My Friend Far, Far Away
A Random Field Approach to Exponential Random Graph Models

Vincent Boucher* and Ismael Mourifié**

First version: November 2012
This version: April 2017

Abstract

We explore the asymptotic properties of strategic models of network formation in very large populations. Specifically, we focus on (undirected) exponential random graph models (ERGMs). We want to recover a set of parameters from the individuals' utility functions using the observation of a *single*, but large, social network. We show that under some conditions, a simple logit-based estimator is coherent, consistent and asymptotically normally distributed under a weak version of homophily. The approach is compelling as the computing time is minimal and the estimator can be easily implemented using pre-programmed estimators available in most statistical packages. We provide an application of our method using the Add Health database.

JEL Codes: C13, C21, D85
Keywords: Network Formation, Random Fields, Spatial Econometrics, Homophily.

* Economics Department, Université Laval and CIRPÉE, vincent.boucher@ecn.ulaval.ca
** Economics Department, University of Toronto, ismael.mourifie@utoronto.ca

Acknowledgements
This paper benefited from many highly interesting discussions. In particular, we want to thank Angelo Mele for his insightful comments on earlier versions of the paper. We also want to thank Yann Bramoullé, Diego Cerdeiro, Arun Chandrasekhar, Tim Conley, Sílvia Gonçalves, Bryan S. Graham, Paul Grieco, Marc Henry, Ilze Kalnina, David Karp, Matthew Lindquist, Yao Luo, Benoit Perron, Joris Pinkse and Eirini Tatsi.

We want in particular to thank Jaap Abbring, Aureo de Paula and two anonymous referees for their numerous and extremely insightful comments and suggestions, which have helped us improving the paper in many important respects.

We would also like to thank the participants at many conferences and seminars, including the SCSE2012, the economics department at Penn State, the economics department at Université de Montreal, the CeMMAP-CIREQ-SciencesPo-X workshop on "The Estimation of Complementarities in Matching and Social Networks.", and the SOFI at Stockholm University. An earlier version of the paper circulated under the title "My Friend Far Far Away. Asymptotic Properties of Pairwise Stable Networks" and provided an interpretation of ERGMs in terms of pairwise stability.

Vincent Boucher thanks the economics faculty and the Cambridge-INET at the University of Cambridge for their support and for many interesting discussions, as well as financial support from FRQSC and SSHRC. A significant part of this work was done while Ismael Mourifié was visiting the economics department at Penn State. Many thanks for their support and hospitality.

This research uses data from Add Health, a program directed by Kathleen Mullan Harris and designed by Harris, J. Richard Udry and Peter S. Bearman 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 their assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (`www.cpc.unc.edu/addhealth`). No direct support was received from grant P01-HD31921 for this analysis.

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 & Fryer Jr (2007), Watts (2007), and Mele (2016)). This raises many interesting questions regarding the cause of this segregation. For instance, we would like to be able to distinguish between the impact of individuals' characteristics (e.g. race) and the impact of individuals' positions in the networks (e.g. popularity).

The shape of social networks also have measurable effects on individuals' choices. Many studies show that an individual's peers can have a significant influence on his actions, ranging from unhealthy consumption choices (e.g. Fortin & Yazbeck (2015) and the references therein) to labor-force participation (e.g. Rolfe et al. (2013), Patacchini & Zenou (2012) and Mourifié & Siow (2014)).[1] However, since most social networks are endogenously formed, the estimated influence of peers is likely to be biased.[2] Understanding how the networks are formed could allow one to control for this endogeneity and suggest policy instruments that would help influence network formation processes.

In this paper, we use Mele (2016)'s network formation model, for which the joint distribution is an exponential random graph model (ERGM). The estimation of ERGMs is typically challenging due to the untractable denominator of the joint distribution. We use random field theory to provide a simple pseudo maximum likelihood estimator (as in Besag (1975) and Strauss & Ikeda (1990)), based on the product of the conditional distributions. To our knowledge, this paper is the first to apply random field theory to empirical models of network formation. The approach is promising and has been recently employed by Leung (2014) in a similar context. Our approach only requires the observation of a

[1] See De Paula et al. (2016) and Boucher & Fortin (2016) for recent reviews.

[2] The literature on peer effects has recently explicitly considered the endogeneity of social networks. See, for example, Badev (2013), Boucher (2016), Goldsmith-Pinkham & Imbens (2013) and Hsieh & Lee (2016).

single, potentially very large, network.

Our model allows for a large set of admissible preferences, which are characterized by intuitive conditions. Specifically, we show that our 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 characteristics of social networks. It formalizes the observation that similar individuals are more likely to interact with each other. As homophily is featured in both theoretical (e.g. Boucher (2015), Bramoullé et al. (2012), and Currarini et al. (2009, 2010)), and empirical (e.g. Mele (2016), 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 American teenagers' friendship networks.

A fundamental challenge in estimating a network formation process is the highly dependent nature of most socio-economic relationships. Consider a friendship: The probability that Alice and Bob are friends depends on their individual characteristics. However, it may also depend on Bob's friendship with Charlotte (who perhaps does not like Alice). The probability that Alice and Bob are friends may then depend on Charlotte's individual characteristics (as well as on her other friendships).

If individuals have homophilic preferences, the probability that Alice and Bob are friends should be primarily influenced by individuals similar to them. If Alice and Bob 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 Bob and Charlotte are friends, there is a greater probability that they go to the same school, or at least live in the same country. Donald, a elderly man, living in a different country (hence having very different individual characteristics than Alice, Bob or Charlotte) probably does not influence the probability that Alice and Bob 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.[3] Some papers require the observation of many (mostly independent) social networks (e.g. Boucher (2015), Currarini et al. (2009, 2010) and Sheng (2014)). One limit of these approaches, however, is that they require the observation of many (an asymptotically infinite number of) independent social networks, which is not always available in existing databases. As the data requirement is large, the computational burden is also important.

Accordingly, many recent approaches only require the observation of a single network, at one point in time (e.g. Graham (2016a) Leung (2014, 2015) and de Paula et al. (2016)). In particular, a significant part of that literature focusses on exponential random graph models (ERGMs, which we formally define in the next section). Chandrasekhar & Jackson (2014, 2015), as well as Mele (2016), followed by Badev (2013), present strategic models of networks formation leading to ERGMs, and focus on Bayesian inference.[4]

Chandrasekhar & Jackson (2014, 2015) focus on asymptotic theory and present conditions on the structure of the network which allow for a consistent estimation of the parameters. By focussing on subgraphs, and under the assumption that the network is sparse, they develop a simple estimator, based on the iterative counting of subgraphs.

Alternatively, Graham (2016a) and Graham (2016b) focus on logit-based models allowing for individual fixed unobserved characteristics. A novative feature in Graham (2016b) is the focus on a dynamic setting where links can be dependent on the previous state of the network. He shows that conditional likelihood, for links that are stable across time, has a tractable denominator.

We contribute to this literature by providing an explicit, easy-to-implement Pseudo-MLE that requires 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 Pseudo-MLE is consistent and asymptotically normally dis-

[3]See Chandrasekhar (2016) and for an excellent recent review.
[4]See also Mele & Zhu (2016).

tributed. To do so, we use the law of large numbers and central limit theorem, following Jenish & Prucha (2009).

The rest of the paper is organized as follows. In Section 2, we present our microeconomic and econometric framework. In Section 3, we review some elements of random fields theory. In Section 4, we derive the asymptotic distribution of our estimator, and in Section 5, we define a class of network formation models suited to our econometric framework. In section 6, we discuss the identification of the model. In Section 7, we provide an application using the Add Health database. We conclude in Section 8.

2. The economy

Let $N = \{1, ..., n\}$ be the set of individuals. Each individual is characterized by a non-stochastic vector of quantitative characteristics $x_i = (x_i^1, ..., x_i^T) \in \mathcal{X} \subset \mathbb{R}^T$, where $T \geq 1$. We define the distance between two individuals as $d(i, j) = d(x_i, x_j)$, where d is a distance on $\mathbb{R}^T$.[5]

In general, the choice of this distance function will be context-dependent (see Example 1 and our empirical application in Section 7). In particular, the distance can represent spatial preferences of individuals.[6] We use $X = (x_1, ..., x_n) \in \mathcal{X}^n$ to denote the matrix of individual characteristics.

We assume that individuals interact in an *undirected* social network. Let $m = \frac{n(n-1)}{2}$ be the number of possible *unordered* pairs of individuals (i, j) for $i \neq j$ in the economy. The set of all pairs will be denoted by S_m, with typical elements $s, r \in S_m$. For any pair (i, j), we let $w_{ij} = 1$ if $i \in N$ and $j \in N$ are linked by a socio-economic relationship (e.g. friendship), and $w_{ij} = 0$ otherwise.

The collection of all relationships, the network, is noted $W_m = \{w_s\}_{s \in S_m}$. For any subset of pairs $\hat{S} \subset S_m$, we note the collection of relationships in $\hat{S}$ as $W_{\hat{S}} = \{w_s\}_{s \in \hat{S}}$. Finally, the set of relationships in $S_m \setminus \{s\}$ is noted by $W_{m,-s}$.

[5] The model can be easily extended to include stochastic and qualitative characteristics, as long as they do not affect the distance. This is discussed in Section 7.

[6] See Henry & Mourifié (2013) for a discussion of spatial preferences in Euclidean space.

We assume that the network W_m is endogenous and determined as a function of individuals' utilities. An individual has preferences over the set of characteristics and the network structure in the economy: $u_i(W_m;\theta)$, where $\theta \in (\theta^1, ..., \theta^K) \in \Theta$ is the set of parameters to be estimated. We assume that Θ is a compact subset of $\mathbb{R}^K$, for $K \geq 1$. Here, the utility function is best interpreted in non-transferable utility setting in the sense that we do not explicitly model transfers between individuals. Assumption 1 resumes the structure of an individual's preferences.

Assumption 1 (Preferences). *For any $i, j \in N$, there exists a function $h_{ij}(W_{m,-ij};\theta)$, three times continuously differentiable in θ, such that:*

$$\begin{aligned} u_i(1, W_{m-ij};\theta) - u_i(0, W_{m-ij};\theta) &= u_j(1, W_{m-ij};\theta) - u_j(0, W_{m-ij};\theta) \\ &= h_{ij}(W_{m,-ij};\theta). \end{aligned}$$

Note that the function $h_{ij}(W_{m,-ij};\theta)$ can be interpreted as the value of a socio-economic relationship between i and j and is assumed to be symmetric with respect to its indices. The payoffs h_{ij} can be best interpreted as the payoff of a local public good, equally valued by i and j. This assumption is imposed for identifiability issues since we focus on undirected network.[7]

We will also make the following assumption:

Assumption 2 (Preferences). *For all pair (i, j), we assume that there exist $\chi_{ij}(\theta), \{\tilde{\chi}_{ij,ik}(\theta)\}_k$ and $\{\tilde{\chi}_{ij,jk}(\theta)\}_k$, symmetric with respect to their indices, such that:*

$$h_{ij}(W_{m,-ij}, \theta) = \chi_{ij}(\theta) + \sum_{k \neq i,j} \tilde{\chi}_{ij,ik}(\theta) w_{ik} + \tilde{\chi}_{ij,jk}(\theta) w_{jk},$$

where $\tilde{\chi}_{ij,ik}(\theta) = 0$ whenever $\max\{d(i,j), d(i,k)\} > \bar{d}$ where $\bar{d} > 0$ is an arbitrarily large (but finite) number, and similarly for $\tilde{\chi}_{ij,jk}(\theta)$.

The assumption that $\tilde{\chi}_{ij,ik}(\theta) = 0$ whenever $\max\{d(i,j), d(i,k)\} > \bar{d}$ implies that the individuals cannot receive positive payoffs from an infinite number of

[7] For undirected networks, one only observes if a link, say between i and j, is created or not. If no link is observed, it is therefore impossible to identify if i or j (or both) "refused the link". Then, if the payoffs were non-symmetric, one can only identify $\min\{h_{ij}, h_{ji}\}$.

links and ensures that $h_{ij}(W_{m,-ij},\theta) < \infty$.[8] Note that this is weaker than assuming that individuals only have a finite number of links, as for instance in de Paula et al. (2016). Indeed, at this point, individuals may have an infinite number of links, however, only a finite number of those links may affect the payoff of any particular link. We will discuss the implications of this assumption in detail in section 6.2.

The specific structure in assumption 2 serves two purposes. First, following Mele (2016), it guarantees the existence of a *potential function* (Monderer & Shapley, 1996), i.e. a function $Q(W_m)$ such that for all $i,j \in N$, $Q(1,W_{m,-ij};\theta) - Q(0,W_{m,-ij};\theta) = h_{ij}(W_{m,-ij};\theta)$. The potential function will be developed explicitly in proposition 1 in section 3.

