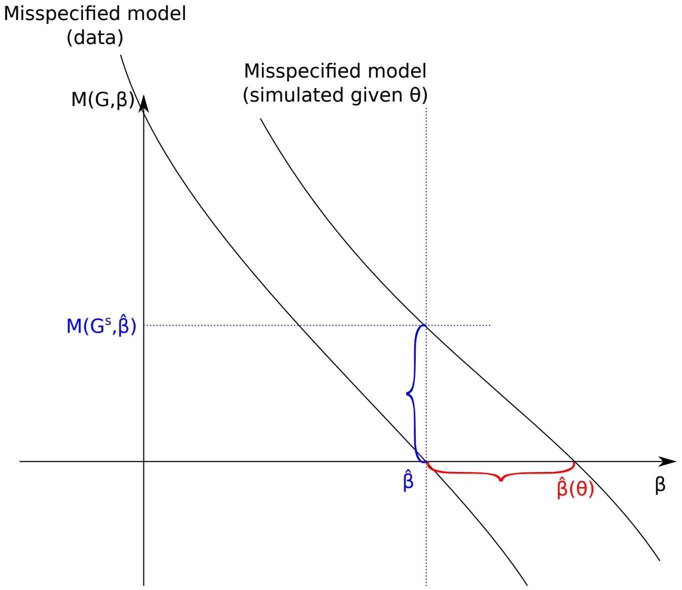
[25] This comes from the Pitman-Koopman-Darmois theorem, which states that, under some regularity conditions, only distributions of the exponential family have sufficient statistics that do not grow with the sample size. See Theorem 5.4 of Lehmann et al. (1991) for an exposition and discussion.

[26] i.e. $\mathbb{E}[\theta|S(G)] = \mathbb{E}[\theta|G]$ for any prior distribution.

[27] Note that identification requires that the dimension of the auxiliary parameter $\beta$ be at least as large as the number of dimensions of $\theta$.

This approach carries the same intuition as the indirect inference approach suggested by Gallant & Tauchen (1996), as displayed in Figure 2. In Gourieroux et al. (1993), one matches the coefficient estimated from the observed data ($\hat{\beta}$ in Figure 2) with the coefficient estimated from the simulated data ($\beta(\theta)$ in Figure 2). This is computationally intensive since it requires the estimated coefficient to be calculated for every simulation. Instead, I propose to follows Gallant & Tauchen (1996) and to match the value of the moments, evaluated at the coefficient estimated from the observed data ($M(G, \hat{\beta})$ in Figure 2).

Figure 2: Link to Indirect Inference



In the context of network formation models, a natural auxiliary model is an exponential random graph model (ERGM):

$$\mathcal{L}(H; X, \theta) = \phi(H; X, \theta) - \ln[\sum_{H'} \exp\{\phi(H'; X, \theta)\}]$$

for some function $\phi$.[28] Before computing the associated summary statistics (or moments, i.e. the derivative of $\mathcal{L}(H;X,\theta)$ with respect to $\theta$), note that $\mathcal{L}(H;X,\theta) \neq \mathcal{L}(H';X,\theta)$ iff $\phi(H;X,\theta) \neq \phi(H';X,\theta)$, so the normalizing constant brings no information. It is therefore sufficient to consider the following summary statistics:[29]

$$S(H) = \frac{\partial \phi(H;X,\cdot)}{\partial \theta}(\hat{\theta})$$

We must now choose a suitable function $\phi(H,X,\theta)$. If $\phi(H;X,\theta)$ is linear in $\theta$ (i.e. $\phi(H;X,\theta) = T(H,X)\cdot\theta$, where $T$ is a vector of the network's statistics), we have $S(H) = T(H,X)$ and the summary statistics $S$ are sufficient statistics for $\mathcal{L}(H;X,\theta)$, since in this case, the ERGM belongs to the exponential family. Unfortunately, this does not imply that $S(H)$ is sufficient for $P(H|X;\theta)$.

As discussed above, it is very hard to find conditions under which $S(H)$ is sufficient for $P(H|X;\theta)$, and the literature mostly uses insufficient statistics in practice.[30] Here, note that Proposition 1 implies that the greatest NE is Pareto-efficient. Therefore, any function $\phi(H;X,\theta)$, monotone with respect to the Pareto order, is a potentially good candidate (e.g. the utilitarian welfare function). I show in section 5, that this choice leads to fairly intuitive summary statistics. I finish this section by discussing the interpretation of ABC based on insufficient summary statistics.

### *4.2. An Interpretation of (Insufficient) Summary Statistics*

As discussed in the previous section, the error of approximation due to the use of insufficient summary statistics for ABC is still not very well understood. Nonetheless, since the requirement of sufficiency is particularly strong, most

---

[28] This is the log-likelihood for an ERGM.

[29] Note that here, $\beta$ has the same number of dimensions as $\theta$. I will therefore slightly abuse notation and let $\beta = \theta$. In general, computing $\hat{\theta}$ is challenging. If it is intuitive to think of $\hat{\theta}$ as being the maximum likelihood estimator of $\mathcal{L}(H;X,\theta)$, it is not needed for the inference procedure. Any carefully chosen value for $\theta$ could be used instead. Note also that if $\phi$ is linear in $\theta$, $S(H)$ will be independent of $\hat{\theta}$.

[30] Park et al. (2015) suggest using a non-parametric kernel density function in order to circumvent the problem of choosing a set of sufficient statistics. The approach is promising for real-valued outcome variables, but it is impracticable for networks.

ABC implementations focus on insufficient statistics, which raises questions about their validity.

In this section, I argue that the even with the use of insufficient summary statistics, ABC still has a meaningful interpretation. Specifically, I provide an alternative interpretation of ABC, based on the notion of *quasi-posterior* distribution, as introduced by Chernozhukov & Hong (2003).

