Network Externalities in Bilateral Link Formation: What's the Value of Friends of Friends?*

Margherita Comola† and Amit Dekel‡

September 22, 2020

Abstract

This paper explores the existence of externalities from network architecture (so-called network externalities) in the context of bilateral link formation. It develops a structural estimation method for static games of network formation with incomplete information. We provide existence, consistency and asymptotic normality results for the proposed two-step estimator, and we document its performance through a simulation exercise. When the estimation method is applied to data on risk-sharing arrangements in a Tanzanian village, results indicate that indirect connections (i.e. friends of friends) matter. The estimated probability of proposing a link to a potential partner increases by 9% for any additional indirect connection provided.

Keywords: Bilateral Link Formation; Network Externalities; Incomplete Information; Risk-sharing

JEL codes: C45; D85; O12

*We thank Antonio Cabrales, Marcel Fafchamps, Rasa Karapandza, Philipp Ketz, Angelo Mele, Dotan Persitz and Angelo Secchi for their helpful comments, and Joachim De Weerdt for making the data available. We acknowledge the financial support of the EUR grant ANR-17-EURE-0001.

†University Paris-Saclay and Paris School of Economics: `margherita.comola@psemail.eu`

‡Paris School of Economics: `amit.dekel@psemail.eu`

1 Introduction

From its very first steps network theory has claimed that the formation of links depend strategically on the entire graph (Jackson and Wolinsky, 1996; Bala and Goyal, 2000). However, evidence based on experimental and observational data still lags behind, and important empirical questions such as the value of indirect connections and their decay rate in real-life situations remain largely unexplored.[1] The goal of this paper is to test whether network architecture helps predicting the formation of bilateral links. In particular, we investigate whether agents choose partners on the basis of their individual characteristics only, or whether indirect connections (i.e. friends of friends) also play a role in these decisions. To answer this question, we develop an estimation protocol suited for a large class of models of bilateral link formation with externalities from indirect connections. The validity of our estimation procedure crucially relies on two features of the underlying game: incomplete information and symmetric equilibrium. In our setting, agents play a simultaneous game of incomplete information where they form bilateral links on the basis of their beliefs. Assuming that these beliefs satisfy a number of regularity conditions (discussed in Section 2), the estimation strategy boils down to a two-step procedure where the first stage consistently estimates agents' beliefs about the emerging network, and the second stage estimates the role of network externalities.[2] This approach requires minimal assumptions on equilibrium selection, and it accommodates data from one single network. We provide existence, consistency and asymptotic normality results for the two-step estimator, and we design an extended simulation exercise to investigate its performance as sample size grows.

We illustrate the procedure using data on risk-sharing arrangements from the Tanzanian village of Nyakatoke.[3] Lacking access to formal insurance, most households in developing countries rely on informal risk-sharing arrangements in face of shocks such as health-related expenses, injuries, funerals and job losses. These arrangements have long captured the attention of economists, for several reasons. On the one hand, the prevalence of the phenomenon makes it of paramount importance for economic development.[4] On the other hand, such arrangements do not take place at the level of the entire society but among pairs of house-

[1] The study of cross-firm collaborative networks suggests that information flows are insignificant for indirect neighbors (Breschi and Lissoni, 2005; Singh, 2005). On the other hand, experimental evidence with dictator games shows that further-away connections are relevant and decay with the inverse of distance (Goeree et al., 2010).

[2] A two-step approach is also taken by König et al. (2014) and Leung (2015).

[3] These data have been the object of numerous articles (De Weerdt and Dercon, 2006; De Weerdt and Fafchamps, 2011; Vandenbossche and Demuynck, 2013; Comola and Fafchamps, 2014)

[4] Coate and Ravallion (1993), Townsend (1994), Udry (1994), Fafchamps and Lund (2003).

holds. By aggregating all these bilateral links, we obtain a graph which is among the most compelling applications of networks in economics.[5]

Much of the economic literature assumes that households voluntarily enter informal risk-sharing arrangements. This implies that links require the consent of the two parties involved, and link formation follows a bilateral process.[6] Nyakatoke data contain detailed information on self-declared risk-sharing links. All adult individuals were asked *"Can you give a list of people [...], who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?"*: this piece of information is used to draw the undirected village network and to investigate the role of network architecture in predicting link formation. Specifically, we test whether agents choose risk-sharing partners on the basis of their individual characteristics only or whether indirect connections (i.e. friends of friends) also play a role in these decisions. Risk-sharing arrangements constitute an intriguing setting where network externalities may combine positive and negative components: friends of friends are beneficial if they broaden social interactions but detrimental if there is competition for scarce resources. Results indicate that Nyakatoke villagers do evaluate potential partners' connections and that the positive component prevails. Our estimates suggest that for a given pair of potential partners ij, the probability that i proposes a link to j increases on average by 0.016 for any additional indirect connection j provides. This increase is sizeable, as it corresponds to approximately 9% of the average fitted probability of link proposal.