Second, assumption 2 allows for the definition of an estimator based on the product of the marginal distributions in the spirit of Besag (1975) and Strauss & Ikeda (1990). A formal treatment is provided in section 3. Example 1 illustrates.

Example 1. *We will use the following as a running example:*

$$h_{ij}(W_{m,-ij};\theta) = -\beta d(i,j) + \sum_{k\neq i,j} \alpha_{ik}w_{ik} + \alpha_{jk}w_{jk}$$

where $\alpha_{ik} = \alpha$ if $d(i,k) < \bar{d}$ and 0 otherwise, and where $\theta \equiv (\alpha,\beta)$. Note that here, we assume that x_i is unidimensional so that $d(i,j) = |x_i - x_j|$. We can also write:

$$h_{ij}(W_{m,-ij};\theta) = \Big\{\alpha n_i(W_{m,-ij}) + \alpha n_j(W_{m,-ij}) - \beta|x_i - x_j|\Big\},$$

where $n_i(W_{m,-ij})$ and $n_j(W_{m,-ij})$ are bounded and represent the number of links of i and j in the network, excluding the potential link between i and j. Then, if $\alpha > 0$, as an individual creates links, it increases the incentive to create additional links.

One can verify that the implied utility function is:

$$u_i(W_m;\theta) = \sum_{j\neq i} -w_{ij}\beta|x_i - x_j| + \sum_{j\neq i}\sum_{k<j,k\neq i} \alpha_{ik}w_{ij}w_{ik} + \alpha_{jk}w_{ij}w_{jk}.$$

since $\alpha_{ij} = \alpha_{ik}$ for all $j,k \neq i$ under assumption 2.

[8] This is true under assumption 5, presented in section 4. We thank an anonymous referee for this pointing this out to us.

Then, i receives direct cost (if $\beta > 0$) of linking equal to $\beta|x_i - x_j|$, but benefits (if $\alpha > 0$) from positive spillover form his other links (i.e. $\alpha w_{ij} w_{ik}$), as well as from j's other links (i.e. $\alpha w_{ij} w_{ik}$).

In other words, in this example, the value of a link between i and j decreases in the distance between them, but increases with the number of friends they have and can be interpreted as preferences for popularity (see Mele (2016)).

Moreover, note that assumption 2 holds, since we can define $\chi_{ij}(\theta) = -\beta|x_i - x_j|$ and $\tilde{\chi}_{ij,ik}(\theta) = \alpha_{ik}$ and $\tilde{\chi}_{ij,jk}(\theta) = \alpha_{jk}$. ■

The link formation process follows directly from Mele (2016) where he presents an intuitive stochastic meeting process and assumes that the data is drawn from its stationary distribution.[9] Starting from any network W_m^0, the meeting process goes as follows for $t = 1, ..., \infty$:[10]

1. Two individuals i and j meet with probability $\lambda(ij; W_{m,-ij}^{t-1}) > 0$;
2. Given such meeting, preferences are affected by a temporary, idiosyncratic shock ε_{ij}^t that affects the value of a link between i and j. The shock ε_{ij}^t is assumed to follow a standardized logistic distribution;[11]
3. Individuals i and j decide to update the status of their socio-economic relationship, letting $w_{ij}^t = 1$ iff $h_{ij}(W_{m,-ij}^{t-1}; \theta) + \varepsilon_{ij}^t \geq 0$.

Note that since preferences and shocks are symmetric (i.e. $h_{ij} = h_{ji}$ and $\varepsilon_{ij}^t = \varepsilon_{ji}^t$), both individuals must agree in step 3. Here, the ε_{ij}^t are temporary idiosyncratic shocks on the value of friendships and can be interpreted as temporary variations of the value of the local public good (i.e. friendship) produced by i and j.

Mele (2016) showed that this process converges to a stationary distribution, which can be written as:

$$\mathbb{P}(W_m|\theta) = \frac{\exp\{Q(W_m; \theta)\}}{\sum_{W'_m} \exp\{Q(W'_m; \theta)\}} \tag{1}$$

[9] See also Badev (2013) for an equilibrium interpretation.

[10] The meeting process is in fact slightly adapted to our setting and notation, assuming a non-directed network, under assumption 1. Note also that the process can be seen as a logit response game (Blume, 1993) on a game where each pair acts as a player.

[11] Note that this is equivalent to assuming that there are iid shock $(\varepsilon_{ij}^0, \varepsilon_{ij}^1)$ distributed according to a type 1 extreme value distribution and such that $u_i(W_m; \theta) + w_{ij}\varepsilon_{ij}^1 + (1 - w_{ij})\varepsilon_{ij}^0$, and similarly for j.

where $Q(W_m)$ is the potential function.

The denominator in (1) is intractable as it sums over the set of all possible network, which has a cardinality of 2^m. This challenge has been the focus of most of the recent literature on empirical network formation models (see Chandrasekhar (2016) for a discussion). We will discuss how we will deal with this intractability in the next section.

3. (Markov) Random Fields

In this section, we briefly introduce some basic definitions and results from the theory of random fields.[12]

Definition 1. *The random field $\{w_{s,m}; x_s \in \mathcal{X}_{S_m}, m \in \mathbb{N}\}$ 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.*

In essence, a random field is a list of random numbers whose indices are mapped onto a space (here, of T dimensions). It can also be viewed as a generalization of a stochastic process. For a stochastic process, observations are located in "time." Here, observations (i.e. pairs) are located in $\mathcal{X}$ (see Besag (1974)).[13]

In order to describe the dependence structure of the random field, it is convenient to define a *neighbourhood*, for each pairs. Formally, a neighbourhood in a random field is defined as follows:

Definition 2. *A neighbourhood $N_s \subset S_m \setminus \{s\}$ of a pair $s \in S_m$ is defined as the smallest collection of pairs such that*

$$\mathbb{P}(w_s | W_{m,-s}) = \mathbb{P}(w_s | W_{N_s}).$$

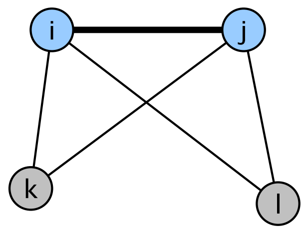
Formally, assumption 2 allows for a natural definition of neighbourhoods on $\{w_{s,m}; x_s \in \mathcal{X}_{S_m}, m \in \mathbb{N}\}$. Define neighboured pairs as those that share an individual, or mathematically: $N_s = \{r \in S_m \setminus \{s\} : r_1 \in s \text{ or } r_2 \in s\}$, as illustrated in Figure 1.[14]

[12] Most of the results presented are rarely treated in econometrics textbooks, but are standard in statistical mechanics and spatial statistics. We refer the interested reader to Cressie (2015) and Koller & Friedman (2009) for additional readings.

[13] The position of the pairs in $\mathcal{X}$ is formally described in the next section.

[14] This is also the definition used by Graham (2016b).

Figure 1: A Pair's Neighbours



Consider the set of all pairs: $S_m = \{(i,j),(i,k),(i,l),(j,k),(j,l),(k,l)\}$. The neighbours of the pair (i,j) are the pairs $\{(i,k),(j,k),(i,l),(j,l)\}$, but not the pair (k,l) since (i,j) and (k,l) have no individuals in common. Then, $N_{(i,j)} = \{(i,k),(j,k),(i,l),(j,l)\}$.

Then, under Assumption 2, we can therefore have:

$$\mathbb{P}(w_{ij}|W_{m,-ij}) = \mathbb{P}(w_{ij}|W_{N_{ij}}).$$

where $N_{ij} = \{ik, jk\}_{k\neq i,j}$.

A random field, joint with the collection of neighbourhoods, is called a *(local) Markov random field*.

Now that we have laid down the basic definitions, we can go back to the intractability of the denominator in equation (1). This intractability makes the use of the MLE approach, based on the maximization of $\ln \mathbb{P}(W_m|\theta)$ with respect to θ, very difficult.

In this paper, we follow the approach of Besag (1975) and Strauss & Ikeda (1990), and present an estimator based on the *conditional* probabilities, i.e. $\mathbb{P}(w_s|W_{m,-s};\theta)$ which would be more *tractable*.

Indeed, Hammersley & Clifford (1971) and Besag (1974) give necessary and sufficient conditions on (1) under which all the conditional distributions are (i) valid, in the sense of being compatible with the joint distribution $\mathbb{P}(W_m|\theta)$ and (ii) tractable, in the sense that they do not depend on the intractable denominator presents in equation (1).

The following result is direct application of Hammersley & Clifford (1971) and Besag (1974) recast into our framework:

Proposition 1. *Suppose that Assumptions 1 and 2 hold, then the potential function can be written as:*

$$Q(W_m;\theta) = \sum_s \chi_s(\theta) w_s + \sum_s \sum_{r<s:r\in N_s} \chi_{r,s}(\theta) w_r w_s, \tag{2}$$

where $r < s$ is for an arbitrary ordering of the pairs, and the conditional probabilities are such that:

$$\mathbb{P}(w_s|W_{m,-s};\theta) = \frac{\exp\{w_s h_s(W_{m,-s},\theta)\}}{1+\exp\{h_s(W_{m,-s},\theta)\}} \tag{3}$$

See also Lee et al. (2001) for a more recent treatment. Example 2 illustrates.

Example 2 (Example 1 continued). *We have:*

$$h_s(W_{m,-s};\theta) = -\beta|x_{s_1} - x_{s_2}| + \sum_{r\in N_s} \alpha w_r.$$

The potential function is therefore:

$$Q(W_m;\theta) = \sum_s -w_s\beta|x_{s_1} - x_{s_2}| + \sum_s \sum_{r<s:r\in N_s} \alpha w_s w_r.$$

■

Proposition 1 give us the true conditional probabilities which do not depend on the intractable denominator.

We can therefore define the following pseudo likelihood (PL) function, in the spirit of Besag (1975) and Strauss & Ikeda (1990):

$$PL \quad = \quad \sum_s \Big(w_s h_s(W_{m,-s};\theta) - \ln[1+\exp\{h_s(W_{m,-s};\theta)\}] \Big) \tag{4}$$

which is not the true log-likelihood function $\ln \mathbb{P}(W_m|\theta)$, unless we have full independence over the random field.

Since the conditional probabilities are well specified, we may expect the maximum of the the PL to be a consistent estimator of the true parameter θ_0 given that it is unique. Besag (1975) provides an heuristic argument in that direction. However, a formal proof of the general case is far from trivial as it requires the use of limits theorems that apply to dependent random fields.

In what follows, we use recent results from Jenish and Prucha (2009), specifically central limit theorems and uniform laws of large numbers in order to show under which conditions the consistency and asymptotic normality of the estimator obtained by maximizing the the PL in (4). Following, Jenish and Prucha (2009) we assume an *increasing domain* asymptotic framework.

4. Limited dependence theorems

In this section, we present two theorems for dependent observations. We show that under ϕ-mixing, the true value $\theta_0 \in \Theta$ can be consistently estimated using the simple PL in (4). These theorems are useful since, as we show in Section 4, there exist simple conditions on h_{ij} that imply ϕ-mixing.[15]

To clarify the exposition, we use the following simplifying notation:

$$\begin{aligned} q_{s,m}(w_{s,m}|W_{m,-s},\theta) &= w_{s,m}h_s(W_{m,-s};\theta) \\ &\quad - \ln[1+\exp\{h_s(W_{m,-s};\theta)\}] \end{aligned}$$

We also use $q_{s,m}(\theta) = q_{s,m}(w_{s,m}|W_{m,-s},\theta)$ when there is no ambiguity. Maximizing the PL is equivalent to maximize:

$$\mathcal{L}_m(\theta) = \kappa(m)^{-1} \sum_{s\in S_m} q_{s,m}(w_{s,m}|W_{m,-s},\theta). \tag{5}$$

where $\kappa(m)$ is a scaling parameter which determines the rate of convergence of our estimator. In general $\kappa(m)$ will be related to the structure of the network which depends on the specification of the payoff functions h_{ij}. $\kappa(m)$ should carefully be chosen to avoid that $\lim_{m\to\infty} \mathbb{E}\mathcal{L}_m(\theta)$ is infinite (i.e., $+/-\infty$) or a trivial function of θ, e.g $\lim_{m\to\infty} \mathbb{E}\mathcal{L}_m(\theta) = 0$. Otherwise, our identification condition (see assumption 4 below) will be trivially violated.[16] In section 5, we provide conditions on h_{ij} such that $\kappa(m) = O(\sqrt{m})$.

[15] Our results can easily be adapted to other mixing definitions, such as α-mixing, as was subsequently done by Leung (2014).

[16] We thank an anonymous referee for pointing this important issue to us.