As discussed in the previous section, the summary statistics in a Bayesian setting play the same role as moments in a classical setting. Consider the following function, which is the objective function for a standard (simulated) method of moments:

$$L(\theta) = -(E(S(H)|\theta) - S(G))'\Omega(E[S(H)|\theta] - S(G))$$

where $E(S(H)|\theta)$ is the average value of the summary statistics for a large number of simulations given $\theta$, and where $\Omega$ is positive definite. Then, following Chernozhukov & Hong (2003) this defines the *quasi-posterior*:

$$\tilde{p}(\theta|X, G) \propto \exp\{L(\theta)\}p(\theta)$$

As discussed in Chernozhukov & Hong (2003), it is often easier to implement MCMC algorithms to simulate $\tilde{p}(\theta|X, G)$ than to maximize $L(\theta)$. In this vision somewhat operational, Bayesian statistics constitutes a convenient alternative to classical estimation of non-convex problems, where identification is challenging, or when the asymptotic properties are not well understood.

The same approach is true here. By construction, $S(\cdot)$ is Bayes sufficient for the quasi-posterior $\tilde{p}(\theta|X, G)$. This implies that the ABC presented in the previous sections converges to the quasi-posterior as $\nu \to 0$. Then, from an operational point of view, ABC procedure can be interpreted as a MCMC procedure for resolution of a classical simulated GMM estimator.

This also allows to see that the difference between ABC and standard Bayesian statistics is similar to the difference between Maximum Likelihood (ML) and

GMM estimations. When the likelihood function has many local maxima, the ML allows to discriminate between them using the value of the likelihood. When using a GMM (based on the derivative of the likelihood) however, identification is lost since all local maxima solve the GMM objective function. In this case, identification can often be restored using a GMM based on higher moments (see for instance the discussion in Gallant & Tauchen (1996)).

Therefore, when summary statistics are based on natural moments, the loss of using insufficient statistics can be seen as a loss of identification, similar to the one due to using GMM instead of ML. I present an illustration in the next section.

## 5. High-School Friendship Networks

In this section, I provide an illustration of the model using friendship formation among teenagers. I use the Add Health database, which contains data on friendships as well as many socio-economic characteristics of high-school teenagers. There are dummy variables indicating whether an individual is female, Hispanic, white, black or Asian, as well as whether their mother or father are currently in the work force. There is also data on mothers' and fathers' years of high school and post-secondary education, and the student's current grade. Table 1 presents descriptive statistics.

I consider the following model:

$$U_i(G; X, \mathcal{E}, \theta) = \sum_{j \neq i} g_{ij} V_{ij}(G_{-ij}, X, \mathcal{E}, \theta)$$

where

$$\begin{aligned} V_{ij}(G_{-ij}, X; \theta) &= \underbrace{\theta_0 + \sum_{l=1}^{10} \theta_l |x_i^l - x_j^l| + \varepsilon_{ij}}_{\text{Private}} + \underbrace{\theta_{11} g_{ji}}_{\text{Reciprocity}} \\ &\quad \underbrace{+ \theta_{12} \sum_{k \neq i,j} g_{ik} \max(g_{jk}, g_{kj})}_{\text{Cliques}} \\ &= \hat{V}_{ij}(G_{-ij}, X; \theta) + \varepsilon_{ij} \end{aligned} \tag{4}$$

Table 1: Descriptive Statistics

| **Variable** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|
| Female | 0.5205 | 0.4999 | 0 | 1 |
| Hispanic | 0.0627 | 0.2425 | 0 | 1 |
| White | 0.8581 | 0.3492 | 0 | 1 |
| Black | 0.1061 | 0.3082 | 0 | 1 |
| Asian | 0.0141 | 0.1178 | 0 | 1 |
| Grade | 8.9552 | 1.6858 | 7 | 12 |
| Education (Mother) | 5.6611 | 2.2787 | 1 | 12 |
| Work (Mother) | 0.7762 | 0.4170 | 0 | 1 |
| Education (Father) | 5.8657 | 2.5208 | 1 | 12 |
| Work (Father) | 0.9450 | 0.2281 | 0 | 1 |

Number of individuals: 782
Number of schools: 14

assuming that $\theta_{11}, \theta_{12} \geq 0$.

We can easily compute:

**Proposition 3.**

$$U_{ij}(G; X, \mathcal{E}, \theta) = g_{ij} V_{ij}(G_{-ij}, X, \mathcal{E}, \theta) + g_{ij}\theta_{12} \sum_{k \neq i,j} g_{ik} \max(g_{jk}, g_{kj})$$

*so assumption 1 holds whenever* $\theta_{11}, \theta_{12} \geq 0$.

Here, $\theta_0$ represent the intrinsic value of friendship, while $\theta_1, ..., \theta_{10}$ are meant to capture homophilic preferences. Homophily is a well-known feature of social networks (e.g. Currarini et al. (2009)) and describes the empirical fact that similar individuals have a higher probability of being linked. In particular, if $\theta_1, ..., \theta_{10}$ are negative, it implies that linking with individuals with different socio-economic characteristics lowers an individual's payoff. The parameter $\theta_{11} \geq 0$ represents the additional benefit of reciprocal links, and $\theta_{12} \geq 0$ represents the benefit of creating cliques.

I assume that the random shocks $\varepsilon_{ij}$ follow a Gumbel(0,1) distribution.[31] This distribution is asymmetric and allows for a better representation of rare events. As shown in Table 2, the number of links is much lower than the number

---

[31] This is also called a log-Weibull or generalized extreme value and referred to in the literature on discrete choice models as the complementary log-log model.

of potential links (i.e. the number of pairs): only about 2.2% of potential friendships are created. Therefore, the use of a distribution where large shocks are less frequent than small ones is more appropriate.

