University of Toronto
Department of Economics

Working Paper 499

My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks

By Vincent BOUCHER and Ismael MOURIFIÃ›

October 01, 2013

My Friend Far Far Away: Asymptotic Properties of Pairwise Stable Networks

Vincent Boucher* and Ismael Mourifié**

Previous version: November 2012

This version: September 2013

Abstract

We explore the asymptotic properties of pairwise stables networks (Jackson and Wolinsky, 1996). Specifically, we want to recover a set of parameters from the individuals' utility functions using the observation of a *single* pairwise stable network. We develop Pseudo Maximum Likelihood estimator and show that it is consistent and asymptotically normally distributed under a very weak version of homophily. The approach is compelling as it provides explicit, easy-to-check conditions on the admissible set of preferences. Moreover, the method is easily implementable using pre-programmed estimators available in most statistical packages. We provide an application of our method using the Add Health database.

JEL Codes: C13,D85

Corresponding author: Vincent Boucher, vincent.boucher@ecn.ulaval.ca

This paper benefited from many highly interesting discussion. We would especially want to thank Tim Conley, Sílvia Gonçalves, Bryan S. Graham, Paul Grieco, Ilze Kalnina, Joris Pinkse, Benoit Perron and Yao Luo for their comments and suggestions. A special thanks also to Yann Bramoullé and Marc Henry for their helpful comments and suggestions. We would also like to thank the participants to many conferences and seminars, including: the SCSE2012, the economic department at PennState, the economic department at U. de Montreal, and the workshop CeMMAP-CIREQ-SciencesPo-X on "The Estimation of Complementarities in Matching and Social Networks".

This research uses data from Add Health, a program project 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.

* Economic Department, Université Laval. Vincent Boucher thanks the economic faculty and the Cambridge-INET at the University of Cambridge for their support and for many interesting discussions.

** Economic Department, University of Toronto. A significant part of this work has been done while Ismael Mourifié was visiting the economic department at PennState. Many thanks for their support and welcome.

1 Introduction

How do social networks form? Specifically, how can we measure the influence of an individual's socioeconomic characteristics on the identity of his peers? We know that many social networks exhibit strong racial or religious segregation (see for instance Echenique and Fryer 2007, Watts 2006, and Mele 2007). This observation raises many interesting questions regarding the cause of this segregation. For instance, we would like to be able to distinguish the impact of the individuals' characteristics (e.g. race), and the impact of the individuals' positions in the networks (e.g. popularity). The shape of the existing social networks also have measurable effects on individuals' choices. Many studies show a strong influence of an individual's peers on his actions, ranging from unhealthy consumption choices (e.g. Fortin and Yazbeck 2011 and the references therein) to labor force participation (e.g. van der Leij et al. 2009, and Patacchini and Zenou 2009). However, since most social networks are endogenously formed, the estimated influence of peers is likely to be biased.[1] Understanding how the networks are formed could then allow us to control for this endogeneity and suggest policy instruments that would help influence network formation processes.

In this paper, we provide a simple Pseudo Maximum Likelihood estimator (PMLE, see Gourieroux et al., 1984) which allows us to recover underlying preference parameters for pairwise stable networks (Jackson and Wolinsky, 1996). The approach is compelling as it requires the observation of a single network. The set of admissible preferences is also large and characterized by intuitive, easy-to-check conditions. Specifically, we show that the estimator is consistent and asymptotically normally distributed provided that individuals' preferences exhibit a weak version of *homophily*. Homophily is one of the most robust empirical fact about social networks. It formalizes the observation that similar individuals are more likely to interact with each other. As homophily is featured by both theoretical (e.g. Bramoullé et al. 2012, and Currarini et al. 2009), and empirical (e.g. Mele 2007, and Christakis et al. 2010) models of network formation, our methodology is applicable to many existing models of network formation. We apply this new methodology to the formation of friendship networks among American teenagers.

A fundamental challenge in estimating a network formation process is the highly dependent nature of most socio-economic relationships. Consider for instance the case of friendship networks.

[1] The literature on peer effects have only recently considered explicitly the endogeneity of social networks. See for instance Goldsmith-Pinkham and Imbens (2011), and Blume et al. (2011).

The probability that Adam and Beth are friends depends on their individual characteristics. However, it may also depends on the fact that Beth is friend with Charlotte (who maybe does not like Adam). The probability that Adam and Beth are friends may then depend on Charlotte's individual characteristics. Hence, the observation "Adam and Beth are friends" depends on Charlotte's characteristics. However, if individuals have homophilic preferences, the probability that Adam and Beth are friends should be primarily influenced by individuals similar to them. If Adam and Beth are high-school teenagers for instance, the probability that they become friends increases if they go the the same school, or if they attend the same classes. Accordingly, if Beth and Charlotte are friends, there is a greater probability that they go to the same school, or at least that they live in the same country. Then, Donald, a elderly man, living in a different country (hence having individual characteristics quite different from those of Adam, Beth and Charlotte) probably does not influence much the probability that Adam and Beth become friends. We generalize this argument and show that homophily implies a generalization of the ϕ-mixing property used in time-series and spatial econometric models. This fact allows us to define a consistent estimation strategy based on a Pseudo Maximum Likelihood estimator.

This paper contributes to the empirical literature on strategic network formation. Two main approaches have been proposed. The first approach requires the observation of many (mostly independent) social networks. Some of those papers are specifically interested in estimating homophilic preferences (see for instance Boucher 2012, and Currarini et al. 2010) and uses standard frequentist approaches, i.e. standard Maximum Likelihood estimators. As these papers assume ex-ante homophily, they are limited in their scope of applications. Looking at a more general set of preferences, Sheng (2012) studies the presence of multiple equilibria when the link formation process is locally dependent. One limit of those approaches is however that they require the observation of many (an asymptotically infinite number of) independent social networks, which is not always available in existing databases.

The second approach requires the observation of only one network, at one point in time. As the observations are highly dependent, standard maximum likelihood methods are not consistent. Accordingly, most papers use a Bayesian approach, and as the likelihood function cannot usually be written explicitly, most papers rely on simulation methods such as Markov Chain Monte Carlo (in particular Christakis et al. 2010, Mele 2010, and Goldsmith-Pinkham and Imbens 2011). If they are less demanding in terms of data, those methods are however quite complex to implement in practice (and not very flexible), and the computing time needed makes them unsuitable for large database.

We contribute to this literature by providing an explicit, easily implementable, PMLE requiring the observation of only one social network, at one point in time. We introduce a weakened notion of homophily, and show that it implies that our PMLE is consistent and asymptotically normally distributed. In order to do so, we use Laws of Large Numbers and Central Limit Theorems due to Jenish and Prucha (2009), as well as estimators for the variance-covariance matrices due to Conley (1999) and Bester et al. (2012).

We also contribute to the literature by discussing the identification of models of network formation based on pairwise stability and by making the link with the bivariate probit with partial observability (Poirier, 1980). We also discuss how our model of network formation can be adapted to study bipartite networks, and games with transfers.

The remaining of the paper is organized as follows. In section 2.1, we present the economy. In section 2.2, we propose an estimator of the equilibrium social network which allows to recover the underlying individuals' preferences. In section 3, we derive the asymptotic distribution of our estimator, and in section 4, we define a class of network formation models suited to our econometric framework. In section 5, we provide an application using the Add Health database, and we discuss policy-making implications and potential avenues for future research in section 6.

2 The basic framework

2.1 The Economy

Let $N = \{1, ..., n\}$ be the set of individuals. Each individual is characterized by a random vector of $T \geq 1$ characteristics $x_i = (x_i^1, ..., x_i^T) \in \mathcal{X}$. We assume that $\mathcal{X} \subset \mathbb{R}^T$ and we define the distance between two individuals as $d(i,j) = d(x_i, x_j)$, where d is a distance on $\mathbb{R}^T$. Finally, we note $x = (x_1, ..., x_n) \in \mathcal{X}^n$ the matrix of individual characteristics. Is it worth noting that the choice of the distance function d is arbitrary. In general, the choice of this distance function will be context-dependent. In particular, the distance can represent spatial preferences of the individuals.[2] We provide an example in section 5.

Let $m = \frac{n(n-1)}{2}$ be the number of possible pairs of individuals (i, j) for $i \neq j$ in the economy.

[2] See in particular Henry and Mourifié (2011) for spatial preferences on the euclidean space.

We assume that individuals interact in a network $g_m = (N, \mathbf{W})$, where $\mathbf{W}$ is a $n \times n$ symmetric matrix that takes values $w_{ij} = 1$ if $i \in N$ and $j \in N$ are linked by a socio-economic relationship (e.g. friendship), and $w_{ij} = 0$ otherwise. For a given set of individuals N, the set of all possible networks is noted $\mathbb{G}_m$. For a given network $g_m \in \mathbb{G}_m$, we will note $ij \in g_m$ if $w_{ij} = 1$. We will also denote by $g - ij$, the network g_m from which we removed the link between i and j. If $ij \notin g_m$, then $g_m - ij = g_m$. We define $g_m + ij$ similarly.

The set of links an individual has is noted $N_i(g_m) = \{j \in N : ij \in g_m\}$. The cardinality of that set is the *degree* of an individual, formally $n_i(g_m) = |N_i(g_m)|$. The *geodesic distance* (or shortest path) between i and j in the network g_m equals the minimal number of existing links in g_m such that j can be reached from i. Let $\rho_{ij}(g_m)$ be the geodesic distance between i and j in the network g_m. We say that i and j are *connected* in g_m if $\rho_{ij}(g_m) < \infty$. If i and j are not connected, we let $\rho_{ij}(g_m) = \infty$. Let $R_{ij}^{g_m} = \{k \in N | \min(\rho_{ik}(g_m), \rho_{jk}(g_m)) < \infty\}$ be the set of individuals connected either to i or to j. For $V \subset N$, we note $g_{m|V}$ the network restricted to individuals in V, i.e. for all $i, j \in V$, we have $(w_{|V})_{ij} = w_{ij}$, while we have $(w_{|V})_{ij} = 0$ if $i \in N \setminus V$ or $j \in N \setminus V$. Let also $x_V \in \mathcal{X}^{|V|}$ be the matrix of individual characteristics of individuals in V.

We assume that the network $g_m = (N, \mathbf{W})$ is endogenous and determined as a function of the individuals' (stochastic) utilities. An individual has preferences over the set of characteristics and the network structure in the economy, i.e. $u_i : \mathbb{G}_m \times \mathcal{X}^n \to \mathbb{R}$. Specifically, we write $u_i(g_m, x; \theta, \varepsilon_i)$ where $\theta \in (\theta^1, ..., \theta^K) \in \Theta$ is the set of parameters to be estimated, and the vector $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{in})$ is the unobserved component of the utility function. It will be convenient to use the following representation of the utility function.