We define the position of a pair s in $\mathcal{X}$ as the average point between the positions of s_1 and s_2 in $\mathcal{X}$, i.e. $x_s \in \mathcal{X}$ such that $x_s = \frac{x_{s_1}+x_{s_2}}{2}$.[17] Using this definition, the distance between two pairs r and s is equal to $d(s,r) = d(x_r, x_s) = d(\frac{x_{s_1}+x_{s_2}}{2}, \frac{x_{r_1}+x_{r_2}}{2})$. We also define $\mathcal{X}_{S_m} \equiv \{x_s : s \in S_m\}$ as the set of pairs' positions in $\mathcal{X}$.

We now turn to the dependence structure of the random field $\{q_{s,m}(\theta); x_s \in \mathcal{X}_{S_m}, m \in \mathbb{N}\}$, defined on the probability space $(\Omega_q, \mathcal{F}_q, \mathbb{P}_q)$, where $\Omega_q = \mathbb{R}$, with associated σ-algebra $\mathcal{F}_q$ and probability measure $\mathbb{P}_q$. 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

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

As discussed in Section 3, this is analagous to standard time-series models. In a stochastic 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.[18] We want to apply the same basic approach when the dependence between A and B goes through $\mathcal{X}$. 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 the law of large numbers and the central limit theorem for dependent observations on random fields developed by Jenish & Prucha (2009) (Theorems 1, 2 and 3). We introduce the following definition:

Definition 3. *For $U \subset \mathcal{X}_{S_m}$ and $V \subset \mathcal{X}_{S_m}$, let $\mathcal{A}_m^U \equiv \sigma(q_{s,m} : x_s \in U\}$ and $\mathcal{B}_m^V \equiv \sigma(q_{s,m} : x_s \in V\}$, the corresponding σ-algebra. Also let $|U|$ and $|V|$ denote the number of pairs located in U and V. We define the $\phi-mixing$ for the random field $\{q_{s,m}(\theta); s \in S_m, m \in \mathbb{N}\}$ with the function:*

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

where $d(U,V) = \inf\{d(s,r) : x_s \in U \text{ and } x_r \in V\}$

[17] 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}.

[18] For example, see White (2001).

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

Assumption 3 (ϕ-mixing).
(3.1) $\bar{\phi}_{1,1}^{1/2}(d) = O(d^{-T+1-\epsilon})$
(3.2) $\bar{\phi}_{k,l}(d) = O(d^{-T+1-\epsilon})$ *for* $k + l \leq 4$
(3.3) $\bar{\phi}_{1,\infty}(d) = O(d^{-T-\epsilon})$ *for some* $\epsilon > 0$.

Recall that $T \geq 1$ is the dimension of $\mathcal{X}$. In words, Assumption 3 implies that $\bar{\phi}_{k,l}(d)$ has to converge *fast enough* to 0. In Section 5, we provide sufficient conditions on the primitive equations of the model under which Assumption 3 holds. For now, we show that the estimation technique is valid as long as 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 additional assumptions:

Assumption 4 (Identification). *There exists a unique* $\theta_0 \in int\ \Theta$ *maximizing* $\lim_{m\to\infty} \mathbb{E}[\mathcal{L}_m(\theta)]$.

This identification condition is discussed in detail in Pötscher & Prucha (1991). Whether or not that assumption holds will strongly depend on the specification of h_i^j; researchers should make sure that it holds in practice. This assumption is discussed in detail in section 6.

The next assumption describes the asymptotic behaviour of the pairs in $\mathcal{X}$.

Assumption 5 (Increasing Domain). *For all* $s_1, s_2 \in N$, $d(s_1, s_2) \geq d_0$ *for some* $d_0 > 0$ *(wlog* $d_0 = 1$*)*

This implies that there exists a minimal distance (in $\mathcal{X}$) between pairs of individuals (and therefore, between individuals). In essence, Assumption 5 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 $\mathcal{X}$ that contains all the pairs of individuals (i.e. $\mathcal{X}_{S_m}$) has to expand as the number of individuals (and therefore pairs) increases. It is worth noting that we require only one observable characteristic for which there exists a minimum distance. For example, one could be interested in friendships between residential neighbours. In that case, there must exist a minimum distance between any neighbouring houses. Accordingly, if the number of households increases, it implies that the residential area increases as well. We will come back on this assumption in the context of our empirical application in Section 7.

Finally, we also need the following standard moment conditions for the payoff functions:

Assumption 6 (Technical conditions I).
(6.1) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} |q_{s,m}(\theta)|^{(1+\eta)}] < \infty$ *for some* $\eta > 0$.
(6.2) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} |\frac{\partial q_{s,m}(\theta)}{\partial\theta}|] < \infty$.

Given the previous assumptions, we have the following:

Theorem 1 (Consistency). *Suppose that Assumptions 4, 5 and 6 hold, and that Assumption 3.2 is respected for* $k = l = 1$. *Then, the estimator* $\hat{\theta} = arg\max_{\theta\in\Theta} \mathcal{L}_m(\theta)$ *converges to* θ_0 *as* $m \to \infty$.

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

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

Since the asymptotic normality of the estimator requires more structure than the one needed for consistency, we need Assumptions 3.1 to 3.3, as well as the following additional technical conditions.

Assumption 7 (Technical conditions II).
(7.1) $B_0(\theta_0) > 0$.
(7.2) $D_0(\theta_0)$ *is invertible.*
(7.3) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} \|D_{s,m}(\theta)\|^{1+\eta}] < \infty$ *for some* $\eta > 0$.
(7.4) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} \|\frac{\partial D_{s,m}(\theta)}{\partial\theta}\|] < \infty$.
(7.5) $\sup_m \sup_s \mathbb{E}[\sup_{\theta\in\Theta} |\frac{\partial q_{s,m}(\theta)}{\partial\theta}|^2] < \infty$
where $D_{m,s}(\theta) = \frac{\partial^2 q_{s,m}(\theta)}{\partial\theta\partial\theta'}$.

These assumptions are standard and are sufficient to show the asymptotic normality of our estimator.[19]

[19]Formally, the proof of Theorem 2 derives the limit distribution for $\sqrt{\kappa(m)}(\hat{\theta} - \theta_0)$. We report the asymptotic distribution of $\hat{\theta}$ for presentation purposes.

Theorem 2 (Asymptotic normality). *Let $m \to \infty$. Under Assumptions 3,4, 5, 6 and 7, the estimator $\hat{\theta} = arg\max_{\theta \in \Theta} \mathcal{L}_m(\theta)$ is normally distributed with its variance-covariance matrix given by $D_0^{-1} B_0 D_0^{-1} / \kappa(m)$.*

The variance-covariance matrix in our setting is analogous a heteroskedasticity and autocorrelation consistent (HAC) variance-covariance matrix in a time-series setting. The estimation of these variances is not straightforward. The estimation of $D_0(\theta_0)$ follows from Theorem 1, since $D_0(\theta)$ has the same dependence structure as $\lim_{m\to\infty} \mathbb{E}\mathcal{L}_m(\theta)$. A consistent estimator then is $D_m(\hat{\theta}) = \kappa(m)^{-1} \sum_{s=1}^{m} D_{s,m}(\hat{\theta})$.

Defining a consistent estimator for $B_0(\theta_0)$ is more challenging. We suggest two approaches to estimate $B_0(\theta_0)$. The first is based on a generalization of standard HAC estimators (Conley (1999)). Although valid, the estimator proposed by Conley (1999) can be very computationally intensive when the number of dimensions of $\mathcal{X}$ increases (say, $T \geq 4$). Bester et al. (2011) propose an alternative approach using the well-known variance cluster (VC) estimator (formally described in the 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 is still valid (provided it is rescaled). 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 is simple to implement.

In this section, we have shown that with ϕ-mixing and some technical conditions, $\theta_0 \in \Theta$ can be recovered using (4). 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 3, as well as for $\kappa(m) = O(\sqrt{m})$.

5. The Role of Homophily

We now turn to economic models of network formation. Recall that Assumption 2 imposes a specific dependence structure. Under those restrictions, 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.[20] We assume the following weak version of homophily:

Assumption 8 (Asymptotic homophily). *We assume that, for all $ij \in S_m$:*

$$d(i,j)^{-\lambda} \sup_{m} \sup_{W_{m,-ij}} h_{ij}(W_{m,-ij};\theta) \to -\infty \text{ as } d(i,j) \to \infty$$

for some $\lambda \geq 1$ and all $\theta \in \Theta$.

Assumption 8 then requires that the probability of a link between two infinitely distant individuals is zero, and that the rate at which the value of the relationship decreases with the distance is sufficiently fast. Note that Assumption 8 only requires that homophily holds asymptotically, hence allowing for a wide range of locally non-homophilic preferences.

We show that Assumption 8 has powerful implications.

Proposition 2. *If assumptions 1, 2, 5 and 8 hold, then assumptions 3 and 6.1 also hold. Moreover a necessary condition for the identification condition (assumption 4) to hold under assumption 8 is $\kappa(m) = O(\sqrt{m})$.*

Assumptions 1, 2 and 8 are assumptions on the preferences, i.e. h_{ij}. Assumption 5 is the increasing domain assumption. Those assumptions imply that assumptions 3 and 6.1, which are high-level assumptions on the random-field $\{q_s(\theta)\}$ also hold. The main benefit of Proposition 2 is therefore that the conditions for ϕ-mixing can be directly verified using the primitive equation of the structural model, as shown in Example 3.

Example 3 (Example 1 continued). *Recall that:*

$$h_{ij}(W_{m,-ij};\theta) = \alpha n_i(W_{m,-ij}) + \alpha n_j(W_{m,-ij}) - \beta|x_i - x_j|$$

and that $x_i, x_j \in \mathcal{X} = \mathbb{R}$ so $d(i,j) = |x_i - x_j|$. Then, Assumption 8 holds as long as $\beta > 0$. ■

[20] Many definitions of homophily exist in the economic literature (see, for example, Currarini et al. (2009, 2010) and Bramoullé et al. (2012)). In particular, some papers explicitly define homophily using a distance function on the space of individual characteristics (for example, Boucher (2015), Johnson & Gilles (2003), Marmaros & Sacerdote (2006), and Iijima & Kamada (2010)).

Note also that proposition 2 provides the order of the convergence rate, which is a necessary condition for identification. We discuss in detail the identification of the model in the next section.

6. Identification.

We now discuss further the implications of the identification condition stated in assumption 4. In order to discuss the intuition, as well as the implications of the assumptions, it helps to concentrate on the special case where $h_{ij}(\theta)$ is linear.

6.1. Linearity in θ.

We now show that, for the special case where $h_s(W_{m,-s}, \theta)$ is linear in θ, i.e. $h_s(W_{m,-s}, \theta) = \Gamma'_{s,m}\theta$ for some vector $\Gamma_{s,m}$, given the adequate choice of $\kappa(m)$, assumption 4 holds whenever $\mathbb{E}[\Gamma_{s,m}\Gamma'_{s,m}]$ is finite and non-singular. Formally:

Proposition 3. *If $h_s(.)$ is linear in θ, i.e. $h_s(W_{m,-s};\theta) = \Gamma'_{s,m}\theta$ such that $\liminf_{m\to\infty} |\lambda_m| > 0$ where λ_m is the smallest eigenvalue of $\mathbb{E}[\Gamma_{s,m}\Gamma'_{s,m}]$, and $\limsup_{m\to\infty} \mathbb{E}[\Gamma_{s,m}\Gamma'_{s,m}] < \infty$ then the limit of the PL, $\lim_{m\to\infty} \mathbb{E}[\mathcal{L}_m(\theta)]$, is uniquely maximized at the true value θ_0.*

Proposition 3 can be seen as a corollary to proposition 1, in the sense that limit of the PL is uniquely maximized at the true value θ_0 mainly because the conditional distributions are well specified. It also provides more classical conditions on the payoff function, as opposed to the generic formulation in assumption 4.[21]

However, proposition 3 still requires somewhat high-level conditions for identification since $\Gamma_{s,m}$ is a function of $W_{m,-s}$. In general, finding sufficient primitive conditions for identification is very hard and is left for future research.

In what follows, we discuss the implication of the identification condition for some special cases. In particular, in the next section, we discuss the link between the identification and the sparsity of the network, as well as the role of $\kappa(m)$ for the identification of the model.

[21] For instance, similar conditions are presented by Qu & Lee (2015).

6.2. Sparsity

As discussed by de Paula et al. (2016), the identification condition is closely linked to the sparsity of the network. We say that a network is sparse if it is not dense, where *dense* means that the number of links is $O(m)$.

Recall that assumptions 2 and 5 rule out the possibility that individuals value an infinite number of links. Together with assumption 8, it implies that the network is sparse.[22] However, we argue that we may still expect the network to be sparse in some cases, without necessarily assuming that $\chi_{ij,ik} = 0$ whenever $\max\{d(i,j), d(i,k)\} > \bar{d}$. Example 4 illustrates.