I now discuss the choice of summary statistics. As discussed in the previous section, a good candidate for the auxiliary model is an ERGM model where $\phi(H;X,\theta)$ is increasing with respect to the Pareto order. Here, a sensible choice is the (expected) utilitarian welfare function:[32]

$$\phi(H;X,\theta) \quad = \quad \sum_{ij} h_{ij}\hat{V}_{ij}(H_{-ij},X,\theta)$$

This leads to the following associated summary statistics:

$$S(H) = M(H,X,\hat{\beta}) = \sum_{ij} h_{ij} Z_{ij}(H_{-ij}) \tag{5}$$

where $Z_{ij}(H_{-ij}) = (X'_{ij},\ 1,\ h_{ji},\ \sum_{k\neq i,j} h_{ik}\max\{h_{jk},h_{kj}\})'$ has the same dimension as $\theta$. Note that since $\phi(H;X,\theta)$ is linear in $\theta$, the summary statistics are independent of $\theta$.

Also note that the intuition behind the construction of the summary statistics is easily visible from (5). The ABC procedure tries to match $r$ moments, one for each dimension of the parameter space, as would be the case for a (simulated) GMM. Table 2 gives the value of such moments in the data.

As a benchmark, I first present the results from a (inconsistent) complementary log-log model, i.e. where $P(g_{ij}=1) = 1-\exp(-\exp(U_{ij}))$. Results are displayed in Table 3. Results show evidence of homophilic preferences on all exogenous characteristics.

I now estimate the true model by simulating from the posterior distribution using Algorithm 2. Results are displayed in Table 4, and details of the implementation as well as diagnostics and convergence tests can be found in

---

[32] Note that the maximum of the utilitarian welfare function is not, in general, a NE. Also, the resulting summary statistics are the same as one would obtain by using $\phi(H;X,\theta) = \sum_{ij} U_{ij}(H_{-ij},X,\theta)$.

Table 2: Network Summary Statistics

| **Summary Statistic (Moments)** | **Value** |
|---|---|
| Number of links × Δ gender | 463 |
| Number of links × Δ Hispanic | 69 |
| Number of links × Δ white | 65 |
| Number of links × Δ black | 25 |
| Number of links × Δ Asian | 26 |
| Number of links × Δ age | 400 |
| Number of links × Δ mother educ. | 2,535 |
| Number of links × Δ mother work | 359 |
| Number of links × Δ father educ. | 2,521 |
| Number of links × Δ father work | 95 |
| Number of links: | 1,175 |
| Number of reciprocal links | 420 |
| Number of cliques | 1,040 |
| Number of pairs | 53,290 |

Table 3: Misspecified Complementary Log-Log Estimates

| **Variable** | Mean | Std |
|---|---|---|
| Gender | -0.3670** | (0.0610) |
| Hispanic | -0.2645** | (0.1311) |
| White | -0.0458 | (0.1584) |
| Black | -1.0359** | (0.2420) |
| Asian | -0.1279 | (0.2000) |
| Grade | -1.1062** | (0.0496) |
| Education (Mother) | 0.0067 | (0.0169) |
| Work (Mother) | -0.1086 | (0.0646) |
| Education (Father) | -0.0418** | (0.0154) |
| Work (Father) | -0.0562 | (0.1112) |
| Reciprocal links | 1.8318** | (0.0699) |
| Cliques | 0.5912** | (0.0206) |
| Intercept | -2.8996** | (0.0740) |

the Appendix. In particular, Figures 3 and 4 of the Appendix show how well the model does in terms of replicating the observed summary statistics, similar to what is proposed in Lemaire et al. (2016). Observed summary statistics are well matched. The posterior distribution for each parameter can also be found in Figures 5 and 6 of the Appendix.

Note that the posterior distribution for the effect of reciprocal links and cliques are small and highly concentrated. This is explained by the equilibrium effects of positive network externalities, and by the explicit consideration of the equilibrium selection mechanism. A small increase in the value of cliques or reciprocal links leads to a large increase in the number of realized cliques at equilibrium.

I conclude in the next section.

## 6. Going Further

The model developed in this paper is quite general and has a wide range of potential applications. In particular, the simulated nature of the estimation procedure allows for much more flexibility with respect to the unobservables. As an illustration, in section 5, I use Gumbel-distributed random errors. However, the model allows for much more general error structures.

As discussed in Pin & Rogers (2015), stochastic models of network formation are particularly good at replicating observed network statistics, more so than strategic models of network formation (as in this paper). However, stochastic models often lack economic intuition. The ABC procedure developed in this paper reconciles the stochastic and strategic approaches. Indeed, the strategic model of network formation could be enriched by a stochastic model (e.g. through the distribution of the unobservables). Since the set of summary statistics used by the ABC procedure is not limited to the ones generated by the strategic model, one can use additional network statistics (e.g. clustering coefficient, average degree, diameter, etc.) in order to recover the parameters of the stochastic network formation model.

Table 4: Posterior Distribution

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|---|---|---|---|---|---|---|
| Sex | 1,950,000 | −0.855 | 0.951 | −19.175 | −0.786 | 1.580 |
| Hisp | 1,950,000 | −0.405 | 0.561 | −11.264 | −0.340 | 0.762 |
| White | 1,950,000 | −0.444 | 0.771 | −16.249 | −0.334 | 1.166 |
| Black | 1,950,000 | −1.432 | 0.980 | −18.840 | −1.385 | 1.102 |
| Asian | 1,950,000 | −0.575 | 0.459 | −5.283 | −0.519 | 0.722 |
| Grade | 1,950,000 | −2.167 | 1.368 | −21.771 | −1.966 | −1.197 |
| M_Educ | 1,950,000 | −1.416 | 3.961 | −38.694 | −0.136 | 0.849 |
| M_Work | 1,950,000 | −0.480 | 0.727 | −16.415 | −0.418 | 1.399 |
| F_Educ | 1,950,000 | −3.457 | 5.516 | −37.948 | −0.556 | 0.514 |
| F_Work | 1,950,000 | −0.075 | 0.576 | −5.076 | −0.050 | 1.499 |
| Intercept | 1,950,000 | −0.131 | 0.763 | −3.594 | 0.007 | 1.510 |
| Recip_Links (logs) | 1,950,000 | −8.449 | 5.707 | −39.991 | −7.230 | −0.082 |
| Cliques (logs) | 1,950,000 | −10.138 | 5.612 | −43.280 | −8.942 | −2.049 |

Simulation-based inference procedures are therefore very promising and could go a long way toward explaining the formation of social and economic networks.