Definition 1 *Given g_m and x, the value for $i \in N$ of a link with $j \in N \setminus \{i\}$ is given by*

$$H_i^j(g_m - ij, x; \theta, \varepsilon_i) = u_i(g_m, x; \theta, \varepsilon_i) - u_i(g_m - ij, x; \theta, \varepsilon_i)$$

Given $H_i^j(g_m - ij, x; \theta, \varepsilon_i)$ for all $i, j \in N$, we want to know what information can be retrieved from the observation of a single network $g_m \in \mathbb{G}_m$, and a set of individual characteristics $x \in \mathcal{X}^n$. We concentrate on the properties of the network g_m and not on the specific dynamic process by which the network is created. For instance, we do not require the links to be added in a specific order to the network. We rather assume that the observed network g_m is *stable*. We are interested in a particular notion of stability, introduced by Jackson and Wolinsky (1996).

Definition 2 *A network* g_m *is* ***Pairwise Stable*** *if, for all* $i, j \in N$*:*

1) $w_{ij} = 1$ *if* [$H_i^j(g_m - ij, x; \theta, \varepsilon_i) \geq 0$ *and* $H_j^i(g_m - ij, x; \theta, \varepsilon_j) \geq 0$]

2) $w_{ij} = 0$ *otherwise*

Then, a link is created iff it is profitable for both individuals involved. Let $PSN \subseteq \mathbb{G}_m$ be the set of pairwise stable networks. The existence and multiplicity of equilibria are discussed in section 4.3. For now, assume that there exists a unique pairwise stable network. Pairwise stability is extensively used in the literature on strategic network formation.[3] Any potential deviation from a pairwise stable network results from a *single* pair of individuals changing the status of *its* link. That is, any admissible deviation is such that $g_m \in \mathbb{G}_m$ goes from g_m to $g_m + ij$ for some $i, j \in N$, or from g_m to $g_m - ij$ for some $i, j \in N$. Pairwise stability can then be viewed as the weakest bilateral extension from the set of individually rational networks.[4] We study the asymptotic properties of pairwise stable networks. In the next section, we present the econometric framework.

2.2 The Econometric Framework

We want to know what information can be retrieved from the observation of a single pairwise stable network. Specifically, suppose that we observe a set of m pairs of individuals. The set of pairs is noted S_m, with typical elements s and r. Any two individuals i and j necessarily belong to some pair s, where $s = (s_1, s_2) = (i, j)$. For each pair, we observe the *status* (linked or not) of the pair and the socio-economic characteristics of the individuals in the pair (age, gender, income...). We formally define the position of a pair $s \in S_m$ in $\mathcal{X}$ as the average point between s_1 and s_2, i.e. $x_s \in \mathcal{X}$ such that $x_s = \frac{x_{s_1} + x_{s_2}}{2}$.[5] Accordingly, the distance between two pairs r and s is equal to $d(s, r) = d(x_r, x_s) = d(\frac{s_1+s_2}{2}, \frac{r_1+r_2}{2})$.

In this section, we show that pairwise stability allows to express the probability of a link's status in terms of the observable socio-economic characteristics. Let $\Sigma_\varepsilon = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}$, we present our first assumption.

[3] See for instance Jackson (2008, chapter 6).

[4] For comparisons between stability concepts on networks, see for instance Bloch and Jackson (2006) and Chakrabarti and Gilles (2007).

[5] This is done without loss of generality. The method is robust to other definitions of a pair's position in $\mathcal{X}$, as long as x_s is located in a given neighbourhood of x_{s_1} and x_{s_2}.

Assumption 1 (Preferences) *For all* $i, j \in N$,

(1.1) $H_i^j(g_m - ij, x; \theta, \varepsilon_i) = h_i^j(g_m - ij, x; \theta) + \varepsilon_{ij}$, *with* $(\varepsilon_{ij}, \varepsilon_{ji})|g_m - ij, x \sim N(0, \Sigma_\varepsilon)$.

(1.2) $h_i^j(g_m - ij, x; \theta)$ *is three times continuously differentiable in* θ.

(1.3) Θ *is a compact subset of* $\mathbb{R}^K$, *for* $K \geq 1$.

Assumption 1.2 and 1.3 are standard technical requirements. Assumption 1.1 deserves more attention. The separability of the error term is quite standard (see Additive Random Utility Models, following McFadden, 1981). Also, as our dependent variable (i.e. the status of a pair) is discrete, only scale-identification can be achieved. There is then no loss of generality in normalizing the diagonal elements of the variance-covariance matrix of the error term. We assume that ε_{ij} follows a normal distribution for convenience (for instance, it allows to present our estimator as a bivariate probit estimator, see below). In general, our method can be adapted to many distributional assumptions. In particular, all our results are valid for any distribution for which the left tail of the cdf distribution is exponentially bounded. Notice that while the $(\varepsilon_{ij}, \varepsilon_{ji})$ are identically distributed, they are not necessarily independent.

The error term ε_{ij} is interpreted as a random shock on the value of the pair for an individual. Hence, the observed binary outcome w_{ij} does not represent the binary choice of a single decision-maker, but ratter the joint choice of both individuals in the pair (i.e. $w_{ij} = 1$ only if the link has positive value for i *and* j). Then, our model leads to a bivariate probit estimator, where we partially observe the choices of the individuals (Poirier, 1980).[6]

Specifically, we want to estimate $\theta \in \Theta$, given the fact that the observed network g_m is pairwise stable. Given definition 2, a link ij is created (i.e. $w_{ij} = 1$) if and only if $H_i^j(g_m - ij, x; \theta, \varepsilon_i) \geq 0$ and $H_j^i(g_m - ij, x; \theta, \varepsilon_j) \geq 0$. Then, under assumption 1.1, the probability that $w_{ij} = 1$ for $i, j \in N$ is equal to $\Phi_2(h_i^j(g_m - ij, x; \theta), h_j^i(g_m - ij, x; \theta), \varrho)$, where Φ_2 is the c.d.f. for the standardized bivariate normal distribution with covariance ϱ. We then propose the following PMLE.[7]

$$\begin{aligned}\mathcal{L}_m(\theta) &= \frac{1}{m}\sum_{ij:i<j} w_{ij} \ln[\Phi_2(h_i^j(g_m - ij, x; \theta), h_j^i(g_m - ij, x; \theta), \varrho)] \\ &+ (1 - w_{ij}) \ln[1 - \Phi_2(h_i^j(g_m - ij, x; \theta), h_j^i(g_m - ij, x; \theta), \varrho)] \qquad (1)\end{aligned}$$

[6] A previous version of the paper was assuming $\varepsilon_{ij} = \varepsilon_{ji}$ which led to an univariate probit model. We thank Bryan S. Graham for having suggested this extension to us.

[7] Since the observations are dependent, the true likelihood of g_m cannot be written as the product of the marginals $\Phi_2(h_i^j(g_m - ij, x; \theta), h_j^i(g_m - ij, x; \theta), \varrho)$. See Gourieroux et al. (1984) or Gourieroux and Monfort (1989, section 8.4) for a general description.

Then our estimator is a "standard" bivariate probit with partial observability.[8] It is well known that the identification of those type of models is tricky and we discuss it in section 5. Another problem is that the estimator $\hat{\theta} = \arg\max_{\theta\in\Theta} \mathcal{L}_m(\theta)$ is not necessarily consistent (as $m \to \infty$) since the observations can be dependent. For instance, $h_i^j(g_m - ij, x; \theta)$ may depends on the number of links i and j have in the network g_m. In the next section, we find sufficient conditions for the consistency and asymptotic normality of $\hat{\theta} = argmax_{\theta\in\Theta}\mathcal{L}_m(\theta)$ when the number of pairs m goes to infinity.

3 Limited Dependence Theorems

In this section, we present two theorems for dependent observations. We show that under ϕ-mixing, $\theta \in \Theta$ can be consistently estimated using the model in (1). Those theorems are useful since, as we show in section 4, there exist simple conditions on h_i^j which imply ϕ-mixing.[9]

We start by introducing the following random variable, for all pairs $s \in S_m$:

$$Z_{s,m} = \begin{cases} 1 \text{ if } H_{s_1}^{s_2}(g_m - ij, x; \theta, \varepsilon_{s_1}) \geq 0 \text{ and } H_{s_2}^{s_1}(g_m - ij, x; \theta, \varepsilon_{s_2}) \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

The random field $\{Z_{s,m}; s \in S_m, m \in \mathbb{N}\}$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \{0,1\}^m$, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on Ω. To clarify the exposition, we use the following simplifying notation:

$$\begin{aligned} q_{s,m}(z_{s,m}|x, g_m, \theta) &= w_s \ln[\Phi_2(h_{s_1}^{s_2}(g_m - ij, x; \theta), h_{s_2}^{s_1}(g_m - ij, x; \theta), \varrho)] \\ &+ (1 - w_s) \ln[1 - \Phi(h_{s_1}^{s_2}(g_m - ij, x; \theta), h_{s_2}^{s_1}(g_m - ij, x; \theta), \varrho)] \end{aligned}$$

so (1) can be written as:

$$\mathcal{L}_m(\theta) = \frac{1}{m} \sum_{s\in S_m} q_{s,m}(z_{s,m}|x, g_m, \theta) \tag{2}$$

We also use $q_{s,m}(\theta) = q_{s,m}(z_{s,m}|x, g_m, \theta)$ when there is no ambiguity.

[8] This estimator is available in many statistical software packages. For instance, in Stata, one can use the command "biprobit" with the "partial" option.

[9] Our results can easily be adapted to other mixing definitions such as α-mixing.

We now turn to the dependence structure of (2). For any two events $A \in \mathcal{A}$ and $B \in \mathcal{B}$, where $\mathcal{A}, \mathcal{B}$ are sub-σ-algebras of $\mathcal{F}$, the ϕ-mixing coefficient is given by

$$\phi(\mathcal{A}, \mathcal{B}) = \sup\{|\mathbb{P}(A|B) - \mathbb{P}(A)|, A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(B) > 0\}$$

This is analogue to standard time-series models. In a time dependent model, the estimation is consistent if $\lim_{r\to\infty} \sup_t \phi(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+r}^{\infty}) = 0$, where $\mathcal{F}_{t_1}^{t_2}$ is the σ-algebra for the realizations from time t_1 to time t_2.[10] We want to apply the same basic approach when the dependence between A and B goes through $\mathcal{X}$. Then, instead of characterizing an observation by its position in time, we define it by its position in $\mathcal{X}$. Since the dependence in $\mathcal{X}$ is more complex than time-dependence, the asymptotic convergence of the ϕ-mixing coefficient is not sufficient. In order to show the consistency and asymptotic normality of $\hat{\theta} = \arg\max_\theta \mathcal{L}_m(\theta)$, we use Laws of Large Numbers and Central Limit theorems for dependent observations on random fields developed by Jenish and Prucha (2009, Theorems 1,2 and 3). Let us introduce the following definition.