Example 4 (Example 1 modified). *Let i and j be the two* farthest *individuals in the population. We have:*

$$h_{ij}(W_{m,-ij}, \theta) = -\beta d(i,j) + \alpha \sum_{k \neq i,j} w_{ik} + w_{jk}.$$

Note that here we assume that, $\alpha_{ik} = \alpha_{jk} = \alpha$ for all i, j, k, irrespective of the distances between them.

Let $d(i,j) = |x_i - x_j|$, since i and j are the two farthest individuals, we know, under assumption 5, that there may be at most $d(i,j) + 1$ individuals in the population,[23] *so:*

$$h_{ij} \leq -\beta d(i,j) + 2\alpha(d(i,j) + 1) = (2\alpha - \beta)d(i,j) + 2\alpha.$$

This upper bound goes to $-\infty$ if $\beta > 2\alpha$, so the two farthest individuals (i and j) will not be linked with probability approaching 1 as $m \to \infty$.

This is confirmed by numerical simulations. Let $\alpha = 1$ and consider the following two specifications:

$$\begin{array}{ll} \textit{Specification 1:} & h_{ij} = n_i + n_j - 3|x_i - x_j| \\ \textit{Specification 2:} & h_{ij} = n_i + n_j - \exp\{|x_i - x_j|\} \end{array}$$

We simulate 1000 draws form the joint distribution (1) using a Gibbs-Sampling algorithm. Figure 2 show $\mathbb{E}(n_i + n_j)^2$, for each specification, as n grows. Overall, the result is independent of the size of the population n, irrespective of the specification. ■

Example 4 shows that sparsity can be achieved even without arbitrarily bounding the number of links, althought the conditions are stronger than those

[22] This can be seen as a corollary to proposition 2.

[23] *Recall here that individuals are located on the line. A similar argument applies for $T \geq 1$.*

of assumption 8. In general, we expect the sufficient conditions on $h_{ij}(W_{m,-ij},\theta)$ to be similar to assumption 8, but where λ is bounded below by a function of T.

We now discuss the intuition for $\kappa(m) = O(\sqrt{m})$ under assumption 8. Example 5 illustrates.

Example 5 (Example 1 modified). *Consider a modification of Example 1, where*

$$h_{ij}(W_{m,-ij},\theta) = -\beta\mathbf{1}\{|x_i - x_j| = 1\} - \infty\mathbf{1}\{|x_i - x_j| > 1\} + \sum_{k\neq i,j}\alpha_{ik}w_{ik} + \alpha_{jk}w_{jk}.$$

In this extreme example, individuals located at a distance greater than one create a link with a probability equal to 0*. In other words, individuals may only create links with their direct (left and right) neighbours. Since, for a population of* $m = n(n-1)/2$ *pairs, there are at most* $n-1$ *pairs located at a distance less than* 1 *(under assumption 5), the pseudo-likelihood (5) can be written as:*

$$\mathcal{L}_m(\theta) = \kappa(m)^{-1}\left[\sum_{s:|s_1-s_2|=1} q_{s,m}(w_{s,m}|W_{m,-ij},\theta)\right]$$

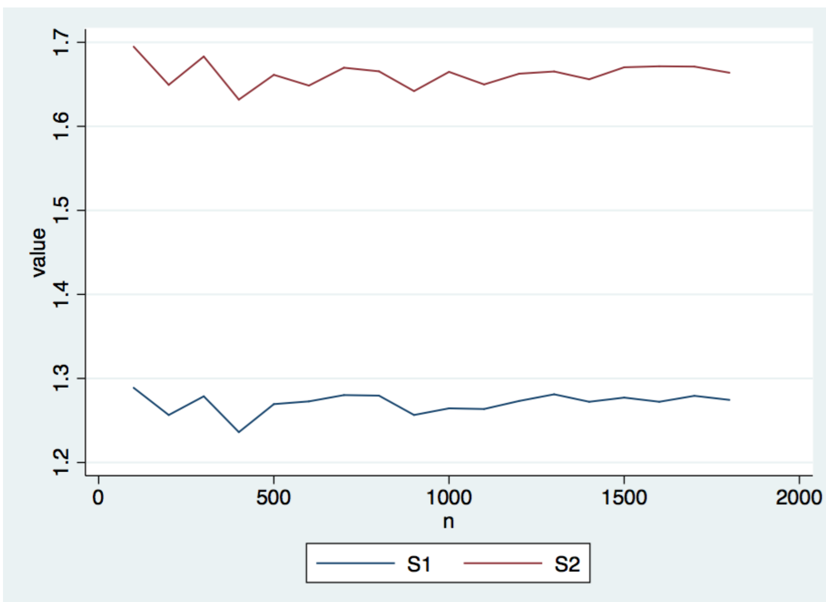
Note that we need only to sum over pairs s *such that* $|s_1 - s_2| = 1$ *since the* $q_{s,m}(\theta) = 0$ *for all the other pairs, with probability* 1 *under assumption 6.1. Since there are only* $n-1$ *pairs* s *such that* $|s_1 - s_2| = 1$*, we immediately get* $\kappa(m) = n-1$. ■

Proposition 2 shows that the intuition in Example 5 extends more generally, i.e. whenever assumption 8 holds. We now turn to our empirical application.

7. Friendship networks

We are interested in the determinants of friendship formation among high school students (e.g. Boucher (2015), Currarini et al. (2009, 2010) and (Mele, 2016)). We use the Add Health database, which provides information on the friendship networks of American high school students. We concentrate on the "saturated sample," which provides information on 3,449 teenagers from 16 schools. In the database, we know if student i nominated j as a friend, and vice-versa. Since we focus of undirected networks, we assume that i and j are friends if i nominated j, *or* j nominated i.

Figure 2: Sparsity (Simulations): $\mathbb{E}(n_i + n_j)^2$ as a function of the number of individuals for S1 and S2.



Specifically, for each n, $\mathbb{E}(n_i + n_j)^2$ is computed as the average (over the 1000 networks simulated), of the average (over the $n(n-1)/2$ pairs) value of $(n_i + n_j)^2$.

We consider two sets of variables. The first set ($x_{ij} \in \mathcal{X} \subset \mathbb{R}^2$) represents the (normalized) location of teenagers' homes.[24] As discussed in Section 4, the asymptotics are well defined on this variable since there must exist a minimal distance between any two individuals' homes. As the number of individuals (and therefore pairs) increases, the size of the residential neighbourhoods must also increase. The second set of variables (z_{ij}) contains other relevant characteristics of the pair (i,j). We use the following specification:

$$z_{ij} = [\texttt{Gender}_{ij},\ \texttt{White}_{ij},\ \texttt{Black}_{ij},\ \texttt{Hisp}_{ij},\ \texttt{Work}_{ij},\ \texttt{Grade}_{ij}]$$

where $\texttt{Gender}_{ij}$ is a binary variable that equals 1 if i and j are of the same gender; $\texttt{White}_{ij}$ is a binary variable that equals 1 if i and j are both white, $\texttt{Black}_{ij}$ is a binary variable that equals 1 if i and j are both black, $\texttt{Hisp}_{ij}$ is a binary variable that equals 1 if i and j are both Hispanic, $\texttt{Grade}_{ij}$ represents the absolute value of the difference between i and j's grade levels, and $\texttt{Work}_{ij}$ gives the sum of the weekly hours worked by i and j. The intuition is that the hours spent working could act as a substitute to (inschool) friendship relations.

We assume the following utility specification:

$$h_{ij}(W_m;\theta) \quad = \quad \theta_0 + [n_i(W_{m,-ij}) + n_j(W_{m,-ij})]\theta_1 + \theta_2 d(x_i,x_j) + z_{ij}\beta \quad (6)$$

where $\beta = [\theta_3,\theta_4,\theta_5,\theta_6,\theta_7,\theta_8]'$. We can show that assumption 2 holds since $h_{ij}(W_m;\theta)$ can also be written as:

$$h_{ij}(W_m;\theta) \quad = \quad \theta_0 + z_{ij}\beta + \theta_2 d(x_i,x_j) + \sum_{k\neq i,j}[w_{ik}\theta_1 + w_{jk}\theta_1]$$

Then, one can define $\chi_{ij}(\theta) = \theta_0 + z_{ij}\beta + \theta_2 d(x_i,x_j)$, $\tilde{\chi}_{ij,ik} = \theta_1$ and $\tilde{\chi}_{ij,jk} = \theta_1$.

The model is therefore a special case of the general model developed in the previous sections. Tables 1 and 2 give a summary of the variables used for

[24]Constructed using normalized GPS coordinates. See appendix for the construction of the distance function as well as details on the scaling factor $\kappa(m)$.

Table 1: Summary Statistics (Pairs)

Variable	Mean	Std	Min	Max
w_{ij}	0.004	0.066	0	1
$n_i + n_j$	8.454	4.855	0	40
`Grade`	0.992	0.849	0	5
`Gender`	0.500	0.500	0	1
`White`	0.255	0.436	0	1
`Black`	0.049	0.216	0	1
`Hisp`	0.117	0.322	0	1
`Work`	17.201	17.370	0	170
$d(i, j)$	13.868	20.940	0	379.215

Number of pairs: 1,624,408

Table 2: Summary Statistics (Individuals)

Variable	Mean	Std	Min	Max
n_i	4.242	3.310	0	21
Female	0.495	0.500	0	1
Grade level	10.194	1.493	7	12
White	0.586	0.493	0	1
Black	0.159	0.366	0	1
Hispanic	0.202	0.401	0	1

Number of individuals: 3,449
Number of residential communities: 16
Number of high schools: 16

individuals, and for pairs of individuals.[25] We now proceed with the estimation.

Following Bester et al. (2011), we use cluster-robust standard errors. Marginal effects are reported in Table 3. The number of friends two individuals have in common has a positive impact on the strength of their friendship. We also observe homophily with respect to grade levels, gender, and racial variables. Geographic distance has a significant negative impact, which is coherent with Assumption 8. Finally, we find a small negative effect of the number of hours worked by friends on the value of the link (this effect is not statistically significant).

[25] Note that the racial variables are not necessarily exclusive (for example, a biracial individual could identify as both black and white). We omitted the racial categories "Asian," "Native" and "Other."

Table 3: Maximum Likelihood Estimates (Logit), Marginal Effects ($\times$ 1000)

Variable	Estimate	S.E.
$n_i + n_j$	0.341^{**}	0.116
`Grade`	-5.061^{**}	0.861
`Gender`	1.530^{**}	0.289
`White`	6.399^{**}	2.342
`Black`	7.012^{**}	2.014
`Hisp`	1.890^{**}	0.610
`Work`	-0.024	0.023
$d(i,j)$	-0.121^{*}	0.059

Note: ** significant at 1%, and * significant at 5%. Standard errors computed using clustered-variance estimator by residential communities. See Appendix for details.

8. Conclusion

In this paper, we have used the theory of Markov random fields in order to develop a simple estimator for an interesting class of ERGMs. 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. The methodology is appealing because it is simple and flexible. We also provided an empirical application using data about the formation American teenagers' friendship networks. We find that there is a positive influence of the individuals' number of friends on link formation, as well as evidence of homophily on gender, grade levels, race, and geographic location.

9. References

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

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 192–236).

Besag, J. (1975). Statistical analysis of non-lattice data. *The statistician*, (pp. 179–195).

Bester, C. A., Conley, T. G., & Hansen, C. B. (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics*, *165*(2), 137–151.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and economic behavior*, *5*(3), 387–424.

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

Boucher, V. (2016). Conformism and self-selection in social networks. *Journal of Public Economics*, *136*, 30–44.

Boucher, V., & Fortin, B. (2016). Some challenges in the empirics of the effects of networks. *in The Oxford Handbook on the Economics of Networks, Editors: Yann Bramoullé, Andrea Galeotti, Brian W. Rogers*.

Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., & Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, *147*(5), 1754–1786.

Chandrasekhar, A. G. (2016). Econometrics of network formation. *in The Oxford Handbook on the Economics of Networks, Editors: Yann Bramoullé, Andrea Galeotti, Brian W. Rogers*.

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*.

Christakis, N. A., Fowler, J. H., Imbens, G. W., & Kalyanaraman, K. (2010). An empirical model for strategic network formation. Tech. rep., National Bureau of Economic Research.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics*, *92*(1), 1–45.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, *77*(4), 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*, *107*(11), 4857–4861.

de Paula, A., Richards-Shubik, S., & Tamer, E. (2016). Identifying preferences in networks with bounded degree. *Working Paper*.

De Paula, A., et al. (2016). Econometrics of network models. *Forthcoming in Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congres*.

Echenique, F., & Fryer Jr, R. G. (2007). A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, (pp. 441–485).

Fortin, B., & Yazbeck, M. (2015). Peer effects, fast food consumption and adolescent weight gain. *Journal of health economics*, *42*, 125–138.

Goldsmith-Pinkham, P., & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, *31*(3), 253–264.

Graham, B. S. (2016a). An empirical model of network formation: detecting homophily when agents are heterogenous. *Forthcoming in Econometrica*.