## 7. References

Aumann, R. J. (1959). Acceptable points in general cooperative n-person games. *Contributions to the Theory of Games*, *4*, 287–324.

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

Bala, V., & Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, *68*(5), 1181–1229.

Beaumont, M. A. (2010). Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics*, *41*(379-406), 1.

Bernheim, B. D., Peleg, B., & Whinston, M. D. (1987). Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, *42*(1), 1–12.

Boucher, V. (2015). Structural homophily. *International Economic Review*, *56*(1), 235–264.

Boucher, V. (2017). Selecting equilibria using best-response dynamics. *Working Paper*.

Boucher, V., & Mourifié, I. (2012,2016). My friend far far away. a random field approach to exponential random graph models. *Working Paper*.

Bramoullé, Y., & Fortin, B. (2010). social networks: econometrics. *The New Palgrave Dictionary of Economics., Basingstoke: Palgrave Macmillan*, *3*.

Chandrasekhar, A. (2015). Econometrics of network formation. In *The Oxford Handbook of the Economics of Networks*, (pp. 303–357).

Chandrasekhar, A. G., & Jackson, M. O. (2014). Tractable and consistent random graph models. Tech. rep., National Bureau of Economic Research.

Chandrasekhar, A. G., & Jackson, M. O. (2015). A network formation model based on subgraphs. *Available at SSRN*.

Chernozhukov, V., & Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, *115*(2), 293–346.

Creel, M., & Kristensen, D. (2015). On selection of statistics for approximate bayesian computing (or the method of simulated moments). *Computational Statistics & Data Analysis*.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, *77*(4), 1003–1045.

De Paula, A. (2015). Econometrics of network models. *Unpublished manuscript, University College London*.

Echenique, F. (2007). Finding all equilibria in games of strategic complements. *Journal of Economic Theory*, *135*(1), 514–532.

Fabrikant, A., Luthra, A., Maneva, E., Papadimitriou, C. H., & Shenker, S. (2003). On a network creation game. In *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, (pp. 347–351). ACM.

Fearnhead, P., & Prangle, D. (2012). Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(3), 419–474.

Forneron, J.-J., & Ng, S. (2015). The abc of simulation estimation with auxiliary statistics. *arXiv preprint arXiv:1501.01265*.

Galeotti, A., Goyal, S., & Kamphorst, J. (2006). Network formation with heterogeneous players. *Games and Economic Behavior*, *54*(2), 353–372.

Gallant, A. R., & Tauchen, G. (1996). Which moments to match? *Econometric Theory*, *12*, 657–681.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, (pp. 457–472).

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, *8*, S85–S85.

Graham, B. S. (2016a). An econometric model of network formation with degree heterogeneity. *Forthcoming in Econometrica*.

Graham, B. S. (2016b). Homophily and transitivity in dynamic network formation. Tech. rep., National Bureau of Economic Research.

Greenberg, E. (2012). *Introduction to Bayesian econometrics*. Cambridge University Press.

Harrison, R., & Muñoz, R. (2008). Stability and equilibrium selection in a link formation game. *Economic Theory*, *37*(2), 335–345.

Hellmann, T. (2013). On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, *42*(1), 211–237.

Jackson, M. O. (2008). *Social and economic networks*, vol. 3. Princeton university press Princeton.

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of economic theory*, *71*(1), 44–74.

Lehmann, E. L., Casella, G., & Casella, G. (1991). *Theory of point estimation*. Wadsworth & Brooks/Cole Advanced Books & Software.

Lemaire, L., Jay, F., Lee, I., Csilléry, K., & Blum, M. G. (2016). Goodness-of-fit statistics for approximate bayesian computation. *arXiv preprint arXiv:1601.04096*.

Leung, M. (2014). A random-field approach to inference in large models of network formation. *Available at SSRN*.

Leung, M. (2015a). A weak law for moments of pairwise-stable networks. *Available at SSRN 2663685*.

Leung, M. P. (2015b). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics*, *188*(1), 182–195.

Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate bayesian computational methods. *Statistics and Computing*, *22*(6), 1167–1180.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, (pp. S13–S29).

Mele, A. (2016). A structural model of segregation in social networks. Tech. rep.

Mele, A., & Zhu, L. (2016). Variational approximations for large strategic ergms. Tech. rep.

Milgrom, P., & Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, (pp. 1255–1277).

Milgrom, P., & Roberts, J. (1994). Strongly coalition-proof equilibria in games with strategic complementarities. Tech. rep., Stanford University, Department of Economics.

Miyauchi, Y. (2016). Structural estimation of pairwise stable networks with nonnegative externality. *Journal of Econometrics*, *195*(2), 224–235.

Newton, J. (2012). Coalitional stochastic stability. *Games and Economic Behavior*, *75*(2), 842–854.

Park, M., Jitkrittum, W., & Sejdinovic, D. (2015). K2-abc: Approximate bayesian computation with kernel embeddings. *stat*, *1050*, 24.

Pin, P., & Rogers, B. (2015). Stochastic network formation and homophily. In *The Oxford Handbook of the Economics of Networks*.

Robert, C. P., Cornuet, J.-M., Marin, J.-M., & Pillai, N. S. (2011). Lack of confidence in approximate bayesian computation model choice. *Proceedings of the National Academy of Sciences*, *108*(37), 15112–15117.