Definition 3 *Let $A, B \subset \Omega$, with corresponding σ-algebra $\mathcal{A}_m$ and $\mathcal{B}_m$. Let also $|A|$ and $|B|$ denote the number of pairs of individuals in A and B. We define the following function:*

$$\bar{\phi}_{k,l}(r) = \sup_m \sup_{A,B} (\phi(\mathcal{A}_m, \mathcal{B}_m), |A| \leq k, |B| \leq l, d(A,B) \geq r)$$

where $d(A,B)$ is the Hausdorff distance on $\mathcal{X}$ for the set of pairs in A and B.

We will show that a sufficient condition for the consistency and the asymptotic normality of $\hat{\theta} = argmax_\theta \mathcal{L}_m(\theta)$ is the following:

Assumption 2 (ϕ-mixing)
(2.1) $\sum_{r=1}^{\infty} r^{T-1} \bar{\phi}_{1,1}^{1/2}(r) < \infty$
(2.2) $\sum_{r=1}^{\infty} r^{T-1} \bar{\phi}_{k,l}(r) < \infty$, *for $k + l \leq 4$*
(2.3) $\bar{\phi}_{1,\infty}(r) = O(r^{-T-\epsilon})$ *for some $\epsilon > 0$.*

Recall that $T \geq 1$ is the dimension of $\mathcal{X}$. In words, not only $\bar{\phi}_{k,l}(r)$ has to converge to 0, but this convergence has to be *fast enough*. In section 4, we give sufficient conditions under which assumption 2 holds. For the moment, we show the validity of the estimation technique given

[10] See for instance White (2001).

that ϕ-mixing is respected. The first theorem concerns the consistency of $\hat{\theta} = arg\max_{\theta\in\Theta}\mathcal{L}_m(\theta)$. First, we need some regularity conditions.

Assumption 3 (Regularity I)

(3.1) There exists a unique $\theta_0 \in int\ \Theta$ *maximizing* $\lim_{m\to\infty}\mathbb{E}[\mathcal{L}_m(\theta)]$.

(3.2) For all $s_1, s_2 \in N$, $d(s_1, s_2) \geq d_0$ *for some* $d_0 > 0$.

(3.3) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} |q_{s,m}(\theta)|^{(1+\eta)}] < \infty$ *for some* $\eta > 0$.

(3.4) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} |\frac{\partial q_{m,s}(\theta)}{\partial\theta}|] < \infty$.

Assumption 3.1 is the identification condition. Assumption 3.2 is the *increasing domain* assumption. It ensures that the distance goes to infinity as the number of individuals goes to infinity. Given the existence of a minimal distance d_0, the sub-space of $\mathcal{X}$ which contains all the individuals has to expand (with respect to d) as the number of individuals increases. This assumption describes how the space of individual characteristics $\mathcal{X}$ is filled as the number of pairs m goes to infinity. Finally, assumption 3.3 and 3.4 require standard moment conditions on the payoff function. We have the following.

Theorem 3.1 (Consistency) *Suppose that assumptions 1 and 3 hold, and that assumption (2.2) is respected for* $k = l = 1$. *Then, the estimator* $\hat{\theta} = arg\max_{\theta\in\Theta}\mathcal{L}_m(\theta)$ *is consistent as* $m \to \infty$.

We still need to derive the asymptotic distribution of $\hat{\theta}$. We define the following matrices:

$$\begin{aligned} D_0(\theta_0) &= lim_{m\to\infty}\mathbb{E}[\frac{\partial^2\mathcal{L}_m(\theta_0)}{\partial\theta\partial\theta'}] \\ B_0(\theta_0) &= lim_{m\to\infty} m\mathbb{E}[\frac{\partial\mathcal{L}_m(\theta_0)}{\partial\theta}\left(\frac{\partial\mathcal{L}_m(\theta_0)}{\partial\theta}\right)'] \end{aligned}$$

Now, since the asymptotic normality of the estimator requires more structure than the one needed for consistency, we need assumptions 2.1-2.3, as well as the following additional regularity conditions.

Assumption 4 (Regularity II)

(4.1) $B_0(\theta_0) > 0$.

(4.2) $D_0(\theta_0)$ *is invertible.*

(4.3) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} \|D_{m,s}(\theta)\|^{1+\eta}] < \infty$ *for some* $\eta > 0$.

(4.4) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} \|\frac{\partial D_{m,s}(\theta)}{\partial\theta}\|] < \infty$.
(4.5) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} |\frac{\partial q_{s,m}(\theta)}{\partial\theta}|^2] < \infty$

where $D_{m,s}(\theta) = \frac{\partial^2 q_{s,m}(\theta)}{\partial\theta\partial\theta'}$. Those assumptions are quite standard and are sufficient to show the asymptotic normality of our estimator.[11]

Theorem 3.2 (Asymptotic Normality) *Let $m \to \infty$. Under assumptions 1, 2, 3 and 4, the estimator $\hat{\theta} = arg\max_{\theta\in\Theta} \mathcal{L}_m(\theta)$ is normally distributed with variance-covariance matrix given by $D_0^{-1}B_0D_0^{-1}/m$.*

The Variance-Covariance Matrix is the equivalent for our setting of the Heteroskedasticity and Autocorrelation Consistent (HAC) variance-covariance matrix. The estimation of those variances is not straightforward. The estimation of $D_0(\theta_0)$ follows from theorems 3.1 and 3.2 since $D_0(\theta)$ has the same dependence structure as $\lim_{m\to\infty} \mathbb{E}\mathcal{L}_m(\theta)$. A consistent estimator is then $D_m(\hat{\theta}) = \frac{1}{m}\sum_{s=1}^m D_{s,m}(\hat{\theta})$. Defining a consistent estimator for $B_0(\theta_0)$ is more challenging. We suggest two approaches to estimate $B_0(\theta_0)$. The first one is based on a generalization of standard HAC estimators and is due to Conley (1999). The estimator $B_m(\theta)$ is formally described in the appendix. Under mixing conditions, $B_m(\theta)$ is a consistent estimator for $B_0(\theta_0)$. Although valid, this estimator can be very computationally intensive when the number of dimensions of $\mathcal{X}$ increases (say, $T \geq 4$). An alternative approach have been suggest by Bester et al. (2012), where they propose to use the well known Variance Cluster (VC) estimator (also formally described in appendix). Although the estimator is not consistent under weak dependence, they show that the estimator converges to a well defined random variable and that the standard t-test are still valid. In other words, under mixing conditions, inference using the VC estimator is valid, even if the estimator itself is not consistent. This estimator has the advantage of requiring little computational time and to be simple to implement.

In this section, we have shown that under ϕ-mixing and some regularity conditions, $\theta \in \Theta$ can be recovered using (1). In the next section, we show that an asymptotic version of the homophily principle is a sufficient condition for ϕ-mixing, as defined in assumption 2.

[11]Formally, the proof of theorem 3.2 derives the limit distribution for $\sqrt{m}(\hat{\theta} - \theta_0)$. We report the asymptotic distribution of $\hat{\theta}$ for presentation purposes.

4 Models of network formation

4.1 A First Example

We now turn to economic models of network formation. We want to find sufficient conditions on $h_i^j(g_m - ij, x; \theta)$ such that assumption 2 holds. To clarify the presentation, we start with a simple example. Assume for the moment that

$$h_i^j = h_i^j[N_i(g_m - ij), N_j(g_m - ij), d(i,j)]. \tag{3}$$

That is, the value of a link depends only on the (direct) links the individuals have, and the distance between them. Given this specific dependence structure, we will show that a weak version of the homophily principle is sufficient to achieve ϕ-mixing.

Homophily is a prominent feature of social networks. It characterizes the empirical fact that similar individuals have a higher probability of being linked.[12] We assume the following:

Assumption 5 (Asymptotic Homophily) *For all* $i, j \in N$,
(5.1) $h_i^j(g_m - ij, x; \theta) \to -\infty$ *as* $d(i,j) \to \infty$.
(5.2) $lim_{d(i,j)\to\infty} \exp\left\{-\frac{h_i^j(g_m - ij, x; \theta)^2}{2d(i,j)}\right\} \in [0,1)$.

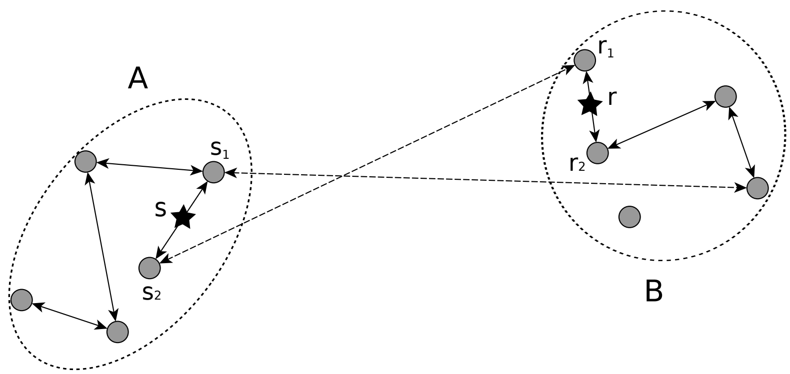
Assumption (5.1) simply says that if the distance between two individuals is infinite, the probability that they form a link is equal to 0. Condition (5.2) limits the asymptotic concavity of h_i^j in d. For example, suppose that $\bar{h}(d) = O(d^\eta)$ for some η. Then, assumption 5.2 holds if $\eta > \frac{1}{2}$, but not if $\eta \leq \frac{1}{2}$. Notice that assumption 5 only requires that homophily holds asymptotically hence allowing for a wide range of non-homophilic preferences.

We show that, under the specification in (3), Asymptotic Homophily is sufficient for ϕ-mixing. Before we present the formal result, we provide a graphical intuition. Consider Figure 1, where we assumed that $\mathcal{X} = \mathbb{R}^2$. Individuals are represented as circles, and pairs as stars.

The ϕ-mixing condition says that, as the distance between A and B tends to infinity, the realizations on A and B (i.e. the status of the pairs within those subsets) are independent. Consider pairs s and r. As the distance between r and s increases, the distance between the

[12]Many definitions of homophily exist in the economic literature, see for instance Currarini et al. (2009) and Bramoullé and Rogers (2010). In particular, some papers explicitly define homophily using a distance function on the space of individual characteristics: for instance, Boucher (2012), Johnson and Gilles (2000), Marmaros and Sacerdote (2006), and Iijima and Kamada (2010).

Figure 1: ϕ-mixing on Networks



individuals within those pairs (i.e. s_1, s_2 and r_1, r_2) increases as well. Under assumption 5, as the distance between, s_2 and r_1 goes to infinity, the probability that they form a link goes to zero. Since, under the specification in (3), payoffs only depends on direct links, the status of s will therefore be independent of the status of r. The argument holds for any pairs in A and B.