From an econometric standpoint, testing whether network architecture predicts link formation has proved to be a complex task. Our paper deals with the case where the researcher observes one single network at one single period and wants to include network covariates in the objective function of agents. In this scenario the structural equation can have multiple solutions (Bjorn and Vuong, 1984; Bresnahan and Reiss, 1991; Tamer, 2003), and the calculation may become intractable due to the combinatory complexity of networks. One solution is provided by the exponential random graph models where a dynamic meeting protocol acts as an equilibrium selection mechanism (Hsieh and Lee, 2016; König, 2016; Mele, 2017; Badev, 2018). Another solution is to condition on classes of models that replicate some observed topological patterns or to limit the degree to which other players can affect one's utility.[7]

[5] Risk-sharing networks have been studied from multiple angles, including the efficiency and sustainability of the resulting arrangements, the determinants of link formation and the structural properties of the network architecture (Genicot and Ray, 2003; Bloch et al., 2008; Jackson et al., 2012; Bramoullé and Kranton, 2007; Ambrus et al., 2014; Ambrus and Elliott, 2018).

[6] Most models of risk sharing and favor exchange assume that agents can refuse transactions that are against their self-interest (Kimball, 1988; Coate and Ravallion, 1993; Kocherlakota, 1996; Bloch et al., 2008; Jackson et al., 2012).

[7] Some papers identify structural parameters by the rate at which various sub-graphs are observed in the

Our paper is close to Leung (2015), who also relies on incomplete information to estimate a simultaneous game of link formation.[8] However, while Leung (2015) considers directed links, we focus on bilateral links. Accordingly, he relies on a non-cooperative Bayesian solution while our equilibrium concept resembles an incomplete-information version of pairwise stability requiring the mutual approval of both players.[9]

This paper's contribution to the existing literature is twofold. First, it provides a methodological infrastructure to estimate a large class of bilateral link formation games under the assumption of incomplete information. This generalizes partial observability models of bilateral link formation (Comola and Fafchamps, 2014) to include network covariates in the objective function of agents. Bilateral network formation models have proved difficult to estimate in presence of externalities because of the non-uniqueness of the resulting equilibrium. The procedure we propose provides a simpler alternative to structural models with complete information which achieve set identification (Miyauchi, 2016; Sheng, 2016; de Paula et al., 2018). Importantly, our procedure can accommodate network data from one single population, and it is computationally very parsimonious so that it suits large networks. As such, it can be of interest beyond the field of development, to all applied economists who want to identify preferences over network topology in a variety of contexts where link formation is 'naturally' bilateral, and the non-cooperative toolbox developed for directed networks (Badev, 2018; Mele, 2017; Leung, 2015) is not applicable. This is certainly the case for the exchange of goods or information in networked markets, digital platforms or industrial R&D networks. Second, our results also advance the knowledge of risk-sharing arrangements by providing first-hand evidence that indirect connections affect linking choices, while previous literature has focused mostly on documenting the number and characteristics of risk-sharing partners.[10]

The paper is organized as follows. Section 2 introduces the theoretical setting. Section 3 presents the estimation strategy. Section 4 describes the simulation exercise. Section 5 applies the estimation method to risk-sharing data from rural Tanzania. Section 6 concludes.

overall network (Chandrasekhar and Jackson, 2016) or by aggregating individuals into 'types' and assuming that agents have preferences only over the type of their partners (de Paula et al., 2018). Along the same lines, Boucher and Mourifie (2017) study a scenario where individual preferences display weak homophyly.

[8]For the estimation of social interaction models with incomplete information, see also Gilleskie and Zhang (2009) and Hoshino (2019).

[9]A network is said to be pairwise stable if no pair of players want to create a new link, and no player wants to sever an existing link (Jackson and Wolinsky, 1996).

[10]An exception is Krishnan and Sciubba (2009), who identify the common features of all equilibrium configurations in a model with negative network externalities and test these predictions against data on labor exchange arrangements in Ethiopia.