Sadegh, M., & Vrugt, J. A. (2014). Approximate bayesian computation using markov chain monte carlo simulation: Dream (abc). *Water Resources Research*, *50*(8), 6767–6787.

Sawa, R. (2014). Coalitional stochastic stability in games, networks and markets. *Games and Economic Behavior*, *88*, 90–111.

Sheng, S. (2012). Identification and estimation of network formation games. *Unpublished manuscript*.

Sheng, S. (2014). A structural econometric analysis of network formation games. Tech. rep.

Topkis, D. M. (1979). Equilibrium points in nonzero-sum n-person submodular games. *SIAM Journal on Control and Optimization*, *17*(6), 773–787.

Topkis, D. M. (1998). *Supermodularity and complementarity*. Princeton university press.

Vannetelbosch, V., & Mauleon, A. (2015). Network formation games. In *The Oxford Handbook of the Economics of Networks*.

Wallace, C., & Young, H. P. (2014). Stochastic evolutionary game dynamics. *Handbook of Game Theory*, *4*.

Wilkinson, R. D. (2013). Approximate bayesian computation (abc) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, *12*(2), 129–141.

Xu, X., & Lee, L.-f. (2015). Estimation of a binary choice game model with network links. *Working Paper*.

Zenou, Y. (2015). Key players. In *The Oxford Handbook of the Economics of Networks*.

## 8. Appendix - Proofs

*Proof of Proposition 1*

**Claim 1: The greatest NE of $\Gamma^P$ is the unique strong NE.**

Let $G^*$ be the greatest NE and suppose that there exists a coalition $S$ with profitable deviation $\hat{G}_S$. This implies that for all $ij \in S$,

$$U_{ij}(\hat{G}_S, G^*_{-S}) \geq U_{ij}(G^*) \tag{6}$$

with strict inequality for at least one $ij$.

Let $G_s^+ = \max\{\hat{G}_S, G_S^*\}$ and $G_s^- = \min\{\hat{G}_S, G_S^*\}$.

Since $G^*$ is a NE, we have $U_{ij}(g_{ij}^-, G^*_{S-ij}, G^*_{-S}) \leq U_{ij}(G^*)$, which implies (using (6)) that $U_{ij}(\hat{g}_{ij}, \hat{G}_{S-ij}, G^*_{-S}) \geq U_{ij}(g_{ij}^-, G^*_{S-ij}, G^*_{-S})$, or:

$$U_{ij}(\hat{g}_{ij}, \hat{G}_{S-ij}, G^*_{-S}) - U_{ij}(g_{ij}^-, G^*_{S-ij}, G^*_{-S}) + U_{ij}(g_{ij}^*, G^*_{S-ij}, G^*_{-S}) \geq U_{ij}(G^*)$$

From assumption 1, this implies:

$$U_{ij}(\hat{g}_{ij}, \hat{G}_{S-ij}, G^*_{-S}) - U_{ij}(g_{ij}^-, G^+_{S-ij}, G^*_{-S}) + U_{ij}(g_{ij}^*, G^+_{S-ij}, G^*_{-S}) \geq U_{ij}(G^*)$$

And from assumption 1 again, it also implies that:

$$U_{ij}(\hat{g}_{ij}, G^+_{S-ij}, G^*_{-S}) - U_{ij}(g_{ij}^-, G^+_{S-ij}, G^*_{-S}) + U_{ij}(g_{ij}^*, G^+_{S-ij}, G^*_{-S}) \geq U_{ij}(G^*) \tag{7}$$

There are two cases: either $g_{ij}^* = 0$, which implies that $g_{ij}^- = 0$, and $g_{ij}^+ = \hat{g}_{ij} = 1$, or $g_{ij}^* = 1$, which implies that $g_{ij}^+ = 1$, and $g_{ij}^- = \hat{g}_{ij} = 0$. In any case, we have:

$$U_{ij}(g_{ij}^+, G^+_{S-ij}, G^*_{-S}) = U_{ij}(\hat{g}_{ij}, G^+_{S-ij}, G^*_{-S}) + [U_{ij}(g_{ij}^*, G^+_{S-ij}, G^*_{-S}) - U_{ij}(g_{ij}^-, G^+_{S-ij}, G^*_{-S})] \tag{8}$$

Therefore, equations (7) and (8) imply that $G_S^+$ is also a profitable deviation for $S$. Note that the deviation is strictly profitable since it must be strict for at least one $ij$.

Note that for all $ij \notin S$, we have $U_{ij}(G^*) \leq U_{ij}(G^+_S, G^*_{-S})$ from assumption 1, so $(G^+_S, G^*_{-S})$ is a Pareto improvement from $G^*$. If $(G^+_S, G_{-S})$ is a NE, this contradicts $G^*$ being the greatest NE.

Suppose that $G^{**} \equiv (G^+_S, G^*_{-S})$ is not a NE. Then, there exists $ij$ and $g_{ij}$ such that $U_{ij}(g_{ij}, G^{**}_{-ij}) > U_{ij}(G^{**})$.

Suppose first that $g^{**}_{ij} = 1$. If $ij \notin S$, it imlies that $g^*_{ij} = 1$ and it is a contradiction since links were added and assumption 1 holds. If $ij \in S$, then $U_{ij}(1, G^{**}_{-ij}) \geq U_{ij}(0, G^*_{-ij}) = 0$ (the coalition was a profitable deviation), but $U_{ij}(1, G^{**}_{-ij}) < U_{ij}(0, G^{**}_{-ij}) = 0$. This is a contradiction of assumption 1.

This shows that $g^{**}_{ij} = 0$ is the only possibility. Then, profitable deviations go by link addition under assumption 1. The process is then repeated until no profitable deviation exists. The final result is necessarily a NE, which contradicts the fact that $G^*$ is the greatest NE.