Before presenting the formal statement, we need to add one more regularity assumption. Recall that a necessary condition for theorems 3.1 and 3.2 was the existence of a minimal distance d_0. However, in order to show that asymptotic homophily is sufficient for ϕ-mixing, we need to be more specific about how the space of individual characteristics is filled as the number of individuals goes to infinity. Specifically, we assume:

Assumption 6 $\lim_{m\to\infty} m d_m^{T+\epsilon} \eta^{d_m} < \infty$ *for all* $\eta \in [0, 1)$ *and for some* $\epsilon > 0$.

where d_m represent the fact that the distance increases as $m \to \infty$ (increasing domain).[13] This is in essence a distributional assumption for the individuals in $\mathcal{X}$. It requires that the tails of the distributions are large enough. If the distribution of individuals on the type is too concentrated, the mixing coefficient $\bar{\phi}_{1,\infty}(r)$ will decrease as m increases, but not enough for assumption 2 to hold. Given this last regularity assumption, we have the following:

Proposition 4.1 *Let* $m \to \infty$. *Suppose that the payoff function is given by (3) for all* $i, j \in N$. *Then, assumptions 1, 5 and 6 imply assumption 2.*

When the payoffs are only dependent through direct links, it is sufficient to show that the probability of a link between an individual in a pair in A and an individual in a pair in B goes to

[13]Specifically assumption 6 must be satisfied for any sequence d_m.

zero fast enough. Since we assumed (assumption 1) that the error term is normally distributed, this probability decreases at exponential rate, which is sufficiently *fast* in the sense of assumption 2.

Assumption 5 is quite natural, and allows to adapt many known theoretical models to our setting. Consider for instance the "Local Spillover" model from Goyal and Joshi (2006):[14]

$$h_i^j(g_m - ij, x) = \psi(n_i(g_m - ij) - 1, n_j(g_m - ij) - 1) - c_{ij}$$

where $\psi : \mathbb{N}^2 \to \mathbb{R}$, and c_{ij} is some positive constant. In this example, the value of a link between i and j is equal to a function of the number of links they have, minus a link-dependent cost. One could adapt their model, and introduce the observed heterogeneity by letting $c_{ij} = d(i,j)$, i.e. a cost equal to the distance between the two individuals in $\mathcal{X}$. Doing so would guarantee the Asymptotic Homophily assumption. We now turn to more general network formation processes.

4.2 More General Models

Proposition 4.1 provides a first encouraging result for the estimation of preferences on networks. However, the specification in (3) excludes many interesting models of network formation. For instance, one could be interested in the following model. Let $\mathbf{C}(g_m, \lambda) = (\mathbf{I} - \lambda\mathbf{W})^{-\mathbf{1}}\mathbf{W}\mathbf{1}$ be the $n \times 1$ vector of Bonacich centrality in the network g, represented by the adjacency matrix $\mathbf{W}$, for some $\lambda \in (0,1)$. The Bonacich centrality accounts for the total number of links (direct and indirect) an individual has, and can be interpreted as a measure of popularity.[15]

Now, define the payoffs as: $h_i^j = h(c_i(g_m - ij, \lambda), c_j(g_m - ij, \lambda), d(i,j))$. This payoff function does not respect the conditions of proposition 4.1 since it depends on indirect links. We will see that we can nonetheless use the same argument to allow for such models. First, we provide some intuition on the class of models which do not respect the ϕ-mixing condition. Suppose that the payoff function is of the following form.[16]

$$h_i^j(g_m - ij, x) = \psi(n_i(g_m - ij), n_j(g_m - ij), L(g_{m,-i-j})) - c_{ij}$$

where $L(g_{m,-i-j}) = \sum_{k \neq i,j} n_k(g_{m,-i-j})$ is the total number of links in the network $g_{m,-i-j}$,

[14]Formally, we are assuming the homogeneity of the function ψ, compared to their original model.
[15]See for instance Mihaly (2009).
[16]This is a loose adaptation of the "Playing the Field" model from Goyal and Joshi (2006)

obtained from g_m by removing all links individuals i and j have in g_m.[17] In that case, the value of a link depends on the whole network, irrespective of the individuals' characteristics. This model does not have the property that the dependence vanishes as the distance between individuals increases, and hence ϕ-mixing is not respected. In order to achieve ϕ-mixing, we have to limit the dependence to the network structure. Specifically:

Assumption 7 (Component Dependence) *For all* $i, j \in N$, $h_i^j(g_m - ij, x; \theta) = h_i^j(g_{m|R_{ij}^g}, x_{R_{ij}^g}; \theta)$

This condition states that the dependence through the network is limited to (finitely) connected individuals. Suppose that the number of individuals in the population is finite. Then, the probability that i and j form a link depends only on the characteristics of the individuals in the same component as i or j.[18] When however, the number of individuals (hence the number of pairs) goes to infinity, we may have two individuals connected through an infinite path. Assumption 7 states that, in that case, those individuals can be treated as disconnected. In other words individuals are unaffected by infinitely distant (in the network) neighbors. Most models of network formation respect this condition as they assume some decay factor.[19] Notice that the previous example where $h_i^j(g_m - ij, x) = \psi(n_i(g_m - ij), n_j(g_m - ij), L(g_{m,-i-j})) - c_{ij}$ does not respect assumption 7. Since h_i^j depends on $L(g_{m,-i-j})$, the payoff function may depend on links between individuals not connected to i nor to j.

Now, by analogy to the specification in (3), we see that it is sufficient for assumption 2 to hold to show that the probability that any two individuals, say s_2 and r_1 are connected through some path goes to zero, i.e. $P(s_2 \leftrightarrow r_1) \to 0$. However, this probability does not only depend on the individuals in pairs in A and B, but also on the individuals in pairs "between" the sets. Figure 2 illustrates.

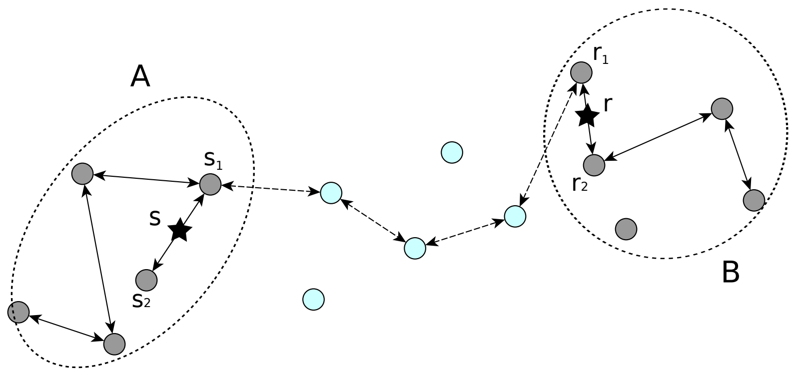
When the number of pairs m (hence the number of individuals n) goes to infinity, there may exists a path of individuals, each of them separated by a finite distance, so $P(A \leftrightarrow B)$ may well be strictly positive. However, since the distance between A and B goes to infinity, this path has to be infinite (i.e. contains an infinite number of individuals). Hence, under assumption 7, the realizations over A and B are independent. Formally,

[17] Specifically, $g_{m,-i-j} = g_m - i1 - ... - in - j1 - ... - jn$.

[18] A component is a maximally connected subnetwork.

[19] Links of degree 1 have more influence than links of degree 2, which have more influence than links of degree 3... and so on. Examples include generalizations the Connection Model from Jackson and Wolinsky (1996), and models based on the Bonacich centrality.

Figure 2: ϕ-mixing on Networks



Proposition 4.2 *Assumptions 1, 5, 6 and 7 imply assumption 2 as $m \to \infty$.*

Proposition 4.2 shows that the class of models that can be estimated using (1) is quite large. It also provide easy to check conditions for applied researchers wanting to estimate some arbitrary model of network formation. In practice, provided that the choosen structural form for $h_i^j(g, x; \theta)$ respects Asymptotic Homophily and Component Dependence, one can estimate $\theta \in \Theta$ using the PMLE defined in (1).

In the next section, we discuss the existence and potential multiplicity of pairwise stable networks.

4.3 Existence and Multiplicity

In the previous sections, we implicitly assumed that the set of pairwise stable networks was non-empty, and unique. In general, this may not be true. General conditions for the existence of a pairwise stable network are well known.[20] One result that is particularly adapted to our setting is the fact that monotone preferences imply the existence of at least one pairwise stable network. Formally:

Definition 4 (Monotonicity) *A payoff function is* ***monotone*** *if for any $g_m, g'_m \in \mathbb{G}_m$ such that $g_m \subseteq g'_m$, we have that $h_i^j(g_m - ij, x, \theta) \leq h_i^j(g'_m, x, \theta)$ for all $i, j \in N$.*

Monotone payoff functions have the convenient property that the set of pairwise stable networks is non-empty, irrespective of the value of the unobserved term ε_{ij}. To see why, consider

[20] For general existence results for pairwise stable networks, see Jackson and Watts (2001) and Chakrabarti and Gilles (2007).

the following simple algorithm. Starting from the empty network, we add links sequentially if $H_i^j(g_m - ij, x; \theta, \varepsilon_i) \geq 0$ and $H_j^i(g_m - ij, x; \theta, \varepsilon_j) \geq 0$. The link creation process stops when there exists no such profitable link creation. Since the payoff function is monotone, the creation of a link increases the value of the existing links so $H_i^j(g_m - ij, x; \theta, \varepsilon_i) \geq 0$ implies that $H_i^j(g_m - ij + kl, x; \theta, \varepsilon_i) \geq 0$ for any link kl. The network generated by this sequential creation of links is then pairwise stable.

Another issue that has not been addressed is the potential existence of multiple equilibria.[21] A specific feature of pairwise stable networks is the complexity of the equilibrium set. In general, one cannot explicitly find the set of pairwise stable networks, as showing existence is already challenging (see Sheng, 2012). Also, recall that, in our model, we assumed that we observe only one equilibrium of the game, and not the other (potential) equilibria. This is a specific feature of the model which differs from the existing literature. In particular, even under presence of multiple equilibria, the likelihood function is always coherent in the sense that $P(w_{ij} = 1) + P(w_{ij} = 0) = 1$. Then, even under multiplicity of equilibria, our estimator remains a well defined extremum estimator, where the objective function is *the probability that the observed network is pairwise stable*.

Let $\mathbb{G}_m^2$ be the powerset of $\mathbb{G}_m$ and let $\Upsilon_\theta : \mathbb{G}_m^2 \to \mathbb{G}_m$ be an equilibrium selection mechanism. That is, for any $E \subseteq \mathbb{G}_m$, $\Upsilon_\theta(E) \in E$. Then, $\mathcal{L}_m(\theta; g, x)$ is a pseudo estimator of $\mathbb{P}(g|x, \theta; g = \Upsilon_\theta(E_x))$, where $E_x \subseteq \mathbb{G}_m$ is the set of pairwise stable networks given x. That is, we maximize the (pseudo) likelihood of $g_m \in \mathbb{G}_m$, conditional of the fact that g_m is selected. Then, our estimator is consistent provided that Υ_θ is independent of θ, i.e. $\Upsilon_\theta = \Upsilon$ for all $\theta \in \Theta$. (This independence assumption plays the same role as Leung's (2013) Sampling Experiment assumption. See his discussion on page 13.) However, the validity of the estimation procedure under the presence of other potential equilibria is unclear if $\theta \in \Theta$ influences the equilibrium selection mechanism. Formally understanding the properties of the estimator under multiple equilibria (and with spatially dependent observations) goes far beyond the scope of this paper and is left for future research.