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

Hammersley, J. M., & Clifford, P. (1971). Markov fields on finite graphs and lattices. *Unpublished*.

Henry, M., & Mourifié, I. (2013). Euclidean revealed preferences: Testing the spatial voting model. *Journal of Applied Econometrics*, *28*(4), 650–666.

Hsieh, C.-S., & Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, *31*(2), 301–319.

Iijima, R., & Kamada, Y. (2010). Social distance and network structures. *Unpublished manuscript, Harvard University*.

Jenish, N., & Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, *150*(1), 86–98.

Johnson, C., & Gilles, R. P. (2003). Spatial social networks. In *Networks and Groups*, (pp. 51–77). Springer.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Lee, J., Kaiser, M. S., & Cressie, N. (2001). Multiway dependence in exponential family conditional distributions. *Journal of multivariate analysis*, *79*(2), 171–190.

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

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

Marmaros, D., & Sacerdote, B. (2006). How do friendships form? *Quarterly Journal of Economics*, (pp. 79–119).

Mele, A. (2016). A structural model of segregation in social networks. *Forthcoming in Econometrica*.

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

Monderer, D., & Shapley, L. S. (1996). Potential games. *Games and economic behavior*, *14*(1), 124–143.

Mourifié, I., & Siow, A. (2014). Cohabitation versus marriage: Marriage matching with peer effects. *Available at SSRN 2541895*.

Patacchini, E., & Zenou, Y. (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics*, *42*(6), 938–949.

Pötscher, B. M., & Prucha, I. R. (1991). Basic structure of the asymptotic theory in dynamic nonlineaerco nometric models, part i: consistency and approximation concepts. *Econometric reviews*, *10*(2), 125–216.

Qu, X., & Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, *184*(2), 209–232.

Rolfe, M., Toomet, O., & van der Leij, M. (2013). On the relationship between unexplained wage gap and social network connections for ethnical groups. *Network Science*.

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

Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, *85*(409), 204–212.

Watts, A. (2007). Formation of segregated and integrated groups. *International Journal of Game Theory*, *35*(4), 505–519.

White, H. (2001). *Asymptotic theory for econometricians*. Academic press New York.

10. Appendix

10.1. Proof of Theorem 1

Lemma 1. *Assumption 3 implies that*

1. $\sum_{d=1}^{\infty} d^{T-1}\bar{\phi}_{1,1}^{1/2}(d) < \infty$.
2. $\sum_{d=1}^{\infty} d^{T-1}\bar{\phi}_{k,l}(d) < \infty$, *for* $k+l \leq 4$.
3. $\bar{\phi}_{1,\infty}(d) = O(d^{-T-\epsilon})$ *for some* $\epsilon > 0$.

Proof. The proof is immediate. Indeed, There exists $M > 0$ such that

$$\begin{aligned}\sum_{d=1}^{\infty} d^{T-1}\bar{\phi}_{1,1}^{1/2}(d) &\leq M\sum_{d=1}^{\infty} d^{T-1}d^{-T+1-\epsilon} \\ &\leq M\sum_{d=1}^{\infty} d^{-\epsilon} < \infty.\end{aligned}$$

Under Assumption 4, it is sufficient to show that:[26]

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

In order to show that this condition holds, it is sufficient to show that the conditions of Theorems 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:

$$\lim \sup_{m\to\infty} \frac{1}{|S_m|}\sum_{s=1}^{m} \mathbb{E}(\bar{q}_{s,m}^{p} 1_{\{\bar{q}_{s,m}^{p} > k\}}) \to 0 \text{ as } k \to \infty,$$

   for some $p \geq 1$, and where $\bar{q}_{s,m} = \sup_\theta |q_{s,m}(w_{s,m}|x, W_m, \theta)|$.

4. Stochastic equicontinuity: For every $\epsilon > 0$,

$$limsup_m \frac{1}{|S_m|}\sum_{s=1}^{m} P(\sup_{\theta'\in\Theta}\sup_{\theta\in B(\theta',\delta)} |q_{s,m}(\theta) - q_{s,m}(\theta')| > \epsilon) \to 0 \text{ as } \delta \to 0,$$

   where $B(\theta', \delta)$ is the open ball $\{\theta \in \Theta : \|(\theta' - \theta)\| < \delta\}$.

[26] See, for example, Gallant and White (1988), p.18.

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 5. Condition 2 is verified by construction, and Conditions 5 and 6 hold from Assumptions 6 and 3. Conditions 3 and 4 hold from the following: Under Condition 5, $\sup_\theta |q_{s,m}(w_{s,m}|W_m, \theta)|$ is $L^{(1+\eta)}$ integrable, which implies the uniform $L^{(1+\eta)}$ integrability of $|q_{s,m}(w_{s,m}|W_m, \theta)|$.

The next lemma shows that Assumption 6 implies Condition 4.

Lemma 2. *Condition 4 is implied by Assumption 6.*

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 6, 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. □

From the previous lemma, Conditions 1-6 are respected, hence Theorem 2 and 3 from Jenish and Prucha (2009) apply. This completes the proof. □

Proof of Theorem 2

We want to show that $\sqrt{\kappa(m)}(\hat{\theta}_m - \theta_0) \Rightarrow N(0, D_0(\theta_0)^{-1}B_0(\theta_0)D_0(\theta_0)^{-1})$. 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

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

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. $\frac{\sigma_m^2}{\kappa(m)} \to B_0(\theta_0)$;
2. $\sigma_m^{-1}Q_m \Rightarrow N(0, I)$;
3. $[\frac{\partial^2 \mathcal{L}_m(\overline{\theta}_m)}{\partial \theta \partial \theta'}] \to_p D_0(\theta_0)$.

Again, we proceed in a series of lemmata.

Lemma 3. *Under Assumption 4,* $\frac{\sigma_m^2}{\kappa(m)} \to B_0(\theta_0)$.

Proof.

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

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 4). Hence, $\frac{\sigma_m^2}{\kappa(m)} \to B_0(\theta_0)$. □

Lemma 4. *Under Assumption 7,* $\sigma_m^{-1}Q_m \Rightarrow N(0, I)$

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. $\liminf_{m\to\infty} \frac{\sigma_m^2}{\kappa(m)} > 0$.

Condition 1 is implied by Assumption 5. Condition 3 holds from Assumption 7, and Condition 4 is implied by Lemma 3. □

Lemma 5. $\frac{\partial^2 \mathcal{L}_m(\bar{\theta}_m)}{\partial\theta\partial\theta'} \to_p D_0(\theta_0)$

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 7 instead of Assumptions 6. □

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

Proof of Proposition 2

Claim 1: Assumption 6.1 holds. Recall that $q_{s,m}(\theta) = q_{s,m}(w_{s,m}|W_{m,-s};\theta)$. We have:

$$
\begin{aligned}
&\sup_m \sup_s \mathbb{E}[\sup_\theta |q_{s,m}(\theta)|^{1+\eta}] \\
&= \sup_m \sup_s \mathbb{E}\Big[\mathbb{E}[\sup_\theta |q_{s,m}(\theta)|^{1+\eta}|W_{m-s}]\Big] \\
&= \sup_m \sup_s \mathbb{E}\Big[P(w_{s,m}=1|W_{m-s};\theta)\sup_\theta |q_{s,m}(w_{s,m}=1|W_{m-s})|^{1+\eta}] + \\
&\qquad\qquad P(w_{s,m}=0|W_{m-s};\theta)\sup_\theta |q_{s,m}(w_{s,m}=0|W_{m-s})|^{1+\eta}]\Big] \\
&= \sup_m \sup_s \mathbb{E}\Big[\frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}} \sup_\theta |h_s(W_{m,-s};\theta) - \ln[1+\exp\{h_s(W_{m,-s};\theta)\}]|^{1+\eta} \\
&+ \frac{1}{1+\exp\{h_s(W_{m,-s};\theta)\}} \sup_\theta \ln[1+\exp\{h_s(W_{m,-s};\theta)\}]^{1+\eta}\Big], \\
&= \sup_m \sup_s \mathbb{E}\Big[\frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}} |h_s(W_{m,-s};\theta^*) - \ln[1+\exp\{h_s(W_{m,-s};\theta^*)\}]|^{1+\eta} \\
&+ \frac{1}{1+\exp\{h_s(W_{m,-s};\theta)\}} \ln[1+\exp\{h_s(W_{m,-s};\tilde{\theta})\}]^{1+\eta}\Big],
\end{aligned}
$$

where $\theta^* = \arg\max\{|h_s(W_{m,-s};\theta) - \ln[1+\exp\{h_s(W_{m,-s};\theta)\}]|\}$ and $\tilde{\theta} = \arg\max\{\ln[1+\exp\{h_s(W_{m,-s};\tilde{\theta})\}]\}$. Under assumptions 2 and 5, we know that $h_s(W_{m,-s};\theta)$

is uniformly bounded from above respect to that $\theta, W_{m,-s}, m$, and s; then the terms $\frac{1}{1+\exp\{h_s(W_{m,-s};\theta)\}} \ln[1+\exp\{h_s(W_{m,-s};\tilde{\theta})\}]^{1+\eta}$ and $\frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}} |\ln[1+\exp\{h_s(W_{m,-s};\theta^*)\}]|^{1+\eta}$ are also bounded. The remaining term to study is

$$
\begin{aligned}
\frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}}|h_s(W_{m,-s};\theta^*)|^{1+\eta} &= \frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}}|h_s(W_{m,-s};\theta)|^{1+\eta}\left|\frac{h_s(W_{m,-s};\theta^*)}{h_s(W_{m,-s};\theta)}\right|^{1+\eta} \\
&\leq \frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}}|h_s(W_{m,-s};\theta)|^{1+\eta}.
\end{aligned}
$$

where the last inequality holds since we have:

$$
\begin{aligned}
\theta^* &= \arg\max\{|\ln \frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}}|\} \\
&= \arg\min \frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}} \\
&= \arg\min h_s(W_{m,-s};\theta).
\end{aligned}
$$

Finally, we have that:

$$
\lim_{h_s(W_{m,-s};\theta)\to\infty} \frac{\exp\{h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}}|h_s(W_{m,-s};\theta)|^{1+\eta} \to 0
$$

uniformly over $\theta, W_{m,-s}, m$, and s. This completes the proof.

Claim 2: $\kappa(m) = O(n)$.

Let's consider:

$$
\kappa(m)\mathcal{L}_m(\theta) = \sum_s^m \Big(w_s h_s(W_{m,-s};\theta) - \ln[1+\exp\{h_s(W_{m,-s};\theta)\}]\Big)
$$

For a fixed individual i let order the rest of the individuals $i_2, ..., i_n$ in the lattice such that $d(i,i_2) \leq d(i,i_3) \leq .... \leq d(i,i_n)$. Notice that we may have multiple individual at the same distance $\bar{d}$ of i. Therefore, we can partition $\{i_2,...,i_n\}$ as follows $\{i_2,...,i_n\} = \{B_1\} \cup ... \cup \{B_{K_i}\}$ such that $\forall (l,l') \in B_k$ $d(i,\tilde{l}) < d(i,l) = d(i,l') \equiv d_k < d(i,\hat{l})$ for $\tilde{l} \in B_{k-1}$ and $\hat{l} \in B_{k+1}$.

In such a case $\kappa(m)\mathcal{L}_m(\theta)$ can be rewritten as:

$$\kappa(m)\mathcal{L}_m(\theta) = \frac{1}{2}\sum_{i}^{n}\sum_{j=i_2}^{i_n}\Big(w_{ij}h_{ij}(W_{m,-ij};\theta) - \ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}]\Big)$$

or equivalently as

$$\kappa(m)\mathcal{L}_m(\theta) = \frac{1}{2}\sum_{i}^{n}\sum_{k=1}^{K_i}\sum_{j\in B_k}\Big(w_{ij}h_{ij}(W_{m,-ij};\theta) - \ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}]\Big)$$

Consider the series:

$$\sum_{j=i_2}^{i_n}\Big(\mathbb{E}w_{ij}h_{ij}(W_{m,-ij};\theta) - \ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}]\Big)$$

Because

$$0 \leq \ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}] \leq \exp\{h_{ij}(W_{m,-ij};\theta)\},$$

we have

$$\begin{aligned}\sum_{j=i_2}^{i_n}\ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}] &\leq \sum_{k=1}^{K_i}\sum_{j\in B_k}\exp\{\bar{h}_i(d_k)\} \\ &= \sum_{k=1}^{K_i}|B_k|\exp\{\bar{h}_i(d_k)\}\end{aligned}$$

where $\bar{h}_i(d_k) = \sup_\theta \sup_{ij:j\in B_k}\sup_{W_m\in\mathbb{W}_m} h_{ij}(W_{m,-ij};\theta)$.