The uniqueness comes from the fact that the NE are ranked according to the Pareto order under Assumption 1 and that at least one $ij$ would strictly prefer the greatest NE. QED

**Claim 2: The behavioural model converges to the greatest NE.**

Note that the behavioural model can be written as follows:

1. A group of pairs $J \subseteq \{ij\}_{j \neq i}$ is selected according to $\lambda(\cdot|G^{t-1})$ such that $\lambda(J|G^{t-1}) > 0$ for all $J \neq \emptyset$.
2. An updating $\hat{G}_J = \{g_{ij}\}_{ij \in J}$ is proposed according to $\hat{\lambda}(\cdot|J, G_{t-1})$ such that $\hat{\lambda}(\hat{G}_J|J, G_{t-1}) > 0$ for all $\hat{G}_J$.
3. $\hat{G}_J$ is accepted iff $U_{ij}(\hat{G}_J, G^{t-1}_{-J}) \geq U_{ij}(G^{t-1})$ for all $ij \in J$.

This dynamics corresponds to the unperturbed dynamics in Sawa (2014). One can see that the meeting process allows for any deviation, for any coalition, with strictly positive probability. Then, any network that is not a strong NE cannot be absorbing. Similarly, any strong NE is an absorbing state. Since the greatest NE is the only strong NE, this shows that the greatest NE is the only absorbing state.

To see that the greatest NE is always reachable from any network, consider the following procedure. Start with $G$, which is not a NE. Consider the set of links such that $g_{ij} = 1$ and $U_{ij}(0, G_{-ij}) > U_{ij}(1, G_{-ij})$. Remove one of those links, and repeat until no link removal is profitable. At this point any profitable deviation must go through link addition, until it reaches a NE. Once a NE is reached, go to the greatest NE. Then, by construction, the greatest NE was obtained by a series of profitable coalitional deviations. QED

*Proof of Proposition 2*

By construction, the procedure goes by link removal, therefore $G^t \geq G^{t+1}$. If, at some step $t$, $U_{ij}(0, G^t_{-ij}) > U_{ij}(1, G^t_{-ij})$, then $U_{ij}(0, G^{t+1}_{-ij}) > U_{ij}(1, G^{t+1}_{-ij})$ from increasing differences (or positive spillovers). Therefore, there is no need to revisit links such that $g^t_{ij} = 0$.

Convergence to the greatest NE follows from Topkis (1979). The convergence time is bounded by the slowest possible convergence, i.e. if only one link is removed at every stage and the greatest NE is the empty network.

*Proof of Proposition 3*

We have:

$$U_i(1, G_{-ij}) - U_i(0, G_{-ij}) = V_{ij}(G_{-ij}) + \theta_{12} \sum_{k \neq i,j} g_{ik} \max\{g_{jk}, g_{kj}\}$$

so

$$U_{ij} = g_{ij} V_{ij}(G_{-ij}) + \theta_{12} g_{ij} \sum_{k \neq i,j} g_{ik} \max\{g_{jk}, g_{kj}\}$$

or,

$$U_{ij} = g_{ij}[\theta_0 + \sum_{l=1}^{10} \theta_l |x^l_i - x^l_j| + \varepsilon_{ij} + \theta_{11} g_{ji} + 2\theta_{12} \sum_{k \neq i,j} g_{ik} \max(g_{jk}, g_{kj})]$$

We immediately see that assumption 1 holds for $\theta_{11}, \theta_{12} \geq 0$.

QED

## 9. Appendix - MCMC Algorithm

### 9.1. Distance Function and Tolerence Level

I scale the value of the different moments so as to make them comparable using the following distance:

$$\rho(S(G), S(H)) = \sqrt{[S(G) - S(H)]' W [S(G) - S(H)]}$$

where $W$ is a diagonal matrix with diagonal elements $W_{ii} = 1/[S_i(G)]^2$. The tolerance level is $\nu = 0.15$.

When the MCMC algorithm generates multiple pseudo-observations $H$, I use the average: $\rho(S(G), \{S(H_k)\}_{k \in K}) = \frac{1}{|K|} \sum_k \rho(S(G), S(H_k))$.

### 9.2. The "Jump" Function

The following description closely follows Sadegh & Vrugt (2014). The only substantial difference is to the random sampling of the dimensions of $\theta$ (see step 1 below). The jump function is such that:

1. Only a subset of the dimensions of $\theta$ are updated. Each dimension is updated with a 20% probability.
2. Conditional on being selected for updating, the "jump" on the dimension $r'$ of $\theta$ uses the information from three other chains to generate the jump and is equal to:

$$(1 + er_1)\sigma \left[ \sum_{m \in \hat{M}} \theta_m^{r',t-1} - \sum_{m \in \tilde{M}} \theta_m^{r',t-1} \right] + er_2$$

   where $er_1 \sim U(-0.1, 0.1)$, $er_2 \sim N(0, 10^{-12})$, $\sigma = 2.38/\sqrt{6d}$) ($d$ is the total number of dimensions of $\theta$ that are updated at this step of the algorithm), $\theta_m^{r',t-1}$ is the dimension $r'$ of $\theta_m^{t-1}$, and $\hat{M}$ and $\tilde{M}$ are random samples of three chains (excluding the chain being updated, and without replacement).

### *9.3. Acceptance Probability During Burn-In*

In order to help the convergence to the stationary distribution during the burn-in phase, Sadegh & Vrugt (2014) propose using the following acceptance probability: $p_{acc} = \min\left(1, \frac{p(\theta_m^*)}{p(\theta_m^{t-1})}\right)$ if $\rho(S(H_m^*), S(G)) \leq \nu$ or if $\rho(S(H_m^*), S(G)) \leq \rho(S(H_m^{t-1}), S(G))$, thus accepting draws with smaller distances even if the distance is larger than the tolerance level.