In the next section, we provides an empirical application of our method using communication networks.

[21] See Bisin et al. (2011), Galichon and Henry (2011), Sheng (2012) and Tamer (2003).

5 Estimation

In this section, we apply the methodology developed in the previous sections to estimate a model of network formation. We first discuss the conditions for the identification of (1). We then apply the model to the formation of friendship networks using the Add Health database.

5.1 Identification in Pairwise Models of Network Formation

In this section, we discuss the identification of network formation models based on pairwise stability. Our discussion applies to most existing empirical models of network formation, e.g Christakis et al. (2009), Goldsmith-Pinkham and Imbens (2011), Mele (2011), and Sheng (2012). Recall that a link exists if and only if it is profitable for both individuals. Specifically,

$$w_{ij} = 1 \text{ iff } H_i^j(\theta) > 0 \text{ and } H_j^i(\theta) > 0$$

This implies that we are facing a problem of partial observability (Poirier, 1980). Of the four possibilities implied by the model, i.e $A = \{H_i^j(\theta) > 0, H_j^i(\theta) > 0\}$, $B = \{H_i^j(\theta) > 0, H_j^i(\theta) < 0\}$, $C = \{H_i^j(\theta) < 0, H_j^i(\theta) > 0\}$, $D = \{H_i^j(\theta) < 0, H_j^i(\theta) < 0\}$, we observe only $\{A\}$, or $\{B, C, D\}$. This creates identification issues. Poirier (1980) discusses potential sources of identification. One is through exclusion restrictions, i.e. $H_i^j(\theta)$ includes variables excluded from $H_j^i(\theta)$. We argue that this approach is problematic for most social network analysis. Suppose that we have the following:

$$\begin{aligned} H_i^j(\theta) &= \theta_1 x_{1,j} + \theta_2 (x_{2,i} - x_{2,j})^2 + \varepsilon_{ij} \\ H_j^i(\theta) &= \theta_1 x_{1,i} + \theta_2 (x_{2,i} - x_{2,j})^2 + \varepsilon_{ji}. \end{aligned}$$

The interpretation is that the value for i of a link with j depends on j's characteristic $(x_{1,j})$, and on the distance between them (i.e. $(x_{2,i} - x_{2,j})^2$ for a characteristic x_2), and similarly for j. However, notice that by redefining $\tilde{H}_i^j = H_j^i$ and $\tilde{H}_j^i = H_i^j$, we have:

$$\begin{aligned} \tilde{H}_j^i(\theta) &= \theta_1 x_{1,j} + \theta_2 (x_{2,i} - x_{2,j})^2 + \varepsilon_{ij} \\ \tilde{H}_i^j(\theta) &= \theta_1 x_{1,i} + \theta_2 (x_{2,i} - x_{2,j})^2 + \varepsilon_{ji}. \end{aligned}$$

I that case, the interpretation of the parameters is reversed. The value for i of a link with j depends on *his own* characteristics, and on the distance between him and j (and similarly for j). As pointed out by Poirier (1980), this is an issue of labelling and prevents global identification of the model.

A specificity of most social networks is that the labelling of the individuals in a pair have no economic meaning. This means that we have no microeconomic foundations which allows to choose one interpretation over the other. There exists however a specific type of networks in which those exclusion restrictions may have some economic justifications. They are called bipartite networks.

In a bipartite network, the population can be separated into distinct subsets such that no link exists within a set. Prominent examples are buyer-seller networks and labour-marker networks (e.g. Kranton and Minehart (2011) and Elliott (2013)).[22] Bipartite networks may provide intuitive exclusion restrictions. Consider a labour-market network, label $i = worker$ and $j = firm$, and define the following preferences.[23]

$$
\begin{aligned}
H_i^j(\theta) &= \theta_1 x_{1,ij} + \theta_2 x_{2,ij} + \varepsilon_{ij} \\
H_j^i(\theta) &= \theta_3 x_{3,i} + \theta_4 x_{2,ij} + \varepsilon_{ji}
\end{aligned}
$$

where $x_{1,ij}$ represents the distance (in Km) between the worker's house and the firm, $x_{2,ij}$ represents the (sector specific) market wage, and $x_{3,i}$ is a measure of the worker's productivity. In that case, we have microeconomic justification for imposing this exclusion restriction, as well as for imposing $\theta_1 < 0$, $\theta_2 > 0$, $\theta_3 > 0$ and $\theta_4 < 0$. Moreover, as every pair is of the type firm-worker, the labelling is no longer arbitrary which allows for the identification of θ.

In most contexts however, networks are not bipartite and we have no ex-ante justification for imposing such exclusion restrictions. Hence, the interpretation of non-symmetric preferences (i.e. such that $h_i^j \neq h_j^i$) is problematic. For that reason, we argue that pairwise analysis of (non-bipartite) network formation models should focus on symmetric preferences.[24] As discussed, in Poirier (1980, p.213), under symmetry and linearity, identification generically holds. In the next

[22] All the results of the paper apply for bipartite network. Notice however that the number of admissible pairs shrinks drastically.

[23] Thanks to Diego Cerdeiro for interesting discussions on that example.

[24] This distinction between bipartite and non-bipartite networks is also important for existence issues as it is much easier to show existence for bipartite networks.

Table 1: Variables' description

Variables	Descriptions
Linked	Binary variable, equals 1 if the individuals are linked.
Popular	Popularity of the two individuals in the pair. Specifically: $n_i(g_m - ij) + n_j(g_m - ij)$.
$\Delta(Grade)$	Difference between the individuals' grade level, in absolute value.
$\mathbb{I}(Gender)$	Binary variable, equals 1 if the individuals are of the same gender.
$\mathbb{I}(Whites)$	Binary variable, equals 1 if both individuals are Whites
$\mathbb{I}(Blacks)$	Binary variable, equals 1 if both individuals are Blacks
$\mathbb{I}(Hisps)$	Binary variable, equals 1 if both individuals are Hispanics
$\Delta(Phys)$	Difference between the individuals' physical attractiveness (see appendix for details)
$\Delta(Psych)$	Difference between the individuals' psychological attractiveness (see appendix for details)
Work	Number of hours worked by the individuals in a week (sum)
$\Delta(Geo)$	Geographical distance between the individuals' residence (normalized, see appendix for details)

section, we implement our approach using the Add Health database.

5.2 Friendship Networks

We use the Add Health database which provides information on friendship networks for high-school teenagers in the US. We concentrate on the "saturated sample", which provides information on 3 449 teenagers, coming from 16 schools. Tables 2 and 3 give a summary of the variables used for individuals, and for pairs of individuals.[25] Table 1 present the variables' definitions. Precise definitions and technical information on constructed variables can be found in appendix.

As discussed previously, friendship networks are not bipartite so we assume $h_i^j = h_j^i$ for all $i, j \in N$. We use the following functional form:

$$\begin{aligned} h_i^j(g_m, x; \theta) &= \theta_1 Popular_{ij} + \theta_2 \Delta(Grade_{ij}) \\ &+ \theta_3 \mathbb{I}(Gender_{ij}) + \theta_4 \mathbb{I}(Whites_{ij}) + \theta_5 \mathbb{I}(Blacks_{ij}) \\ &+ \theta_6 \mathbb{I}(Hisps_{ij}) + \theta_7 \Delta(Phys_{ij}) + \theta_8 \Delta(Psych_{ij}) \\ &+ \theta_9 Work_{ij} + \theta_{10} \Delta(Geo_{ij}) + \theta_{11} \end{aligned} \quad (4)$$

[25]Notice that the racial variable are not necessarily exclusive. We also omitted racial categories "Asian", "Native" and "Other".

Table 2: Descriptive Statistics for the Individuals

Variable	Mean	Std.	Min	Max
$n_i(g_m)$	4.2422	3.309961	0	21
Gender (Female=1)	0.4954	0.5001	0	1
Hispanic	0.2017	0.4013	0	1
White	0.5868	0.4925	0	1
Black	0.1594	0.3661	0	1
Grade	10.1940	1.4932	7	12
Hours Worked	8.3202	11.8225	0	100
Physical	3.5429	0.8510	1	5
Psychological	3.5583	0.8192	1	5

Number of Individuals: 3 449
Number of Communities: 16

Table 3: Descriptive Statistics for the Pairs

Variable	Mean	Std.	Min	Max
Linked	0.0012	0.0345	0	1
Popular	8.4807	4.6771	0	41
$\Delta(Grade)$	1.6448	1.3248	0	5
$\mathbb{I}(Gender)$	0.4999	0.5000	0	1
$\mathbb{I}(Whites)$	0.3433	0.4748	0	1
$\mathbb{I}(Blacks)$	0.0254	0.1573	0	1
$\mathbb{I}(Hisps)$	0.0406	0.1973	0	1
$\Delta(Phys)$	0.9016	0.8115	0	6
$\Delta(Psych)$	0.8620	0.7888	0	6
Work	16.6356	16.7135	0	200
$\Delta(Geo)$	72.8096	44.3516	0	100

Number of Pairs: 5 946 076
Number of Community Clusters: 256

Table 4: Bivariate Probit with Partial Observability (marginal effects)

Variable	dy/dx ($\times 1000$)	Std. Err. ($\times 1000$)
Popular	$0.0495^{\dagger}$	(0.0305)
$\Delta(Grade)$	-0.6405^{*}	(0.2636)
$\mathbb{I}(Gender)$	0.2008^{*}	(0.0871)
$\mathbb{I}(Whites)$	0.8543^{**}	(0.2536)
$\mathbb{I}(Blacks)$	0.9059^{*}	(0.3595)
$\mathbb{I}(Hisps)$	0.2983	(0.1931)
$\Delta(Phys)$	-0.0282	(0.0228)
$\Delta(Psych)$	-0.0291^{*}	(0.0115)
Work	$-0.0034^{\dagger}$	(0.0018)
$\Delta(Geo)$	-0.0388^{*}	(0.0165)

$\varrho = 0.9992$ (0.0001)
(pseudo) log-Likelihood: -40313.979

Significance levels: † = 10%, * = 5%, and ** = 1%

where $\theta_1 > 0$, and θ_{11} represents the intrinsic value of a link. The restriction $\theta_1 > 0$ is needed to ensure that preferences are monotonic, which implies the existence of a Pairwise Stable network (see section 4.3). Following Bester et al. (2012), we estimate the specification in (4) using Cluster-Robust standard errors. Marginal effects are reported in Table 4. Notice that while we assumed $\theta_1 > 0$ in (4), we did not used that restriction for the estimation.