All proofs are relegated to Appendix B.

2 The Model

2.1 The game

Let $N = \{1, 2, ..., n\}$ be a set of agents who play to form an undirected network. For agent i, let $X_i = [X_{i,1}, ..., X_{i,q}]$ be a vector of individual attributes of dimension $[1 \times q]$ and $\epsilon_i = [\epsilon_{i,1}, ..., \epsilon_{i,i-1}, 0, \epsilon_{i,i+1}, ..., \epsilon_{i,n}]$ be a $[1 \times n]$ vector of shocks with all other agents. We assume that the shocks $\{\epsilon_{ij} \mid i, j \in N, i \neq j\}$ are independently drawn from the standard normal distribution (ϵ_{ij} does not necessarily equal ϵ_{ji}). The set of attributes vectors $X = \{X_1, ..., X_n\}$ is common knowledge, while the set of shocks is private information, i.e. only i knows ϵ_i.

Agents play a simultaneous-move game of bilateral network formation, where everyone announces independently the links he wishes to form, and the resulting network is given by the mutually announced links (Myerson, 1991). The action of agent i is represented by a binary vector of length n, where the jth entry ($j \neq i$) equals 1 if i proposes j to form a link and 0 otherwise.[11] Formally, we denote i's action by $S_i = [S_{i,1}, ..., S_{i,i-1}, 0, S_{i,i+1}, ..., S_{i,n}] \in \{0, 1\}^n$. The actions of all agents stacked on top of each other can be interpreted as an adjacency matrix of a directed network of link proposals:

$$S = \begin{bmatrix} 0 & S_{1,2} & \dots & S_{1,n} \\ S_{2,1} & 0 & \dots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,1} & \dots & S_{n,n-1} & 0 \end{bmatrix} \tag{1}$$

In turn, this proposal network gives rise to an undirected network G, where a link between two agents exists if and only if both propose to each other: $G_{ij} = S_{ij} \cdot S_{ji}$.

For a given network G, the utility of agent i is given by:

$$u_i(X, G; \theta_0) = \sum_{j \neq i} G_{ij} \cdot (v_{ij}(X, G_{-i}; \theta_0) + \epsilon_{ij}) \tag{2}$$

where G_{-i} indicates G with the i^{th} row and column deleted, $\theta_0 \in \Theta$ is a $[p \times 1]$ vector of parameters from a compact set Θ, and $v_{ij}(\cdot)$ is linear in θ_0 and anonymous (i.e. the agents'

[11] Since an agent cannot form a link with herself, the ith entry always equals 0.

labels are inconsequential). Estimating the parameters in θ_0 is the goal of the procedure described in Section 3.

Note that this functional form may accommodate different types of externalities from indirect connections, provided that the term $v_{ij}(X, G_{-i}; \theta_0) + \epsilon_{ij}$, which represents the marginal utility of i from having a link with j, depends on G only through G_{-i}. In other words, this separability condition requires that the decision of i to propose to j is independent from her linking status with all other agents $k \neq j$. The assumption of payoffs separability is frequently made for the estimation of discrete games within and beyond the field of networks (Leung, 2015; de Paula et al., 2018; Menzel, 2016; Lewbel and Tang, 2015). In our context this assumption rules out, for example, network formation models where agents collect benefits only through the shortest geodesic paths (because the lengths of paths between i and j depend on i's linking status with other agents) or models where agents derive utility from having friends in common (Jackson et al., 2012).

2.2 Equilibrium

In this incomplete information game agents do not know in advance which links will be formed in equilibrium, hence they cannot directly maximize their utility. Instead, they develop beliefs about the structure of the network that will be formed and maximize their expected utility given these beliefs. Let $\sigma^{S_{-i}}$ be a $[(n-1) \times n]$ matrix representing i's beliefs about the probabilities that each agent $j \neq i$ will propose to another agent $k \neq j$. Given these beliefs, i chooses to propose to j if and only if his expected marginal utility is positive:

$$S_{ij} = \mathbb{1}\Big\{\sigma_{ji}^{S_{-i}} \cdot (\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{S_{-i}}] + \epsilon_{ij}) \geq 0\Big\} \tag{3}$$

As $\sigma^{S_{-i}}$ represents beliefs, all its entries are weakly positive, thus Equation 3 is equivalent to:

$$S_{ij} = \mathbb{1}\Big\{\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{S_{-i}}] + \epsilon_{ij} \geq 0\Big\} \tag{4}$$