From Jenish and Prucha (2009) (Lemma A.1 (iii)), there are at most Cd^{T-1} (where $C > 0$ is a positive constant) individual located at a distance $d_k \in [d, d+1)$ from i, so we have:

$$
\begin{aligned}
\sum_{j=i_2}^{i_n} \ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}] &\leq C\sum_{d=1}^{K_i} d^{T-1}\exp\{\bar{h}_i(d+1)\} \\
&\leq C\sum_{d=1}^{\infty} d^{T-1}\exp\{\bar{h}_i(d+1)\}.
\end{aligned}
$$

Using the Cauchy criteria for convergent series, we can see that the latter summation converges if $\lim_{d\to\infty}\Big[(T-1)\frac{\ln d}{d}+\frac{\bar{h}_i(d+1)}{d}\Big]<0$, which is true for $\lambda=1$ in assumption 8.

Consider now the series:

$$
\sum_{j=i_2}^{i_n} \mathbb{E}w_{ij}h_{ij}(W_{m,-ij};\theta) \leq \sum_{j=i_2}^{i_n} \mathbb{E}w_{ij}\big|h_{ij}(W_{m,-ij};\theta)\big|
$$

We have:

$$
\begin{aligned}
\sum_{j=i_2}^{i_n} \mathbb{E}w_{ij}\big|h_{ij}(W_{m,-ij};\theta)\big| &= \sum_{j=i_2}^{i_n} \mathbb{E}\Big[\big|h_{ij}(W_{m,-ij};\theta)\big|\mathbb{E}[w_{ij}|W_{m,-ij}\}]\Big], \\
&\leq \sum_{j=i_2}^{i_n} \big|\bar{h}_{ij}\big|\mathbb{E}\Big[P(w_{ij}=1|W_{m,-ij})\Big], \\
&\leq \sum_{j=i_2}^{i_n} \big|\bar{h}_{ij}\big| \sup_{W_m\in\mathbb{W}_m} P(w_{ij}=1|W_{m,-ij}), \\
&= \sum_{j=i_2}^{i_n} \big|\bar{h}_{ij}\big| \sup_{W_m\in\mathbb{W}_m} \frac{\exp\{h_{ij}(.)\}}{1+\exp\{h_{ij}(.)\}}, \\
&\leq \sum_{j=i_2}^{i_n} \big|\bar{h}_{ij}\big| \frac{\exp\{\bar{h}_{ij}\}}{1+\exp\{\bar{h}_{ij}\}}.
\end{aligned}
$$

where $\bar{h}_{ij}=\sup_{W_m\in\mathbb{W}_m} h_{ij}(W_{m,-ij};\theta)$. The last inequality holds because $\frac{\exp\{x\}}{1+\exp\{x\}}$ is monotone in x.

$$\begin{aligned}\sum_{j=i_2}^{i_n} |\bar{h}_{ij}| \frac{\exp\{\bar{h}_{ij}\}}{1+\exp\{\bar{h}_{ij}\}} &= \sum_{k=1}^{K_i} \sum_{j \in B_k} |\bar{h}_{ij}| \frac{\exp\{\bar{h}_{ij}\}}{1+\exp\{\bar{h}_{ij}\}} \\ &\leq \sum_{k=1}^{K_i} |B_k| \big|\tilde{h}_i(d_k)\big| \frac{\exp\{\tilde{h}_i(d_k)\}}{1+\exp\{\tilde{h}_i(d_k)\}},\end{aligned}$$

where $\tilde{h}_i(d_k) = argmax_{\{\bar{h}_{ij}:d(i,j)=d_k\}} \big|\bar{h}_{ij}\big| \frac{\exp\{\bar{h}_{ij}\}}{1+\exp\{\bar{h}_{ij}\}}$. Notice that $\tilde{h}_i(d_k)$ is finite, since $\bar{h}_{ij}$ is bounded from above and $\big|\bar{x}\big| \frac{\exp\{x\}}{1+\exp\{x\}} \to 0$ when $x \to \infty$. We can use similar derivations that earlier when invoking Lemma A.1 (iii) from Jenish and Prucha (2009) and then get the following:

$$\sum_{j=i_2}^{i_n} \mathbb{E} w_{ij} |h_{ij}(W_{m,-ij};\theta)| \leq C \sum_{d=1}^{\infty} d^{T-1} \frac{\exp\{\tilde{h}_i(d)\}}{1+\exp\{\tilde{h}_i(d)\}} |\tilde{h}_i(d)|.$$

Again, using the Cauchy criteria for convergent series, we can see that the latter summation converges if $\lim_{d\to\infty} \Big[(T-1)\frac{\ln d}{d} + \frac{\ln \big|\tilde{h}_i(d)\big|}{d} + \frac{\tilde{h}_i(d)}{d} - \frac{\ln(1+\tilde{h}_i(d))}{d}\Big] < 0$, which is true for $\lambda = 1$ in assumption 8.

So, we have shown that

$$\Big|\sum_{j=i_2}^{i_n} \mathbb{E}[q_{ij}(\theta)]\Big| = \Big|\sum_{j=i_2}^{i_n} \mathbb{E}\Big[w_{ij} h_{ij}(W_{m,-ij};\theta) - \ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}]\Big]\Big| < \infty$$

As can be seen in the proof of the Claim 1, $\mathbb{E}[q_{ij,m}(\theta)]$ is a weighting average of $q_{ij,m}(w_{ij,m} = 1|W_{m,-ij};\theta) = \ln \frac{\exp\{h_{ij}(W_{m,-ij};\theta)\}}{1+\exp\{h_{ij}(W_{m,-ij};\theta)\}}$ and $q_{ij,m}(w_{ij,m} = 0|W_{m,-ij};\theta) = -\ln[1+\exp\{h_{ij}(W_{m,-ij};\theta)\}]$ where the weights are all probabilities (non-negative). Because $q_{ij,m}(w_{ij,m} = 1|W_{m,-ij};\theta)$ and $q_{ij,m}(w_{ij,m} = 1|W_{m,-ij};\theta)$ are all non-positive uniformly over $\theta, W_{m,-ij}$, m, and ij then $\mathbb{E}[q_{ij,m}(\theta)] \leq 0$.

Therefore there exist A_i a non-positive finite constant, i.e. $\infty < A_i \leq 0$ such that $E[Z_i] \equiv \sum_{j=i_2}^{i_n} \mathbb{E}[q_{ij,m}(\theta)] = A_i$. $\kappa(m)\mathcal{L}_m(\theta)$ can be therefore rewritten as a summation from 1 to n of Z_i.

We have

$$\mathbb{E}[\mathcal{L}_m(\theta)] = \frac{1}{\kappa(m)}\sum_{i}^{n} A_i$$

Therefore, $\kappa(m) = O(n)$. This completes the proof.

Claim 3: Assumption 3 holds.

The objective is to show that, as the distance (in X) between the pairs located in U, and the pairs located in V goes to infinity, we have that $|P(A^U|B^V) - P(A^U)|$ goes to 0 (at the appropriate rate). The proof is rather lengthy and proceeds in a series of steps.

In order to clarify the argument, we briefly explain each step.

- Step 1: Since ϕ-mixing is defined as the distance between sets (i.e. $d(U,V)$) of pairs goes to infinity, we describe what happens to the distance between the individuals as $d(U,V)$ grows. This allows to separate three cases for $s : x_s \in U$, and $r : x_r \in V$
- Step 2 (First case): $d(s,r) = O(d(s_1,s_2))$
- Step 3 (Second case): $d(s,r) = O(d(r_1,r_2))$
- Step 4 (Third case): $d(s,r) = O(\min\{d(s_1,r_1), d(s_1,r_2), d(s_2,r_1), d(s_2,r_2)\})$

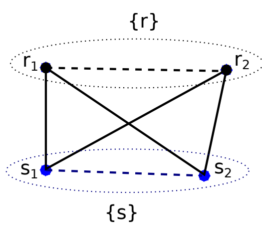
We also make the following remarks and notations:

- Technically A^U gives realizations of $\{q_{ij}\}_{ij}$ for individuals i and j, such that the pair (i,j) is located in the subspace U of X (i.e. $x_{ij} \in U \subset X$). To lighten the text, we will often use "i belongs to U", instead of "i belongs to a pair located in U", when there is no confusion. We will also often use U instead of $\{s\}_{s:x_s \in U}$.
- The sets A^U and B^V are realisations over the random-field $\{q_s(\theta)\}$, which is itself a function of the random-field $\{w_s\}$. Formally, $q_s(\theta) = w_s h_s(W_{m,-s};\theta)-$

$\ln(1+\exp\{h_s(W_{m,-s};\theta)\})$. From assumption 2, the function $h_s(W_{m,-s};\theta)$ only depends on neighbourhood links, so $q_s(\theta) = f(w_s, W_{N(s)})$.

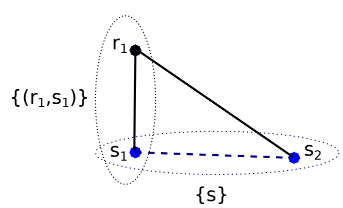
- We will expand the notion of neighbourhoods to set of pairs: $N(U) = \cup_{s:x_s \in U} N(s) \setminus \{s\}_{s:x_s \in U}$ is the set of neighboured pairs of pairs located in U (excluding pairs located in U).

Figure 3: Pairs do not share any individual



Note: Neighbouring pairs of pair s are represented by solid lines.

Figure 4: Pairs share an individual



Note: Neighbouring pairs of pair s are represented by solid lines.

Step 1: Distance between sets of pairs, and distance between individuals.

By definition, $d(U,V) = d(s^*, r^*) = \inf\{d(s,r)|s : x_s \in U, r : x_r \in V\}$. First, note that if $U \cap V \neq \emptyset$, i.e. there is a pair located in U, which is also located in V. We have $d(U,V) = 0$. We can therefore abstract for this case, and concentrate on the case where $U \cap V = \emptyset$.

Second, note that any couple of pairs (r,s) such that $x_s \in U$ and $x_r \in V$ is either such that s and r do not share any individual (as in Figure 3), or such that they share an individual (as in Figure 4).

Consider first the case where s and r don't share any individual, i.e. $s \cap r = \emptyset$, as in Figure 3. As $d(s,r)$ grows, there are three (exhaustive, but non mutually exclusive) possibilities:

(1) $d(s_1, s_2)$ goes to infinity, formally: $d(s,r) = O(d(s_1, s_2))$

(2) $d(r_1, r_2)$ goes to infinity, formally: $d(s,r) = O(d(r_1, r_2))$

(3) $\min\{d(s_1, r_1), d(s_1, r_2), d(s_2, r_1), d(s_2, r_2)\}$ goes to infinity, formally: $d(s,r) = O(\min\{d(s_1, r_1), d(s_1, r_2), d(s_2, r_1), d(s_2, r_2)\})$

To see why the order is as above, consider the first case (the other cases are symmetric): $d(s,r) = O(d(s_1, s_2))$. Consider Figure 3, where we let $d(s_2, r_2) = \epsilon_1 > 0$ and $d(r_1, r_2) = \epsilon_2 > 0$. This forces the distance between s_1 and s_2 to go to infinity as $d(s,r)$ goes to infinity. We have: $d(r_1, s_2) \leq \epsilon_1 + \epsilon_2$. Moreover, we have: $d(s_1, r_1) \leq d(s_1, s_2) + d(s_2, r_1)$, so:

$$d(s,r) \leq \epsilon_1/2 + d(s_1, s_2)/2 + (\epsilon_1 + \epsilon_2)/2$$

which implies: $d(s,r) = O(d(s_1, s_2))$. This argument can be repeated for all possible cases, and also for the case where s and r share an individual:

Consider the case where s and r share one individual, i.e. $s \cap r \neq \emptyset$, as in Figure 4. Similarly to the case where $s \cap r = \emptyset$, we have that as $d(s,r)$ grows, there are two exhaustive, but non mutually exclusive possibilities:

(1) $d(s,r) = O(d(s_1, s_2))$

(2) $d(s,r) = O(d(r_1, r_2))$

Note that here, in either case, we necessarily have that $d(s,r) = O(d(s_2, r_1))$.

This first Step shows that we need to look at three cases when $d(s,r)$ goes to infinity:

(1) $d(s,r) = O(d(s_1, s_2))$

(2) $d(s,r) = O(d(r_1, r_2))$

(3) $d(s,r) = O(\min\{d(s_1,r_1), d(s_1,r_2), d(s_2,r_1), d(s_2,r_2)\})$

where $s : x_s \in U$ and $r : x_r \in V$. We will study those cases in Steps 2, 3 and 4 below.