Also during the burn-in phase, I remove extreme chains using the following procedure. Looking at the distribution of distance at every step, I look at the chain with the largest distance. If this largest distance is greater than two standard deviations above the mean, I replace the chain with another randomly selected chain. I repeat until more than 90% of the chains have surpassed the tolerance level. Chains below the tolerance level are then are randomly assigned values from chains that passed the tolerance level.

All chains reached the threshold after 1315 steps of the MCMC algorithm. (The actual burn-in is considered to be the 69,999 first steps; see below.)

### *9.4. Prior Distribution*

The prior distributions are flat, i.e. all parameters are normally distributed, with a mean equal to the complementary log-log estimates (see Table 3) and a variance of 100.

### *9.5. Convergence Tests*

Following Sadegh & Vrugt (2014), I consider that the MCMC algorithm reaches the stationary distribution when the Gelman and Rubin's convergence test (Gelman & Rubin, 1992) is less than 1.2 for all parameters, dropping the first half of the MCMC steps. The output of the test is displayed in Table 5. The final number of steps is 21,000. For any statistic of the stationary distribution, I drop the first 69,999 steps, so all draws from steps 70,000 to 120,000 are assumed to be from the stationary distribution.

Table 5: Gelman and Rubin's Convergence Test (for second half of the 120,000 steps

| **Variable** | **Value** |
|---|---|
| Gender | 1.20 |
| Hispanic | 1.03 |
| White | 1.04 |
| Black | 1.09 |
| Asian | 1.00 |
| Grade | 1.10 |
| Education (Mother) | 1.09 |
| Work (Mother) | 1.04 |
| Education (Father) | 1.07 |
| Work (Father) | 1.04 |
| Intercept | 1.08 |
| Reciprocal links (logs) | 1.00 |
| Cliques (logs) | 1.00 |

### *9.6. Computing Time*

The estimation was performed on a Thinkstation D40, with two Intel Xeon E5 2609 @ 2.40GHz (total of 8 cores), with 40GB of RAM in about 4 days.

### *9.7. Summary Statistics*

The distribution of the summary statistics at the stationary distribution (the posterior distribution) is displayed in Figures 3 and 4. Observed values are displayed in black.

### *9.8. Posterior Distribution*

The posterior distributions for all parameters are displayed in Figures 5 and 6.

Figure 3: Distribution of the Summary Statistics at the Posterior Distribution

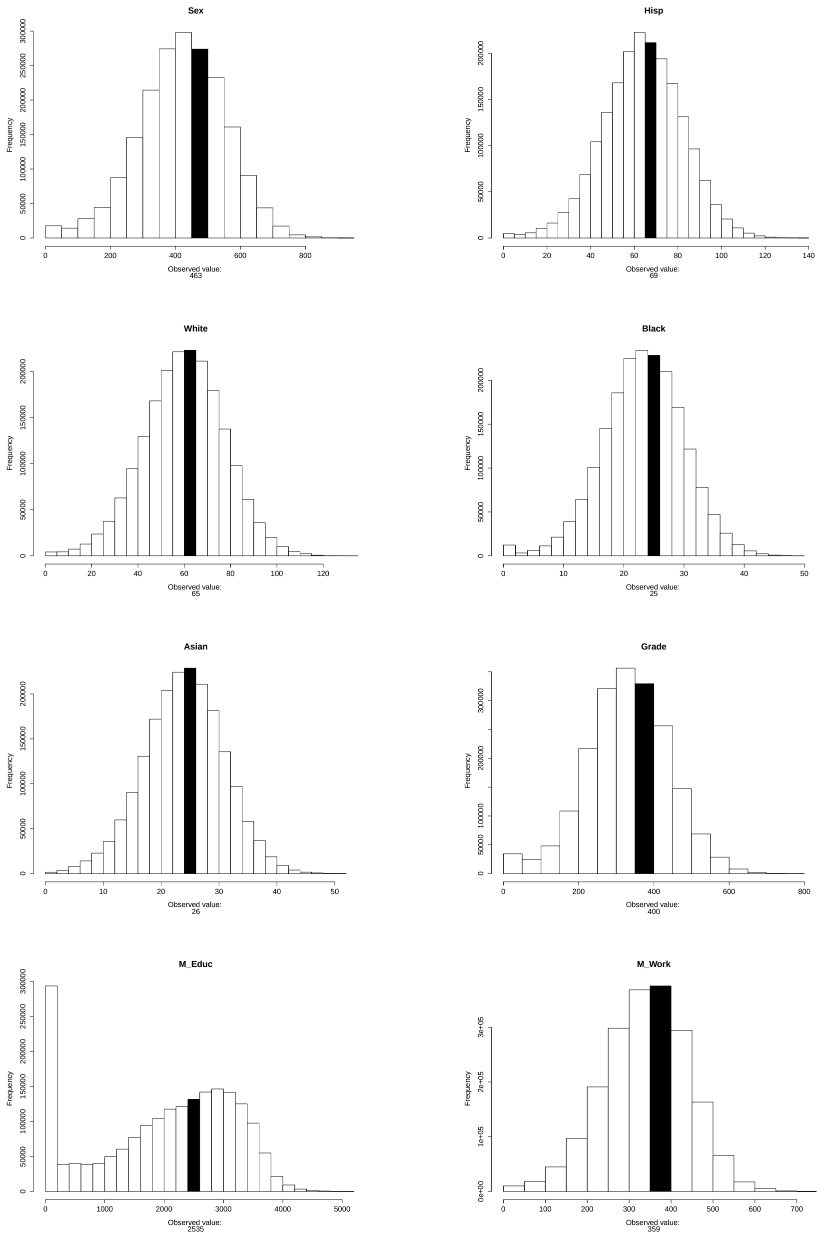


Figure 4: Distribution of the Summary Statistics at the Posterior Distribution (continued)