Since $v_{ij}(\cdot)$ depends on G_{-i} and not S_{-i}, we can in fact condition on a coarser set of beliefs which provide information only on G_{-i}. Let σ^G be a $[n \times n]$ matrix representing agents' common beliefs about the linking probabilities among all pairs of agents, and let $\sigma^{G_{-i}}$ denote the same matrix but with its i^{th} row and column deleted. Then:

$$S_{ij} = \mathbb{1}\Big\{\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}] + \epsilon_{ij} \geq 0\Big\} \tag{5}$$

Note that the decision rule we obtained in Equation (5) does not involve i's linking decisions with respect to agents other than j or his beliefs about proposals extended to him. This is illustrated in Figure 1. i's relevant beliefs are depicted as a complete weighted network on all agents besides him. The idea that i does not take into account beliefs about proposals extended to him is captured by the absence of directed edges pointing at him. The questions about whether to propose to each of the other agents, depicted by the dotted directed edges, are answered independently.

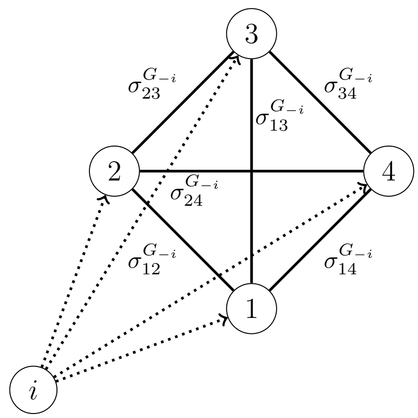


Figure 1: An example of proposal decisions

Given the decision rule of Equation 5, from the perspective of a different agent $k \neq i$, who knows X but does not know ϵ_i, the probability that i would propose to j depends on the distribution from which ϵ_{ij} is drawn:

$$P(S_{ij} = 1|X, \sigma^{G_{-i}}) = P\left(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}] + \epsilon_{ij} \geq 0\right) \tag{6}$$

$$= \Phi\left(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}]\right) \tag{7}$$

where Φ is the CDF of the standard normal distribution. Given that the ϵ values are independent, the probability that a link between i and j is formed (conditional on X and σ^G) is the product of the two proposal probabilities:

$$P(G_{ij} = 1|X, \sigma^G) = \underbrace{\Phi\left(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}]\right)}_{i \text{ proposes to } j} \cdot \underbrace{\Phi\left(\mathbb{E}[v_{ji}(X, G_{-j}; \theta_0)|X, \sigma^{G_{-j}}]\right)}_{j \text{ proposes to } i} \tag{8}$$

This naturally leads to a stability concept. A belief matrix σ^G is said to be an equilibrium if it satisfies the following condition for all $i, j \in N$:

$$\sigma_{ij}^G = \Phi(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|\sigma^{G_{-i}}]) \cdot \Phi(\mathbb{E}[v_{ji}(X, G_{-j}; \theta_0)|\sigma^{G_{-j}}]) \tag{9}$$

A belief matrix constitutes an equilibrium if for all i and j, the probability that all agents $k \neq i, j$ assign to the event that i and j will be linked in G equals the true probability of this event given that i and j optimally respond to $\sigma^{G_{-i}}$ and $\sigma^{G_{-j}}$, respectively. Using this notion of equilibrium belief matrix we define an equilibrium network. A network is said to be an equilibrium network if it satisfies the following condition for all $i, j \in N$:

$$G_{ij} = \underbrace{\mathbb{1}\left\{\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}] + \epsilon_{ij} \geq 0\right\}}_{S_{ij}} \underbrace{\mathbb{1}\left\{\mathbb{E}[v_{ji}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}] + \epsilon_{ji} \geq 0\right\}}_{S_{ji}} \tag{10}$$

and σ^G is an equilibrium matrix. Since S_{ij} (S_{ji}) takes the value 1 if and only if the expected marginal utility of i (j) from linking with j (i) is positive, this means that G is an equilibrium network if for all links that exist the expected marginal utility of both agents is positive, and for all links that do not exist the expected marginal utility of at least one of the agents is negative (where the expectation is taken with respect to some equilibrium belief matrix). Our stability concept therefore parallels the Pairwise stability solution concept (Jackson and Wolinsky, 1996), only that it is adjusted to the incomplete information settings in the sense that the conditions for Pairwise stability are required to hold in expectation, and an additional restriction is imposed over beliefs.[12]