Step 2: Consider first the case where $d(U,V) = O(d(s_1,s_2))$ for some $s : x_s \in U$. We have, for $A^U \in \mathcal{A}_m^U$ and $B^V \in \mathcal{B}_m^V$:

$$|P(A^U|B^V) - P(A^U)| \leq \sum_{t=1}^{k} |P(A^{s_t}|B, A^{s_1}, ..., A^{s_{t-1}}) - P(A^{s_t}|A^{s_1}, ..., A^{s_{t-1}})| \quad (7)$$

from the chain rule, where $A^U = (A^{s_1}, ..., A^{s_k})$.

Notice that, given a fixed m, for any pair s q_s is a discrete random variable because it has a countable range. Indeed, recall that q_s is a function of a countable finite number of binary random variable, i.e. $q_s = \ln \frac{\exp\{w_{s,m} h_s(W_{m,-s};\theta)\}}{1+\exp\{h_s(W_{m,-s};\theta)\}}$. Let $\{a_1^s, ..., a_{\tilde{m}}^s\}$ the support of q_s where $\tilde{m}$ is the cardinality of this support. Notice that because of the homophily assumption (assumption 8) when the distance between the pairs $s = (s_1, s_2)$ converge to infinity, i.e. $d(s_1,s_2) \to \infty$ then the support is $\{-\infty, 0\}$.

There are three cases:

1. $A^s \in (-\infty, 0)$. Then, letting $d(s_1,s_2) \to \infty$, there exists $K > 0$ such that $P(q_s = A^s) = 0$ for all $d(s_1,s_2) > K$, since A^s falls outside the range of q_s. We therefore have: $P(A^{s_t}|B, A^{s_1}, ..., A^{s_{t-1}}) = P(A^{s_t}|A^{s_1}, ..., A^{s_{t-1}}) = 0$.
2. $A^s = -\infty$ but under Assumption 6.1, we know that $\sum_{k=1}^{\tilde{m}} |a_k^s| P(q_s = a_k^s) < \infty$. Then, $P(q_s = a_k^s) = 0$ whenever $a_k^s = -\infty$ otherwise q_s will not have a finite moment. We therefore have: $P(A^{s_t}|B, A^{s_1}, ..., A^{s_{t-1}}) = P(A^{s_t}|A^{s_1}, ..., A^{s_{t-1}}) = 0$
3. $A^s = 0$ This only happens as $d(s_1,s_2) = \infty$, where the range only has two elements, i.e. $-\infty$ and 0. Since $P(-\infty) = 0$, we have: $P(q_s = A^s) = 1$. Therefore, we have: $P(A^{s_t}|B, A^{s_1}, ..., A^{s_{t-1}}) = P(A^{s_t}|A^{s_1}, ..., A^{s_{t-1}}) = 1$ when $d(s_1,s_2) = \infty$ while $P(A^{s_t}|B, A^{s_1}, ..., A^{s_{t-1}}) = P(A^{s_t}|A^{s_1}, ..., A^{s_{t-1}}) = 0$ when $d(s_1,s_2) < \infty$.

Step 3: Consider now the case where $d(U,V) = O(d(r_1,r_2))$ for some $r : x_r \in$

V. For $A^U \in \mathcal{A}_m^U$ and $B^V \in \mathcal{B}_m^V$, we need to consider:

$$|P(A^U|B^V) - P(A^U)|$$

for all A^U and B^V such that $P(B^V) > 0$. Similarly to Step 2, we have: $P(q_r = B^r)$ has a range in $(-\infty, 0)$ if $d(r_1, r_2) < \infty$ so whenever there exists r such that $B^r \in (-\infty, 0)$, there exists $K > 0$ such that $P(B) = 0$ for all $d(r_1, r_2) > K$. In those cases, ϕ-mixing does not apply. Similarly, if $B^r = -\infty$, we also have $P(B) = 0$ from assumption 6.1 (which must hold for all pairs in the random field). Finally, suppose that $B^r = 0$, we have that $P(B^r = 0) = 0$ for $d(r_1, r_2) < \infty$ and $d(r_1, r_2) < \infty$ for $P(B^r = 0) = 1$ for $d(r_1, r_2) = \infty$.

In all cases, whenever there exists $r : x_r \in V$ such that $d(U, V) = O(d(r_1, r_2))$, ϕ-mixing holds trivially, either because $P(B) = 0$, or because $P(B)$ is a Dirac distribution.

Step 4:

It remains to check the case where $d(U, V)$ it not $O(d(s_1, s_2))$, or $O(d(r_1, r_2))$, for any pairs s or r such that $s : x_s \in U$ and $r : x_r \in V$. That is, in (7), the elements of the sum for which $d(U, V)$ it not $O(d(s_1, s_2))$ (step 2), and assuming that no pair $r : x_r \in V$ are such that $d(U, V) = O(d(r_1, r_2))$.

We define the following set:

$$C_{U,V} = \{s \in S_m | s_1 \in r : x_r \in U \text{ and } s_2 \in t : x_t \in V\}$$

That is, the set of pairs (i, j) such that i is in a pair located in U, and such that j is in a pair located in V. We also let $D_{U,V} = 0$ if $W_{C_{U,V}} = 0$, and $D_{U,V} = 1$ otherwise. Then, $D_{U,V} = 0$ implies that there is no link between an individual in U and an individual in V.

From Step 1 above, and since $d(U, V)$ it not $O(d(s_1, s_2))$, or $O(d(r_1, r_2))$, for any pairs s or r such that $s : x_s \in U$ and $r : x_r \in V$, we necessarily have that $d(U, V) = O(d(i, j))$ for all $(i, j) \in C_{U,V}$.

We will use the following result.

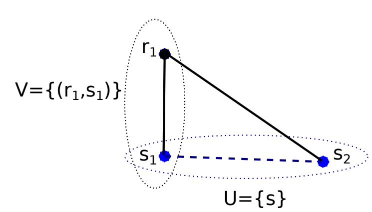
Step 4.1 $P(A^U|B^V, D_{U,V} = 0) = P(A^U|D_{U,V} = 0)$.

Equivalently: $P(A^U|B^V, W_{C_{U,V}} = 0) = P(A^U|W_{C_{U,V}} = 0)$. Indeed, typically, $W_{C_{U,V}}$ does contain information also contained in B^V. We claim that it contains all the information contained in B^V. Note that $W_{C_{U,V}} = 0$ means that all pairs between an individual in a pair in U, and an individual in a pair in V are not linked.

Now, by construction of the random-field $\{q_s\}$ from $\{w_s\}$ (that is: $q_s = f(w_s, W_{N(s)})$), B^V depends only on $W_{V\cup N(V)}$, while A^U depends only on $W_{U\cup N(U)}$. Note that this heavily relies on Proposition 1. There are four cases:

1. $U \cap V \neq \emptyset$: Since it implies that $d(U, V) = 0$, we can disregard this case (see Step 1). We can therefore interpret the three following cases as being conditional on $U \cap V = \emptyset$.
2. $U \cap N(V) \neq \emptyset$: In words, this is the set of pairs in U that have an individual in a pair in V. In Figure 5 for instance, this is pair s. Clearly, any pair in $U \cap N(V)$ has an individual in a pair in U and an individual in a pair in V. Therefore, $U \cap N(V) \subseteq C_{U,V}$.
3. $V \cap N(U) \neq \emptyset$: Similarly, in Figure 5, this is a pair in V that has an individual in U, i.e. (r_1, s_1). We also have: $V \cap N(U) \subseteq C_{U,V}$.
4. $N(U) \cap N(V) \neq \emptyset$: We need to describe the set $N(U) \cap N(V)$. Any pair $r \in N(U)$ has one (and only one) individual in a pair in U: say $r_1 \in (r_1, s_1) \in U$. Similarly, $r \in N(V)$ has one individual in a pair in V. There are two cases:

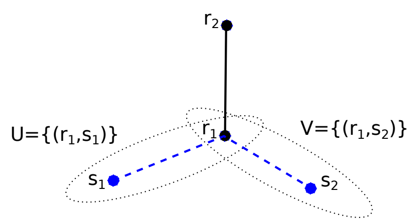
Figure 5: Cases 2 and 3: $V \cap N(U) \neq \emptyset$



Note: Neighbouring pairs of pair s are represented by solid lines.

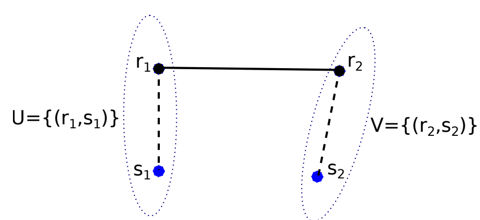
(a) $r_1 \in (r_1, s_2) \in V$: then $(r_1, s_2) \in V \cap N(U)$. That is: U and V share an individual (here r_1). However, this implies that $(r_1, s_1) \in C_{U,V}$ and $(r_1, s_2) \in C_{U,V}$ so conditional on $W_{C_{U,V}} = 0$, the realization of w_r (which is in the intersection of $N(U)$ and $N(V)$) brings no additional information on either (r_1, s_1) or (r_1, s_2). See Figure 6.

Figure 6: Case 4(a): $N(U) \cap N(V) \neq \emptyset$: r_1 belongs to a pair in U, as well as in a pair in V.



(b) $r_2 \in (r_2, s_2) \in V$: r is a link between an individual in a pair in U and an individual in a pair in V, therefore $r \in C_{U,V}$. See Figure 7.

Figure 7: Case 4(b): $N(U) \cap N(V) \neq \emptyset$: r_1 belongs to a pair in U, r_2 belongs to a pair in V.



Therefore, in all cases, B^V provides no information on A^U, conditional on $W_{C_{U,V}} = 0$.

Step 4: (continued)

Consider $A^U \in \mathcal{A}_m^U$ and $B^V \in \mathcal{B}_m^V$. Let partition U in U_1 and $U_2 = U \setminus U_1$ such that U_1 contains the pairs s for which $d(U, V) = O(d(s_1, s_2))$ (provided U_1 is non-empty).

We have:

$$
\begin{aligned}
|P(A^U|B^V) - P(A^U)| &= |P(A^{U_1}|A^{U_2}, B^V)P(A^{U_2}|B^V) - P(A^{U_1}|A^{U_2})P(A^{U_2})| \\
&\leq |P(A^{U_1}|A^{U_2}, B^V)P(A^{U_2}|B^V) - P(A^{U_1}|A^{U_2})P(A^{U_2}|B^V)| \\
&\quad +|P(A^{U_1}|A^{U_2})P(A^{U_2}) - P(A^{U_1}|A^{U_2})P(A^{U_2}|B^V)| \\
&\leq |P(A^{U_1}|A^{U_2}, B^V) - P(A^{U_1}|A^{U_2})|P(A^{U_2}|B^V) \\
&\quad +|P(A^{U_2}) - P(A^{U_2}|B^V)|P(A^{U_1}|A^{U_2})
\end{aligned}
$$

So the mixing coefficient $|P(A^U|B^V)-P(A^U)|$ is bounded by $|P(A^{U_1}|A^{U_2})P(A^{U_2})- P(A^{U_1}|$ and $|P(A^{U_2}) - P(A^{U_2}|B^V)|$. The first quantity was treated in Step 1.

We now proceed to show that $|P(A^{U_2}|B^V)-P(A^{U_2})| \leq 2[P(D_{U,V} = 1|B^V)+ P(D_{U,V} = 1)]$.

$$
\begin{aligned}
P(A^{U_2}|B^V) - P(A^{U_2}) &= P(A^{U_2}|B^V, D_{U,V} = 1)P(D_{U,V} = 1|B^V) \\
&\quad +P(A^{U_2}|B^V, D_{U,V} = 0)P(D_{U,V} = 0|B^V) \\
&\quad -P(A^{U_2}|D_{U,V} = 1)P(D_{U,V} = 1) \\
&\quad -P(A^{U_2}|D_{U,V} = 0)P(D_{U,V} = 0)
\end{aligned}
$$

And therefore, using Step 4.1:

$$
\begin{aligned}
|P(A^{U_2}|B^V) - P(A^{U_2})| &\leq |P(A^{U_2}|B^V, D_{U,V} = 1)P(D_{U,V} = 1|B^V) - P(A^{U_2}|D_{U,V} = 1)P(D_{U,V} = 1)| \\
&\quad +P(A^{U_2}|D_{U,V} = 0)|P(D_{U,V} = 0|B^V) - P(D_{U,V} = 0)| \\
&\leq |P(D_{U,V} = 1|B^V) + P(D_{U,V} = 1)| \\
&\quad +P(A^{U_2}|D_{U,V} = 0)|P(D_{U,V} = 0|B^V) - P(D_{U,V} = 0)| \\
&\leq 2[P(D_{U,V} = 1|B^V) + P(D_{U,V} = 1)].
\end{aligned}
$$

This shows that the ϕ-mixing coefficient is bounded by: $2[P(D_{U,V} = 1|B^V)+ P(D_{U,V} = 1)]$. In other words, the mixing coefficient is bounded by the probability that there exists a link between an individual in U and an individual in V.

Recall that we consider the case where $d(U,V) = O(d(i,j))$ for all $(i,j) \in C_{U,V}$.