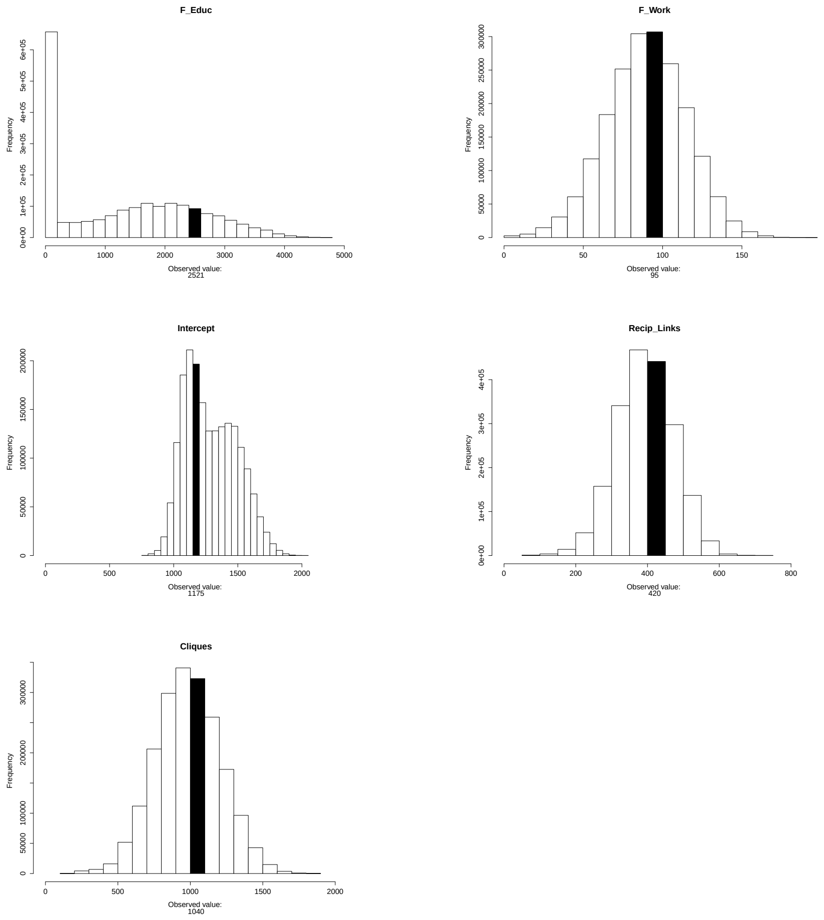


Figure 5: Posterior Distributions

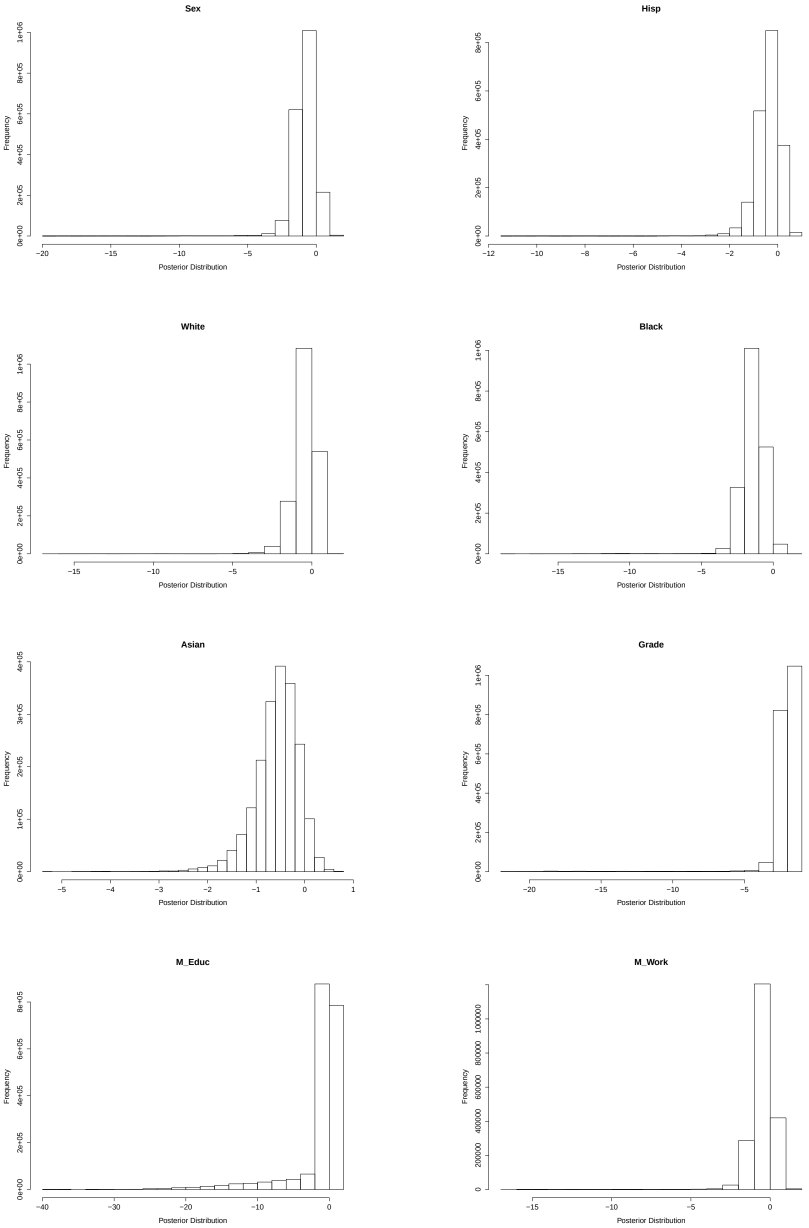


Figure 6: Posterior Distributions (continued)