We focus on symmetric equilibria, defined as equilibria in which all pairs of agents that are observationally equivalent have the same linking probabilities. Formally, an equilibrium σ^G is symmetric if for all $i, j \neq k, l \in N$:

$$(X_i = X_k \textit{ and } X_j = X_l) \textit{ or } (X_i = X_l \textit{ and } X_j = X_k) \implies \sigma_{ij}^G = \sigma_{kl}^G \tag{11}$$

Figure 2 illustrates the idea of a symmetric belief matrix. Agents in this network have a single binary attribute – being either black or white – depicted by the colors of the nodes. Beliefs are depicted by weights on edges and their values by their color (i.e. all red beliefs equal each other, and all blue beliefs equal each other). All pairs consisting of two black agents have the same σ^G value (red), and the same holds for pairs of white and black agents

[12] A similar stability concept can be found in Song and van der Schaar (2015) under the name 'stable equilibrium'.

(blue) and pairs of two white agents (green). The described beliefs are therefore symmetric.

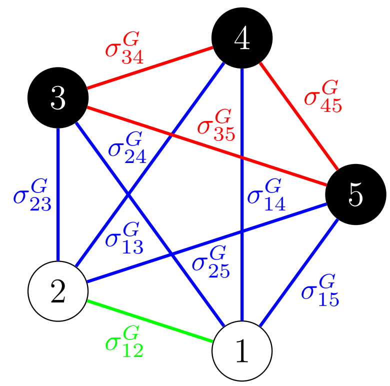


Figure 2: Example of a symmetric belief matrix

The following proposition establishes the existence of a symmetric equilibrium.

Proposition 1 (Existence)**.** *For any discrete* X *and* θ_0*, there exists a symmetric equilibrium.*

In what follows we concentrate on symmetric equilibria and to simplify the terminology we refer to a symmetric equilibrium σ^G as "equilibrium beliefs".

2.3 Example

Consider the case where 3 agents have one binary attribute X_i, and their utility function is as follows:

$$v_{ij}(X, G_{-i}; \theta_0) = \theta_1 + \theta_2 X_i + \theta_3 |X_i - X_j| + \theta_4 \frac{1}{n-1} \sum_{k \neq i} G_{jk} \tag{12}$$

with $\theta_0 = [-1, 1, -0.5, 1]'$. The term $|X_i - X_j|$ represents a measure of similarity between i and j. It thus accounts for homophily. The term $\frac{1}{n-1} \sum_{k \neq i} G_{jk}$ represents the average number of indirect connections (i.e. paths of length 2) that i gains by forming a link with j. It thus accounts for externalities from the network topology.

Columns 1 and 2 in Table 1 present all possible ordered pairs in the 3-agent network. Columns 3 and 4 report the binary attributes of agent i and j respectively. Column 5 reports $|X_i - X_j|$. The third term in the utility function $\frac{1}{n-1} \sum_{k \neq i} G_{jk}$ depends on the network structure G. Its expected value therefore depends on the beliefs about the network structure σ^G.

Let us consider a given set of beliefs which are reported in column 6. Column 7 uses these beliefs to compute $\frac{1}{n-1}\sum_{k\neq i}\sigma_{jk}^{G_{-i}}$. Using columns 3, 5 and 7 and the functional form we can now compute the expected value of v_{ij} for all pairs of agents. This is reported in column 8. Now, given that the ϵ_{ij} values are drawn independently from the standard normal distribution, the probability that i would propose to j (that is, that $\mathbb{E}[v_{ij}] + \epsilon_{ij} \geq 0$) is $\Phi(\mathbb{E}[v_{ij}])$. This is reported in column 9. Finally, the probability that a link exists in G is the product of the proposal probabilities of the two agents involved. This is reported in column 10.

1	2	3	4	5	6	7	8	9	10
i	j	X_i	X_j	$\lvert X_i - X_j\rvert$	σ^G	$\frac{1}{n-1}\sum_{k\neq i}\sigma_{jk}^{G_{-i}}$	$\mathbb{E}[v_{ij}]$	$\Phi(\mathbb{E}[v_{ij}])$	$\Phi(\mathbb{E}[v_{ij}])\cdot\Phi(\mathbb{E}[v_{ji}])$
1	2	0	1	1	0.027	$0.5\cdot 0.255$	-1.3725	0.0850	0.027
1	3	0	1	1	0.027	$0.5\cdot 0.255$	-1.3725	0.0850	0.027
2	1	1	0	1	0.027	$0.5\cdot 0.027$	-0.4865	0.3133	0.027
2	3	1	1	0	0.255	$0.5\cdot 0.027$	0.0135	0.5054	0.255
3	1	1	0	1	0.027	$0.5\cdot 0.027$	-0.4865	0.3133	0.027
3	2	1	1	0	0.255	$0.5\cdot 0.027$	0.0135	0.5054	0.255