Table 4 shows a positive effect of popularity. More connected individuals have a higher probability of creating friendship relations. Not surprisingly, being of the same grade level has a positive influence on the formation of friendship relations. We also find evidence of racial segregation, which seems to be stronger for Blacks than for Whites and Hispanics. Distance in terms of personality traits have negative impact, as for an increase in the number hours worked. This last finding suggests a substitution between the time spent with friends, and the time spent working. Geographical distance also have a negative impact. The estimated covariance of the error term is fairly high ($\varrho = 9992$). This suggests that the errors are strongly correlated, and that the benefit of using a bivariate probit instead of a univariate probit is small (see next section for a discussion).

In the next section, we present alternative specifications and discuss some practical consideration while estimating the model.

5.3 Alternative Specifications and Practical Considerations

As this application shows, the approach used in this paper is promising as it has the advantage of being intuitive, flexible, and simple to implement. In this section, we discuss alternative specifications and the link with games with transfers, as well as large databases and subsampling issues.

5.3.1 Bivariate Probit, Probit, and TU Games

As we mentioned in section 2, our analysis also applies to the case where we impose $\varepsilon_{ij} = \varepsilon_{ji} \sim N(0,1)$. In that case, the resulting estimator is no longer a bivariate probit, but an univariate probit. In general, the bivariate probit is more flexible as is includes more unknown parameters. However, under the assumption that $h_i^j = h_j^i$, the bivariate and univariate probit estimators only differ by one parameter: ϱ. It is then worthwhile to look at the univariate probit estimator.

Interestingly, the probit estimator can be interpreted in terms of games with transferable utilities (TU games).[26] Specifically, lets define a TU-Pairwise Stable network as follows:

Definition 5 *A network g_m is **TU-Pairwise Stable** if, for all $i, j \in N$:*
1) $w_{ij} = 1$ if [$H_i^j(g_m - ij, x; \theta, \varepsilon_i) + H_j^i(g_m - ij, x; \theta, \varepsilon_j) \geq 0$]
2) $w_{ij} = 0$ otherwise

With the additional assumptions that $H_i^j = H_j^i = h_i^j + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0,1)$, we have:

$$w_{ij} = 1 \text{ if } [\ h_i^j(g_m - ij, x; \theta) + \varepsilon_{ij} \geq 0\ ]$$

and the resulting (pseudo) estimator is a probit. Table 5 displays estimation results for a probit model, using the specification in (4). Results are quite similar to those of Table 4, with the distinction that the marginal effects have twice the magnitude. The reason it that marginals effects reported in Table 4 capture the effects on the value received by *one* of the individuals in the pair, while those of Table 5 capture the effects on the value of the pair, which is shared by *both* individuals. TU games also have the (non trivial) advantage that the existence of a (TU)-pairwise stable network holds for larger sets preferences (compared with the existence of a pairwise stable network).

We now discuss a sampling procedure for large databases.

[26]Sheng (2012) also makes the distinction between pairwise stability for games with and without transfers.

Table 5: Probit, i.e. assuming $\varepsilon_{ij} = \varepsilon_{ji}$ (marginal effects)

Variable	dy/dx ($\times$1000)	Std. Err. ($\times$1000)
Popular	0.0990†	0.0610
$\Delta(Grade)$	-1.2810*	0.5273
$\mathbb{I}(Gender)$	0.4015*	0.1742
$\mathbb{I}(Whites)$	1.7086**	0.5072
$\mathbb{I}(Blacks)$	1.8119*	0.7190
$\mathbb{I}(Hisps)$	0.5967	0.3862
$\Delta(Phys)$	-0.0565	0.0456
$\Delta(Psych)$	-0.0581*	0.0229
Work	-0.0068†	0.0037
$\Delta(Geo)$	-0.0776*	0.0329
(pseudo) log-Likelihood: -40313.978		

Significance levels: † = 10%, * = 5%, and ** = 1%

5.3.2 Sampling

Recall that the number of observation over a sample of n individuals is equal to the number of pairs in the sample, i.e. $n(n-1)/2$. For instance, in our application, $n = 3449$ which leads to a model with 5 946 076 observations. For large databases, the estimation could then be prohibitively time consuming. For such large databases, we suggest to use a random sample. Table 6 reports the results for the bivariate probit model with a random subsample of 25% of the database. The final sample has 1 486 519. Notice that the results are quite similar to those of Table 4. If this approach is compelling however, one has to keep in mind that the constructed variable dependent on the network structure have to be constructed *over the whole sample*. In our case, this implies that the variable *Popular* was constructed *before* the subsampling procedure.

We now briefly conclude.

6 Conclusion

In this paper, we have developed a micro-founded econometric model of network formation which requires the observation of only one social network. We have shown that an asymptotic version of homophily is sufficient for ϕ-mixing, which implies that the estimation of the underlying preference parameters can be achieved using a simple Pseudo Maximum Likelihood estimator. The methodology is appealing as it is simple, and flexible. We also discussed identification and

Table 6: Bivariate Probit with Partial Observability (marginal effects): Subsample

Variable	dy/dx (×1000)	Std. Err. (×1000)
Popular	0.0557†	0.0320
$\Delta(Grade)$	-0.6595*	0.2909
$\mathbb{I}(Gender)$	0.2027*	0.0977
$\mathbb{I}(Whites)$	0.8141**	0.2451
$\mathbb{I}(Blacks)$	0.9760*	0.3946
$\mathbb{I}(Hisps)$	0.3206	0.1967
$\Delta(Phys)$	0.0060	0.0243
$\Delta(Psych)$	-0.0317*	0.0385
Work	-0.0027†	0.0016
$\Delta(Geo)$	-0.0400*	0.0189
$\varrho = 0.9989$ (0.0002)		
(pseudo) log-Likelihood: -10084.732		

Significance levels: † = 10%, * = 5%, and ** = 1%

estimation issues and compared our baseline model with games on bipartite networks and games with transferable utilities. We provided an empirical application using the formation of friendship networks among American teenagers. We found positive influence of popularity on link formation, as well as evidence of homophily on gender, race, and geographic localisation.

References

Bester A.C., Conley T.G. and Hansen C.B. (2010), "Inference with Dependent Data Using Cluster Covariance Estimators" *Working Paper*

Bisin A., Moro A and Topa G. (2011) "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions", *Working Paper*

Blume L. Brock W., Durlauf S. and Ioannides Y. (2011) "Identification of Social Interactions", Chapter 18 in Social Economics, Benhabib, Bison and Jackson, North-Holland.

Boucher V. (2012) "Structural Homophily", *Working Paper*

Bramoullé Y., Djebbari H., and Fortin B. (2009) "Identification of peer effects through social networks", Journal of Econometrics, 150, 41-55

Bramoullé Y. and Rogers B. (2010) "Diversity and Popularity in Social Networks", *Working Paper*

Chakrabarti S. and Gilles R.P. (2007) "Network Potentials", The Review of Economic Design, 11, 13-52

Christakis N., Fowler J., Imbens G.W. and Kalyanaraman K. (2010) "An Empirical Model for Strategic Network Formation", *Working Paper*

Conley T. (1999) "GMM Estimation with Cross Sectional Dependence", Journal of Econometrics, 92, 1-45

Currarini S., Jackson M. O., Pin P. (2009) "An Economic Model of Friendship: Homophily, Minorities, and Segregation", Econometrica, 77, 1003-1045

Currarini S., Jackson M. O., Pin P. (2010) "Identifying the roles of race-based choice and chance in high school friendship network formation," Proceedings of the National Academy of Sciences of the United States of America, 107(11): 4857-4861

Echenique F. and Fryer R.G. (2007) "A Mesure of Segregation Based on Social Interactions" The Quarterly Journal of Economics, 122(2), 441-485

Elliott M. (2013) "Inefficiencies in Networked Markets", *Working Paper*

Galichon A. and Henry M. (2011) "Set identification in models with multiple equilibria", Review of Economic Studies, 78(4), 1264-1298

Gallant R.A. and White H. (1988) "A unified theory of estimation and inference for nonlinear dynamic models", B. Blackwell, 155p.

Goldsmith-Pinkham P. and Imbens G.W. (2011), "Social Networks and the Identification of Peer Effects", *Working Paper*

Gourieroux C. and Monfort A. (1995) "Statistics and Econometric Models: Volume 1, General Concepts, Estimation, Prediction and Algorithms", Cambridge University Press, 524p.

Gourieroux C., Monfort A., Trognon A. (1984) "Pseudo Maximum Likelihood Methods: Theory" Econometrica, 52(3), 681-700

Goyal S. and Joshi S. (2006) "Unequal Connections", International Journal of Game Theory, 34, 319-349

Henry M. and Mourifié I. (2011) "Euclidean revealed preferences: testing the spatial voting model" Journal of Applied Econometrics, *Forthcoming*

Iijima R. and Kamada Y. (2010) "Social Distance and Network Structures", *Working Paper*

Jackson M. O. (2008) Social and Economic Networks, Princeton University Press

Jackson M. O. and Watts (2001) "The existence of Pairwise Stable Networks", Seoul Journal of Economics, 14(3), 299-321

Jackson M. O. and Wolinsky (1996) "A Strategic Model of Social and Economic Networks", Journal of Economic Theory, 71, 44-74

Jenish N. and Prucha I.R. (2009) "Central limit theorems and uniform laws of large numbers for arrays of random fields", Journal of Econometrics, 150, 86-98

Johnson C. and Gilles R. P. (2000) "Spatial Social Networks", Review of Economic Design, 5, 273-299

Kranton R. E. and Minehart D. F. (2011) "A Theory of Buyer-Seller Networks", The American Economic Review, 91(3), 485-508

van der Leij M., Rolfe M. and Toomet O. (2009) "On the Relationship Between Unexplained Wage Gap and Social Network Connections for Ethnical Groups", *Working Paper*

Leung M. (2013) "Two-Step Estimation of Network-Formation Models with Incomplete Information", *Working Paper*

Marmaros D. and Sacerdote B. (2006) "How Do Friendships Form?", The Quarterly Journal of Economics, 121(1), 79-119

McFadden (1981) "Econometric Models of Probabilistic Choice", in C. Manski and D. McFadden (editors), Structural Analysis of Discrete Data with Econometric Applications, Cambridge, Mass.:M.I.T.Press.

Mele A. (2010) "A Structural Model of Segregation in Social Networks", *Working Paper*

Patacchini E. and Zenou Y. (2009) "Ethnic Networks and Employment Outcomes", Regional Science and Urban Economics, *Forthcoming*

Pinkse J. and Slade M.E. (1998), "Contracting in Space: An application of spatial statistics to discrete-choice models", Journal of Econometrics, 85, 125-154

Wang H., Iglesias E.M., and Wooldridge J.M. (2009) "Partial Maximum Likelihood Estimation of a Spatial Probit Model", *Working Paper*

Watts A. (2007) "Formation of Segregated and Integrated Groups", International Journal of Game Theory, 35:505-519

White H. (2001) "Asymptotic Theory for Econometricians", Academic Press, 264p.