Since there are at most $4kl$ such pairs, where $|U| = k$ and $|V| = l$, the second bound is bounded by:

$$(16kl) \sup_{W_m \in \mathbb{W}_m} \sup_{\{i \in S_m^U, j \in S_m^V\}} P(w_{ij} = 1 | W_{m,-ij}).$$

where $d(U,V) = O(d(i,j)$.

It is then sufficient to show that, as $d(i,j) \to \infty$, we have:

$$\begin{aligned} \sup_{A^U, B^V} |P(A^U | B^V) - P(A^U)| &\leq kl \sup_{W_m \in \mathbb{W}_m} \sup_{\{i \in S_m^U, j \in S_m^V\}} P(w_{ij} = 1 | W_{m,-ij}) \\ &= kl \sup_{W_m \in \mathbb{W}_m} \sup_{\{i \in S_m^U, j \in S_m^V\}} \frac{\exp\{w_{ij} h_{ij}(W_{m,-ij}; \theta)\}}{1 + \exp\{h_{ij}(W_{m,-ij}; \theta)\}} \end{aligned}$$

$$\overline{\phi}_{k,l}(d) \leq kl \sup_m \sup_{W_m \in \mathbb{W}_m} \sup_{\{i \in S_m^U, j \in S_m^V : d(i,j) > d\}} \frac{\exp\{w_{ij} h_{ij}(W_{m,-ij}; \theta)\}}{1 + \exp\{h_{ij}(W_{m,-ij}; \theta)\}}. \tag{8}$$

Under Assumption 8, we have that

$$\overline{\phi}_{k,l}(d) = O(d^{-2T+1-\epsilon})$$

for any $k, l < \infty$, which is sufficient for Assumptions 3.1 and 3.2.

The case of Assumption 3.3 is different. Consider the bound (8) for $k = 1$, and denote by s the only pair in U:

$$\begin{aligned} \overline{\phi}_{1,l}(d) &\leq l \sup_m \sup_{W_m \in \mathbb{W}_m} \sup_{\{i \in s, j \in S_m^V : d(i,j) > d\}} \frac{\exp\{w_{ij} h_{ij}(W_{m,-ij}; \theta)\}}{1 + \exp\{h_{ij}(W_{m,-ij}; \theta)\}} \\ &= l F[\tilde{h}(d)] \end{aligned}$$

where $F[\tilde{h}(d)] \equiv \sup_m \sup_{W_m \in \mathbb{W}_m} \sup_{\{i \in s, j \in S_m^V : d(i,j) > d\}} \frac{\exp\{w_{ij} h_{ij}(W_{m,-ij};\theta)\}}{1 + \exp\{h_{ij}(W_{m,-ij};\theta)\}}$.

Here, we implicitly assume that it is possible to have l pairs located at a distance of exactly d from s. For $d_0 > 0$, as l is increasing, this is impossible. From Jenish and Prucha (2009) (Lemma A.1(iii)), there can be at most Cd^{T-1}

pairs located at a distance in $[d, d+1)$, where C is a finite constant. The other pairs must be located at a distance of at least $d+1$. In fact, there can be at most $C(d+\iota)^{T-1}$ pairs located at a distance in $[d+\iota, d+\iota+1)$... and so on. We should therefore rewrite our bound as:

$$\overline{\phi}_{1,\infty}(d) \quad \leq \quad C \sum_{\iota=0}^{\infty} C_{d+\iota}(d+\iota)^{T-1} F[\tilde{h}(d+\iota)]$$

From Assumption 8, this can be rewritten as

$$\overline{\phi}_{1,\infty}(d) \quad \leq \quad C \sum_{\iota=0}^{\hat{\iota}(d)} (d+\iota)^{T-1} F[\tilde{h}(d+\iota)] + CK \sum_{\iota=\hat{\iota}(d)+1}^{\infty} (d+\iota)^{-T-\epsilon}$$

where K is finite and positive. The function $\hat{\iota}(d)$ maps to a finite and positive integer and represents the first value of ι such that $F[\tilde{h}(d+\iota)] \leq K(d+\iota)^{-2T+1-2\epsilon}$ holds. Note that $\hat{\iota}(d)$ is necessarily non-increasing in d, hence $\sup_d \hat{\iota}(d) = \hat{\iota}(0)$. Notice that $\hat{\iota}(0)$ is necessarily finite under Assumption 8.

We therefore have:

$$\overline{\phi}_{1,\infty}(d) \quad \leq \quad C \sum_{\iota=0}^{\hat{\iota}(0)} (d+\iota)^{T-1} F[\tilde{h}(d+\iota)] + CK \sum_{\iota=0}^{\infty} (d+\iota)^{-2T-2\epsilon}$$

The first term is a finite sum of sequences of d, each of which is (at least) $O(d^{-T-\epsilon})$, so the first term is $O(d^{-T-\epsilon})$. For a fixed d, $\sum_{\iota=0}^{\infty}(d+\iota)^{-T-\epsilon}$ converges. Then, the rest of the series converges to zero. This means that for every $\epsilon > 0$, there is a positive integer R such that for all $r \geq r' \geq R$ we have $\sum_{\iota=r}^{r'}(d+\iota)^{-T-\epsilon} < \epsilon$. Thus, $\sum_{\iota=0}^{\infty}(d+\iota)^{-T-\epsilon} \approx \sum_{\iota=0}^{R}(d+\iota)^{-T-\epsilon}$. Since every term of the last finite summation is $O(d^{-T-\epsilon})$, we have finally $\overline{\phi}_{1,\infty}(d) = O(d^{-T-\epsilon})$. This completes the proof. □

10.2. Proof of Proposition 3.

For simplicity, we omit the dependence on m. Let $h_s(W_{m,-s};\theta) = \Gamma_s'\theta$ where Γ_s is a vector of explanatory variables. We know from proposition 1 that the conditional distribution is well specified, i.e.

$$\mathbb{P}(w_s|W_{m,-s};\theta) = \frac{\exp\{w_s\Gamma_s'\theta\}}{1+\exp\{\Gamma_s'\theta\}}.$$

Therefore, if θ_0 is the true value of interest. The true conditional distribution takes the form $\mathbb{P}(w_s|W_{m,-s};\theta_0) = \frac{\exp\{w_s\Gamma_s'\theta_0\}}{1+\exp\{\Gamma_s'\theta_0\}}$.

The proof proceeds by a series of claims. Recall that we have: $q_s(w_s;\theta) = \ln\mathbb{P}(w_s|W_{m,-s};\theta) = w_s\Gamma_s'\theta - \ln[1+\exp\{\Gamma_s'\theta\}]$.

Claim 1. *If $\mathbb{E}[\Gamma_s\Gamma_s']$ is nonsingular, then $q_s(w_s;\theta) \neq q_s(w_s;\theta_0)$ for $\theta \neq \theta_0$ with positive probability under θ_0.*

Indeed, If $\mathbb{E}[\Gamma_s\Gamma_s']$ is nonsingular for $\theta \neq \theta_0$, $\mathbb{E}[\{\Gamma_s'(\theta-\theta_0)\}^2] = (\theta-\theta_0)'\mathbb{E}[\Gamma_s\Gamma_s'](\theta-\theta_0) > 0$. This implies that $\Gamma_s'(\theta-\theta_0) \neq 0$ and then $\Gamma_s'\theta \neq \Gamma_s'\theta_0$ with positive probability under θ_0. Because the function $\frac{e^x}{1+e^x}$ and $\frac{1}{1+e^x}$ and $\ln(x)$ are strictly monotonic, the result holds given that $q_s(w_s;\theta)$ can be written also as follows:

$$q_s(w_s;\theta) \quad = \quad w_s\ln\mathbb{P}(w_s=1|W_{m,-s};\theta) + (1-w_s)\ln\mathbb{P}(w_s=0|W_{m,-s};\theta).$$

Claim 2. *If $\mathbb{E}[|\Gamma_s\Gamma_s'|] < \infty$ then $\mathbb{E}[|q_s(w_s;\theta)|\big|W_{m,-s}] < \infty$.*

Indeed, because we can show that $\max(|x-\ln[1+e^x]|, |\ln[1+e^x]|) < 1+x^2$, we have then $\mathbb{E}[|q_s(w_s;\theta)|\big|W_{m,-s}] \leq 1+\theta'\mathbb{E}[\Gamma_s\Gamma_s']\theta$.

Claim 3. *If*

$$\big|\ln\mathbb{P}(w_s=1|W_{m,-s};\theta)\big|\mathbb{P}(w_s=1|W_{m,-s};\theta_0) + \\ \big|\ln\mathbb{P}(w_s=0|W_{m,-s};\theta)\big|\mathbb{P}(w_s=0|W_{m,-s};\theta_0) < \infty$$

for each $\theta \in \Theta$, then if $q_s(w_s;\theta) \neq q_s(w_s;\theta_0)$ for $\theta \neq \theta_0$ we have $\mathbb{E}[q_s(w_s;\theta)] < \mathbb{E}[q_s(w_s;\theta_0)]$.

Indeed, we have

$$\mathbb{E}[|q_s(w_s;\theta)|\big|W_{m,-s},\theta_0] = \ln\mathbb{P}(w_s=1|W_{m,-s};\theta)\mathbb{P}(w_s=1|W_{m,-s};\theta_0)$$
$$+\ln\mathbb{P}(w_s=0|W_{m,-s};\theta)\mathbb{P}(w_s=0|W_{m,-s};\theta_0).$$

Thus,

$$\begin{aligned}
&\mathbb{E}[|q_s(w_s;\theta)|\big|W_{m,-s},\theta_0] - \mathbb{E}[|q_s(w_s;\theta_0)|\big|W_{m,-s},\theta_0] \\
&= \ln\left\{\frac{\mathbb{P}(w_s=1|W_{m,-s};\theta)}{\mathbb{P}(w_s=1|W_{m,-s};\theta_0)}\right\}\mathbb{P}(w_s=1|W_{m,-s};\theta_0) + \ln\left\{\frac{\mathbb{P}(w_s=0|W_{m,-s};\theta)}{\mathbb{P}(w_s=0|W_{m,-s};\theta_0)}\right\}\mathbb{P}(w_s=0|W_{m,-s};\theta_0) \\
&< \ln\{\mathbb{P}(w_s=1|W_{m,-s};\theta) + \mathbb{P}(w_s=0|W_{m,-s};\theta)\} = 0.
\end{aligned}$$

The inequality holds using Jensen's Inequality for strictly concave function. Then, we have $\mathbb{E}[|q_s(w_s;\theta)|] - \mathbb{E}[|q_s(w_s;\theta_0)|] < 0$ by iterated expectation.

Claim 4. *If $\mathbb{E}[\Gamma_{s,m}\Gamma'_{s,m}]$ is finite and nonsingular and $\kappa(m) = O(n)$ then $\lim_{m\to\infty}\mathbb{E}[\mathcal{L}_m(\theta)] < \lim_{m\to\infty}\mathbb{E}[\mathcal{L}_m(\theta_0)]$ for $\theta \neq \theta_0$.*

Indeed, a direct implication of claims 1, 2, and 3 is that $\mathbb{E}[q_s(w_s;\theta)] < \mathbb{E}[q_s(w_s;\theta_0)]$. Then we have $\frac{1}{\kappa(m)}\sum_s^m \mathbb{E}[q_s(w_s;\theta)] < \frac{1}{\kappa(m)}\sum_s^m \mathbb{E}[q_s(w_s;\theta_0)]$. And then this remains valid at the limit if we choose the right order of $\kappa(m)$.

Bester et al. (2012)

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

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

where c_s is the group in which $s \in S$ is located. They show in their Proposition 1(ii) that one can use the t-statistic of the cluster-variance estimator, provided that it is rescaled by a factor of $\sqrt{C/C-1}$.

This approach has the advantage of being fast and easy to implement. In practice, the construction of those groups is not necessarily straightforward. Bester et al. (2012) recommend the use of a relatively small number of large groups. An important requirement, however, is a boundary condition that 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 $\mathcal{X}$.

Application: Distance and Convergence Rate

The Add Health database provides information on the (normalized) location of individuals' homes. Following the survey's definition, individuals can only be friends if they belong to the same community. This is similar to Example 5 . We therefore dropped the links between individuals from different neighbourhoods. Also, since the GPS coordinates are normalized to 100 for confidentiality issues, it does not go well with assumption 5. Indeed under assumption 5, (and, more generally, in reality) more populated communities should occupy more physical space. We therefore rescale the distances between pairs using $\sqrt{n_c}$ where n_c is the number of individuals in the community c. Indeed, from Jenish & Prucha (2009), lemma A.1 (ii), there are $O(d)^T$ individuals located in a ball of radius d. Here, $T = 2$, so an coherent scaling is such that: $n_c = d^2$, or $d = \sqrt{n_c}$.