Table 1: Example

Note that in this example $\sigma_{ij}^G = \Phi(\mathbb{E}[v_{ij}])\Phi(\mathbb{E}[v_{ji}])$ for all i and $j \neq i$. This means that the beliefs σ^G in column 6 are equilibrium beliefs. Also note that all pairs of agents which are observationally equivalent have the same linking probabilities, e.g. the pairs $\{1,2\}$ and $\{1,3\}$ have the same linking probability under σ^G. This means that the beliefs σ^G are symmetric.

2.4 Equilibrium Selection

As described by equation 8, for a given set of beliefs the model implies a unique probability for each link to be formed. Recall that linking decisions are based on beliefs and the ϵ values are independent from one another. This implies that, conditional on beliefs, the model also predicts a unique probability distribution over networks. In particular, the probability that a given network G is formed under the belief matrix σ^G is given by the product of the probabilities of the linking statuses among all pairs of agents under that belief matrix:

$$P(G|X,\sigma^G) = \prod_{i,j>i}^{n}\left[\left(P(G_{ij}=1|X,\sigma^{G_{-i}})\right)^{G_{ij}}\cdot\left(P(G_{ij}=0|X,\sigma^{G_{-i}})\right)^{1-G_{ij}}\right] \tag{13}$$

Hence, conditional on beliefs the likelihood is well defined, that is, it is a function that provides a unique output for any input (X, θ_0, σ^G). An econometric model is defined as "complete" if for any (X, θ_0) the likelihood is well defined (Tamer, 2003). If it were the case that for any X and θ_0 our model would have implied, for instance, a *unique* symmetric equilibrium, then imposing the assumption that agents play a symmetric equilibrium would have yielded a complete model. While Proposition 1 establishes the existence of a symmetric equilibrium for any X and θ_0, it does not guarantee its uniqueness. Moreover, it can be shown that for some values of X and θ_0 multiplicity of symmetric equilibria indeed occurs.[13]

The problem of incompleteness arises from the fact that the model we specified thus far is mute about the process by which an equilibrium is selected when multiplicity occurs. To complete the model, on top of assuming that agents play a symmetric equilibrium, we need to specify an equilibrium selection mechanism. Let $\omega(X, \theta_0)$ denote the set of symmetric equilibria for X and θ_0 and let Σ^G denote its corresponding random variable. We assume that each $\sigma^G \in \omega(X, \theta_0)$ is selected with probability $P(\Sigma^G = \sigma^G)$ according to some probability distribution. This assumption can be interpreted as a public signal, which, with probability $P(\Sigma^G = \sigma^G)$ informs agents that σ^G is selected. Given this selection mechanism the likelihood is well defined for each X and θ_0 (without conditioning on a particular σ^G):

$$P(G|X) = \sum_{\sigma^G \in \omega(X,\theta_0)} P(\sigma^G) \cdot \prod_{i,j>i}^{n} \Bigg[\Big(P(G_{ij} = 1 | X, \sigma^{G_{-i}}) \Big)^{G_{ij}} \times \Big(P(G_{ij} = 0 | X, \sigma^{G_{-i}}) \Big)^{1-G_{ij}} \Bigg] \tag{14}$$

Note that we do not make assumptions on the shape of the probability distribution over symmetric equilibria. We therefore do not impose any restrictions on the network formation game. The assumption above merely formalizes the process by which equilibria are selected when multiplicity occurs. An alternative, more substantial, assumption that is commonly made when the econometrician observes many repetitions of the game ("many-market asymptotics") is that the probability distribution over equilibria is degenerate. This guarantees that the equilibrium being played in all repetitions of the game is the same. We are able to avoid this assumption and achieve point identification in a large-network case ("large market asymptotics") thanks to the assumption that only *symmetric* equilibria are allowed to be selected. This will become clear in Section 3.2, where we discuss ways to recover beliefs from the data.

[13] Consider for instance the example in Subsection 2.3 with $\theta_0 = [-1, 0, 0, 7]$. There exists a symmetric equilibrium in which $\sigma_{ij} = 0.037$ for all i and $j \neq i$, as well as one in which $\sigma_{ij} = 0.986$ for all i and $j \neq i$.