7 Appendix

Proof of Theorem 3.1

Under assumption 3.1, it is sufficient to show that:[27]

$$\sup_{\theta\in\Theta} |\mathcal{L}_m(\theta) - \mathbb{E}(\mathcal{L}_m(\theta))| \rightarrow_{a.s.} 0, \text{ as } m \rightarrow \infty.$$

In order to show that this condition hold, it is sufficient to show that the conditions of theorem 2 and 3 from Jenish and Prucha (2009) hold. Specifically,

1. $d(r,s) > d_0 > 0$ for any $r, s \in S_m$

2. $(\Theta, \|.\|)$ is a totally bounded metric space.

3. Domination:
$$\lim_{m\rightarrow\infty} \sup \frac{1}{|S_m|} \sum_{s=1}^{m} \mathbb{E}(\bar{q}^p_{s,m} 1_{\{\bar{q}^p_{s,m} > k\}}) \rightarrow 0 \text{ as } k \rightarrow \infty,$$
for some $p \geq 1$, and where $\bar{q}_{s,m} = \sup_\theta |q_{s,m}(z_{s,m}|x, g_m, \theta)|$.

4. Stochastic equicontinuity: For every $\epsilon > 0$,
$$limsup_m \frac{1}{|S_m|} \sum_{s=1}^{m} P(\sup_{\theta'\in\Theta} \sup_{\theta\in B(\theta',\delta)} |q_{s,m}(\theta) - q_{s,m}(\theta')| > \epsilon) \rightarrow 0 \text{ as } \delta \rightarrow 0,$$
where $B(\theta', \delta)$ is the open ball $\{\theta \in \Theta : \|(\theta' - \theta)\| < \delta\}$.

5. $\sup_m \sup_{s\in S_m} \mathbb{E}[\sup_{\theta\in\Theta} |q_{s,m}(\theta)|^{(1+\eta)}] < \infty$ for some $\eta > 0$.

6. $\sum_{d=1}^{\infty} d^{T-1} \bar{\phi}_{1,1}(d) < \infty$.

Condition 1 is implied by assumption 3.2. Condition 2 is verified by construction, and condition 5 and 6 are just assumption 3.3 and ϕ-mixing(2). Conditions 3 and 4 hold from the following: Under condition 5, $\sup_\theta |q_{s,m}(z_{s,m}|x, g_m, \theta)|$ is $L^{(1+\eta)}$ integrable which implies the uniform $L^{(1+\eta)}$ integrability of $|q_{s,m}(z_{s,m}|x, g_m, \theta)|$.

The next lemma shows that assumption 3.4 implies condition 4.

Lemma 7.1 *Condition 4 is implied by assumption 3.4.*

[27] see for instance Gallant and White (1988), pp.18.

Proof From the mean value theorem, we can write

$$q_{s,m}(\theta) \quad = \quad q_{s,m}(\theta') + \frac{\partial q_{s,m}(\tilde{\theta})}{\partial \theta}(\theta - \theta'),$$

Thus,

$$\begin{aligned} |q_{s,m}(\theta) - q_{s,m}(\theta')| &\leq |\frac{\partial q_{s,m}(\tilde{\theta})}{\partial \theta}| \|(\theta - \theta')\| \\ &\leq \sup_{\theta \in \Theta} |\frac{\partial q_{s,m}(\theta)}{\partial \theta}| \|(\theta - \theta')\|. \end{aligned}$$

According to Proposition 1 of Jenish and Prucha (2009), $q_{s,m}(\theta)$ is L_0 stochastically equicontinuous on Θ if the following *Cesàro* sums is finite. i.e

$$limsup_m \frac{1}{|S_m|} \sum_{s=1}^{m} \mathbb{E}(\sup_{\theta \in \Theta} |\frac{\partial q_{s,m}(\theta)}{\partial \theta}|) < \infty.$$

However, under assumption 3.4, each term of the *Cesàro* sums is finite, in the sense that $\sup_m \sup_{s \in S_m} \mathbb{E}[\sup_{\theta \in \Theta} |\frac{\partial q_{s,m}(\theta)}{\partial \theta}|] < \infty$. This fact completes the proof. $\square$

From the previous lemma, conditions 1-6 are respected, hence theorem 2 and 3 from Jenish and Prucha (2009) apply. This completes the proof. $\square$

Proof of Theorem 3.2

We want to show that $\sqrt{m}(\hat{\theta}_m - \theta_0) \Rightarrow N(0, D_0(\theta_0)^{-1} B_0(\theta_0) D_0(\theta_0)^{-1})$. From the mean value theorem, we have that

$$\begin{aligned} \frac{\partial \mathcal{L}_m(\hat{\theta}_m)}{\partial \theta} &= \frac{\partial \mathcal{L}_m(\theta_0)}{\partial \theta} + \frac{\partial^2 \mathcal{L}_m(\overline{\theta}_m)}{\partial \theta \partial \theta'} \\ 0 &= \frac{\partial \mathcal{L}_m(\theta_0)}{\partial \theta} + \frac{\partial^2 \mathcal{L}_m(\overline{\theta}_m)}{\partial \theta \partial \theta'}(\hat{\theta}_m - \theta_0). \end{aligned}$$

and

$$\begin{aligned} \sqrt{m}(\hat{\theta}_m - \theta_0) &= -\sqrt{m}[\frac{\partial^2 \mathcal{L}_m(\overline{\theta}_m)}{\partial \theta \partial \theta'}]^{-1} \frac{\partial \mathcal{L}_m(\theta_0)}{\partial \theta} \\ &= -[\frac{\partial^2 \mathcal{L}_m(\overline{\theta}_m)}{\partial \theta \partial \theta'}]^{-1} [\frac{\sigma_m}{\sqrt{m}}][\sigma_m^{-1} Q_m], \end{aligned}$$

where $\sigma_m^2 = Var(Q_m)$ and $Q_m = \sum_{s=1}^{m} \frac{\partial q_{s,m}(\theta_0)}{\partial \theta}$.

Then, it is sufficient to show the following:

1. $\frac{\sigma_m^2}{m} \to B_0(\theta_0)$;

2. $\sigma_m^{-1} Q_m \Rightarrow N(0, I)$;

3. $[\frac{\partial^2 \mathcal{L}_m(\bar{\theta}_m)}{\partial\theta\partial\theta'}] \to_p D_0(\theta_0)$.

Again, we proceed in a series of lemmata.

Lemma 7.2 *Under assumptions 3.1,* $\frac{\sigma_m^2}{m} \to B_0(\theta_0)$.

Proof

$$\begin{aligned}
\frac{1}{m}\sigma_m^2 &= \frac{1}{m} Var(m\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta}) \\
&= m\mathbb{E}[\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta}\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta'}] + m\mathbb{E}[\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta}]\mathbb{E}[\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta'}] \\
&= m\mathbb{E}[\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta}\left(\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta}\right)'].
\end{aligned}$$

where the last inequality holds since $\mathbb{E}[\frac{\partial \mathcal{L}_m(\theta_0)}{\partial\theta}] = 0$, as θ_0 maximizes $\mathbb{E}[\mathcal{L}_m(\theta)]$ (Assumption 3.1). Hence, $\frac{\sigma_m^2}{m} \to B_0(\theta_0)$. □

Lemma 7.3 *Under assumptions 1, and 4,* $\sigma_m^{-1} Q_m \Rightarrow N(0, I)$

Proof It is sufficient to show that the conditions for theorem 1 from Jenish and Prucha (2009) hold. Specifically,

1. $d(r, s) > d_0 > 0$ for any $r, s \in S_m$.

2. ϕ-mixing on Random Fields.

3. $\sup_m \sup_{s \in S_m} \mathbb{E}[\sup_{\theta\in\Theta} |\frac{\partial q_{s,m}(\theta)}{\partial\theta}|^2] < \infty$.

4. $\liminf_{m\to\infty} \frac{\sigma_m^2}{m} > 0$.

Condition 1 is implied by assumption 3.2. Condition 3 is just assumption 4.5, and condition 4 is implied by lemma 7.2. □

Lemma 7.4 $\frac{\partial^2 \mathcal{L}_m(\bar{\theta}_m)}{\partial\theta\partial\theta'} \to_p D_0(\theta_0)$

Proof The proof is identical to the proof for the consistency of $\hat{\theta}$, replacing $q_{s,m}(\theta)$ by $D_{s,m}(\theta)$, and using assumptions 4.3 and 4.4 instead of assumptions 3.3 and 3.4. □

Putting together lemmata 7.2, 7.3 and 7.4 completes the proof. □

Proof of Proposition 4.1

Let $H_i^j = h_i^j[N_i(g), N_j(g), d(i,j)] + \varepsilon_{ij}$ where $\varepsilon \sim N(0,1)$. We show that under assumption 5 and 6, ϕ-mixing is respected. Recall that $\phi(\mathcal{A},\mathcal{B}) = \sup\{|P(A|B) - P(A)|, A \in \mathcal{A}, B \in \mathcal{B}, P(B) > 0\}$. Formally, A and B are subsets of pairs, i.e. $A, B \in S_m$. Let $i \in s \in A$ and $j \in s \in B$.

We have that $P(A) = P(A|\exists ij \in g)P(\exists ij \in g) + P(A|\nexists ij \in g)P(\nexists ij \in g)$ and $P(A|B) = P(A|B \cap \exists ij \in g)P(\exists ij \in g) + P(A|B \cap \nexists ij \in g)P(\nexists ij \in g)$. Since the payoff function only depends on direct links, $P(A|B \cap \nexists ij \in g) = P(A|\nexists ij \in g)$. Hence, we can rewrite

$$\phi(\mathcal{A},\mathcal{B}) = \phi(\mathcal{A},\mathcal{B}|\exists ij \in g)P(\exists ij \in g)$$

Since, for any A, B, $\phi(\mathcal{A},\mathcal{B}) \in [0,1]$, we have that $\phi(\mathcal{A},\mathcal{B}) \leq P(\exists ij \in g)$. Let $\bar{h}(d) = \sup_\theta \sup_g \sup_{ij} h_i^j(g,x,d;\theta)$ and $\underline{h}(d) = \inf_\theta \inf_g \inf_{ij} h_i^j(g,x,d;\theta)$. We then have that $\bar{\phi}_{k,l} \leq 4kl\Phi[\bar{h}(d)]$ since there can be a maximum of $2k$ individuals in A and $2l$ individuals in B. That is, the sum of the probabilities for each possible pairs between A and B, and for the maximal value for h_i^j. Notice that by the properties of the Hausdorf distance, if $d(i,j) \geq c$ for some $c > 0$ and all $i \in s \in A$ and $j \in r \in B$, then $d(A,B) \geq c$.