3 Estimation

Imagine we observe a single network G and agents' attributes X. Let us assume that G is formed according to the model specified above, that is, the network results from all agents behaving optimally given the symmetric equilibrium belief σ^G and their realization of the error terms ϵ_i that we do not observe. Our goal is to estimate and conduct inference on the true parameter vector θ_0. In what follows we describe the building blocks of our procedure.

3.1 Log-likelihood function

Let us denote by δ_{ij} a function that takes X_i, X_j and returns a vector of covariates of dimension $[1 \times (p-k)]$ (e.g. i's attributes and the distance between i and j's attributes, in the example above). Denote by γ_{ij} a function that takes i's beliefs about the emerging network (possibly together with X) and returns a vector of covariates of dimension $[1 \times k]$ (e.g. the number of length-two paths i gains from linking with j, in the example above). To facilitate an intercept, assume that δ_{ij} always returns 1 as a first element. We call the first type of covariates 'exogenous' as they do not depend on the network structure, and the second type 'endogenous', as they do. $v_{ij}(\cdot)$ is a linear function of the exogenous and endogenous covariates:

$$v_{ij}(X, G_{-i}; \theta_0) = [\delta_{ij}(X_i, X_j), \gamma_{ij}(X, G_{-i})] \cdot \theta_0 \tag{15}$$

The expected value of v_{ij} conditional on X and the event that σ^G is the selected equilibrium is therefore:

$$\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}] = [\delta_{ij}(X_i, X_j), \gamma_{ij}(X, \sigma^{G_{-i}})] \cdot \theta_0 \tag{16}$$

Suppressing some of the input arguments, we can now rewrite Equation (8) as:

$$P(G_{ij} = 1|X, \sigma^G) = \Phi([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta_0) \cdot \Phi([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta_0) \tag{17}$$

Since $\{\epsilon_{ij}|i, j \in N, i \neq j\}$ are drawn independently from one another, conditional on X and the event that σ^G is selected, the likelihood of observing a network G is:

$$L(\theta, \sigma^G) = \prod_{i,j>i}^{n} \Bigg[\Big(\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\big) \cdot \Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\big) \Big)^{G_{ij}} \\ \times \Big(1 - \Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\big) \cdot \Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\big) \Big)^{1-G_{ij}} \Bigg] \tag{18}$$

By taking the log of this expression and dividing by the number of observations we obtain the following log-likelihood function:

$$l(\theta, \sigma^G) = \frac{2}{n(n-1)} \sum_{i,j>i}^{n} \Bigg[\Big(G_{ij} \cdot \log \Big(\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\big) \cdot \Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\big) \Big) \Big) \\ + \Big(\Big(1 - G_{ij}\Big) \cdot \log \Big(1 - \Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\big) \cdot \Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\big) \Big) \Big) \Bigg] \tag{19}$$

This function depends on the unobserved beliefs σ^G. We therefore cannot directly proceed to maximize it with respect to θ. Instead, we follow a two-step procedure, where in the first stage we consistently estimate the symmetric equilibrium beliefs (Subsection 3.2), and in the second stage we plug the estimated beliefs into the log-likelihood function to recover the estimands (Subsection 3.3).

Two comments about the log-likelihood function are in place. First, note that if we rule out endogenous covariates from the marginal utility the model boils down to a bivariate probit with partial observability (Poirier, 1980).[14] This model has been used to model bilateral link formation in the absence of network externalities (Comola and Fafchamps, 2014). Second, note that under uniqueness of equilibria, resorting to recovering σ^G from the data is not strictly necessary. Instead, we could analytically calculate the unique equilibrium beliefs for any candidate θ that is being considered by the optimization algorithm and evaluate the log-likelihood function at these beliefs. This method breaks down under multiplicity: even if the task of calculating all equilibria for a candidate θ is feasible, we cannot determine which is played in the data without imposing additional restrictions.

[14] Partial observability occurs when a positive outcome for one response variable is only observed if the other response variable is also positive. In our context the decision rules of the two agents can be interpreted as two partially observed latent binary response variables, where the θs are by construction the same across the two equations.