Now, we know that the Chernoff bound for Φ is such that $\Phi[\bar{h}(d)] \leq \frac{1}{2}\exp\{-\frac{1}{2}\bar{h}(d)^2\}$ for $\bar{h}(d) < 0$, which is true for d big enough from assumption (5.1). Then, a sufficient condition for assumption (2.1) and (2.2) is $\bar{\phi}_{k,l}(d) \leq 2kl\exp\{-\frac{1}{2}\bar{h}(d)^2\}$ for $k + l \leq 4$ or equivalently:

$$d^{T-1}\bar{\phi}_{k,l}(d) \leq 2kld^{T-1}\exp\{-\frac{1}{2}\bar{h}(d)^2\}$$

for all $d > \bar{d}$ for some $\bar{d} > 0$ and $k + l \leq 4$. Then, assumption (2.1) and (2.2) hold if $\sum_{d=1}^{\infty} d^{T-1}\exp\{-\frac{1}{2}\bar{h}(d)^2\}$ converges. According to Cauchy's rule, this last sum converges if $\lim_{d\to\infty}\exp\left\{-\frac{\bar{h}(d)^2}{2d}\right\} \in [0,1)$. Which is true under assumption (5.2).

Now, ϕ-mixing (3) is different since $l = \infty$, so the upper bound goes to infinity. Specifically, condition (3) implies that there exists $C > 0$, $\overline{m}$, $\bar{d}$ such that $d^{T+\epsilon}\bar{\phi}_{1,m}(d) \leq C$, for some $\epsilon > 0$, for all $m > \overline{m}$ and $d > \bar{d}$. Using the Chernoff bound again we have that $d^{T+\epsilon}\bar{\phi}_{1,m}(d)$ is bounded when m goes to infinity if

$$\lim_{m\to\infty} m d_m^{T+\epsilon} \exp\left\{-\frac{\bar{h}(d_m)^2}{2d_m}\right\} < \infty$$

assuming the increasing domain assumption 6, and the asymptotic homophily assumption (5.2). $\square$

Proof of Proposition 4.2

Let $P(A \leftrightarrow B)$ be the probability that there exist a path between an individual in a site in A and and individual in a site in B. Using the same argument as in the proof of proposition 4.1, we have that $\phi(\mathcal{A},\mathcal{B}) \leq P(A \leftrightarrow B)$. The probability $P(A - B)$ is however not trivial to compute. Instead, we use the fact that $P(A - B) = P(\exists k - B : ik \in g)$ for some i in a site in A. Since k is connected to B, there are two possibilities: (1) the distance between k and B is finite, or (2) the distance between k and B is infinite, and k is reached from B using an infinite number of links.

We start with the second possibility. In that case, from assumption 7, the realization on B does not depends on k, hence $P(A|B) = P(A)$.

Now, suppose that the distance between k and B is finite. Then, as in the proof of proposition 4.1, we can write

$$d^{T-1}\bar{\phi}_{k,l}(d) \leq 2kmd^{T-1}\exp\{-\frac{1}{2}\bar{h}(d)^2\}$$

for all $d > \bar{d}$ for some $\bar{d} > 0$. The remaining of the proof is omitted as it is identical to the proof of proposition 4.1. $\square$

Conley's (1999) estimator

Conley (1999) provides an estimator when $\mathcal{X} \subset \mathbb{R}^2$ and $\{Z_{s,m}; s \in S_m, m \in \mathbb{N}\}$ is α-mixing and stationary. This approach has also been recently used by Wang et al. (2010) in the context of the estimation of a spatial probit. We propose to extend Conley's (1999) estimator for $\mathcal{X} \subset \mathbb{R}^T$, where $T \geq 1$.

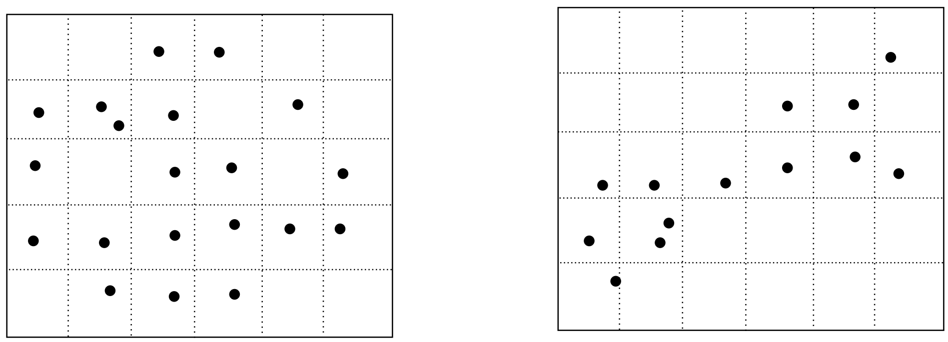
We consider a compact subset of the space of individual characteristics, i.e. $Y \subset \mathcal{X}$. We define a random process Λ on a regular lattice on Y such that $\Lambda_y = 1$ if the location $y = (y_1, ..., y_T)$ is sampled, and $\Lambda_y = 0$ otherwise. We assume that Λ is independent of the underlying random field, has finite expectation, and is stationary. Intuitively, since the lattice is regular, it gives an idea of the dependence structure between the observations. Consider Figure 1 below, where

$\mathcal{X} = \mathbb{R}^2$ for presentation purposes. Sampled pairs are represented by the black circles.

Figure 3: Regular Lattice and Dependence Structure

(a) Uniform Dependence Structure

(b) Directed Dependence Structure



In Figure 1a, sites are distributed more or less uniformly in Y. In Figure 1b however, the dependence structure seems to be more directed. Now, lets define $\bar{y} = (\bar{y}_1, ..., \bar{y}_T)$ to be the maximal location for Λ_y in every dimension. Notice that this quantity is well defined since Y is compact. For instance, for the lattice in Figure 1, $\bar{y} = (6, 5)$.

Now, let $\hat{q}_y(\theta) = \frac{1}{n(y)} \sum_{s \in y} q_{s,m}(\theta)$, where $s \in y$ is a sampled pairs s in location y, and $n(y)$ is the number of sampled pairs in location y. We define the following process, for any location y:

$$R_y(\theta) = \begin{cases} \frac{\partial \hat{q}_y}{\partial \theta}(\theta) & \text{if } \Lambda_y = 1 \\ 0 & \text{otherwise} \end{cases}$$

Let m^* be the number of sampled locations.[28] We can now present our proposed estimator, based on a generalization of Conley (1999):

$$\begin{aligned} B_m(\theta) &= \frac{1}{m^*} \sum_{y_1=0}^{\tilde{y}_1} ... \sum_{y_T=0}^{\tilde{y}_T} \sum_{y'_1=y_1+1}^{\bar{y}_1} ... \sum_{y'_T=y_T+1}^{\bar{y}_T} \Gamma_{\tilde{y}}(y) \left[R_{y'}(\theta) R'_{y'-y}(\theta) + R_{y'-y}(\theta) R'_{y'}(\theta) \right] \\ &- \frac{1}{m^*} \sum_{y_1=1}^{\bar{y}_1} ... \sum_{y_T=1}^{\bar{y}_T} R_y(\theta) R'_y(\theta) \end{aligned} \tag{5}$$

Where $\tilde{y} < \bar{y}$, and $\Gamma_{\tilde{y}}(y)$ is a kernel function. For instance, Conley (1999) proposed to use $\tilde{y} = o(\bar{y}^{1/3})$, i.e. a bound of the same order as the cubic root of $\bar{y}$, and the following Bartlett

[28]A simple way to compute m^* is to count the number of times $\Lambda_y = 1$.

window kernel:

$$\Gamma_{\tilde{y}}(y) = \begin{cases} (1-\frac{|y_1|}{\tilde{y}_1})...(1-\frac{|y_T|}{\tilde{y}_T}) & \text{for } |y_1| < \tilde{y}_1, ..., |y_T| < \tilde{y}_T \\ 0 & \text{otherwise} \end{cases}$$

As in the estimation of HAC variances, the precise choice of $\tilde{y}$ and $\Gamma_{\tilde{y}}(y)$ will depend on the specific application. With that regard, we can easily show that the estimator in (5) when $T = 1$ is equivalent to a HAC estimator.

Lets rewrite the estimator for $T = 1$:

$$\begin{aligned} B_m(\theta) &= \frac{2}{m}\sum_{k=0}^{\tilde{y}_1}\sum_{y=k+1}^{\bar{y}_1}\Gamma_{\tilde{y}}(k)R_y(\theta)R'_{y-k}(\theta) - \frac{1}{m}\sum_{y=1}^{\bar{y}_1}R_y(\theta)R'_y(\theta) \\ &= \hat{\gamma}(0) + 2\sum_{k=1}^{\tilde{y}_1}\Gamma_{\tilde{y}}(k)\hat{\gamma}(k) \end{aligned}$$

where $\hat{\gamma}(0) = \frac{1}{m}\sum_{y=0}^{\bar{y}_1} R_y(\theta)R'_y(\theta)$ is the estimation of the variance of the process R_y, and $\hat{\gamma}(k) = \frac{1}{m}\sum_{y=k+1}^{\bar{y}_1} R_y(\theta)R'_{y-k}(\theta)$ the estimation of the autocovariance of the process R_y. Then, in one dimension our proposed estimator become exactly the HAC variance estimator for the covariance stationary process R_y, using the Bartlett kernel. In our case here, under some ϕ mixing conditions we may ensure that $\gamma(k) \to 0$ as $k \to \infty$.

Bester et al. (2012)

Let $\mathcal{X}$ be partitioned into groups, or clusters: $c = 1, ..., C$. Bester et al. (2012) propose to use the following CV estimator:

$$\hat{B}_m(\theta) = \frac{1}{m}\sum_{s\in S}\sum_{r\in S}\mathbb{I}(c_s = c_r)\frac{\partial q_{s,m}(\theta)}{\partial\theta}\left(\frac{\partial q_{s,m}(\theta)}{\partial\theta}\right)'$$

Where c_s is the group in which $s \in S$ is located. This is the usual Cluster-Variance estimator. It has the advantage of being easy and fast to implement. In practice, the constructions of those groups is not necessarily straightforward. Bester et al. (2012) recommend to use a relatively small number of large groups. An important requirement however is a boundary condition which states that most of the pairs in groups are located in the interior (i.e. not on the boundary) of those groups in $\mathcal{X}$. Specifically, let $\partial(c_s)$ be the boundary of the group c_s, and $\bar{c}_m$ be the average number of pairs in a group, then one should have $\partial(c_s) < \bar{c}_m^{(T-1)/T}$, where $T \geq 1$ is the dimension of $\mathcal{X}$.

Constructed variables

$\Delta(Phys)$ and $\Delta(Psys)$: Those variables are constructed from the interviewer's subjective appreciation of the teenager's physical and personality attractiveness. Specifically: "How physically attractive is the respondent?" and "How attractive is the respondent's personality?", both ranging from 1 (very unattractive) to 5 (very attractive).

$\Delta(Geo)$: The Add Health database provides information on (normalized) geographical location of the individuals' residences within communities. Following the survey's definition, friendships are restricted within communities. (Individuals can be friends only if they belong to the same communities.) Then for individuals from the same community, we set the geographical distance between them according to the variable, and we normalized the distance between individuals from different communities to 100. Notice that this implies that Asymptotic Homophily is trivially respected by construction of the database.