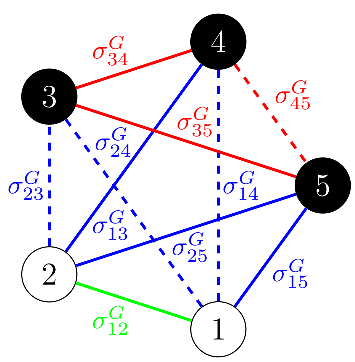
3.2 Estimating Beliefs

Under the assumption that beliefs satisfy the symmetric equilibrium condition, producing a consistent estimate of the beliefs $\hat{\sigma}^G$ is straightforward. Consider a set of observationally equivalent pairs of agents. In a symmetric equilibrium, the belief that any of these pairs are linked is identical (due to symmetry) and correct (since it is an equilibrium). Thus, the proportion of pairs within this set that are linked in the observed network is a consistent estimator for the belief that any of the pairs in the set are linked. In the case of discrete attributes, the estimator for the belief that i and j are linked $\hat{\sigma}_{ij}^G$ is defined as:

$$\hat{\sigma}_{ij}^G \equiv \frac{\sum_{l,k>l} \Big[G_{kl} \cdot \mathbb{1}\big\{ (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k) \big\} \Big]}{\sum_{l,k>l} \Big[\mathbb{1}\big\{ (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k) \big\} \Big]} \tag{20}$$

Proposition 2. *When X is discrete and has finite support, and the selected equilibrium is symmetric, $\hat{\sigma}_{ij}^G$ is consistent for σ_{ij}^G for all $i, j \in N$ such that $i \neq j$.*

Figure 3 provides an example of how this estimator is calculated. As in figure 2, the colors of the agents depict their one dimensional binary attribute (being either black or white) and the colors of the edges and weights illustrate which pairs of agents have identical ex-ante linking probabilities (due to symmetry). The type of the edges illustrate which links are realized in the observed network - full lines describe realized links and dashed lines describe unrealized ones. The $\hat{\sigma}^G$ matrix presents the estimated beliefs. Concentrating on the black pairs, for instance, since two out of the three potential links between this type of pairs are realized we estimate the belief that these pairs are linked by $\frac{2}{3}$.



$$\hat{\sigma}^G = \begin{bmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{2}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{2}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{2}{3} & \frac{2}{3} & 0 \end{bmatrix}$$

Figure 3: Example of beliefs estimation

To get a better understanding of the advantages of this estimation method, it is useful to contrast our "large-market" framework with an alternative "many-markets" framework. Assume we were to observe many repetitions of the game over a constant set of agents ("many-markets"). The same pairs of agents are expected to have the same ex-ante linking probabilities across games, regardless of anonymity of preferences or symmetry of beliefs. As mentioned in Subsection 2.4, this only holds when agents are guaranteed to play the same equilibrium across games, which can be obtained by assuming a degenerate equilibrium selection mechanism. Thus, the proportion of games in which a given pair is linked gives a consistent estimate for the belief that this pair would be linked as the number of games increases to infinity. In our context of "large-market" framework we can relax the assumption that the the equilibrium selection mechanism is degenerate and estimate symmetric beliefs from one single network realization. This broadens the applicability of our estimator, as many network datasets depict one population (Goyal et al., 2006; Mele, 2017).[15]

Two concerns that this estimator might raise, however, are the following. First, since the denominator sums up pairs that are exactly identical, it is only applicable to cases where all attributes in X are discrete (for continuous attributes the denominator evaluates to zero). Second, since the estimator divides the set of observations into bins of identical pairs of agents, when the sample size is small, the number of attributes is high, and their support is large, we run the risk of not having enough observations within each bin to obtain meaningful estimates. Both of these concerns are addressed in Appendix A. Subsection A.1 allows for the inclusion of continuous attributes, thereby resolving the first concern. Subsection A.2 discusses smoothing of discrete variables, which helps alleviating the second.

3.3 Estimating Preferences

Once $\hat{\sigma}^G$ is computed, plugging it into Equation (19) and maximizing with respect to θ yields our estimates $\hat{\theta}$ of θ_0. Since $\hat{\sigma}^G$ is consistent $\hat{\theta}$ is also consistent under standard regularity conditions. Below we state the consistency and asymptotic normality results for the second-stage estimator which is demonstrated in the Appendix.

Proposition 3 (Consistency)**.** *Under standard regularity conditions, $\hat{\theta}$ is consistent for θ_0.*

Since the endogenous covariates are computed based on the estimated beliefs rather than the true ones, standard errors should be adjusted. Proposition 4 shows how this can be done

[15] Our estimation procedure also carries over to the case of multiple networks, provided that beliefs are separately estimated for each sub-population.

under the innocuous assumption that the aggregate values of the true endogenous covariates and the estimated ones are identical.

Proposition 4 (Asymptotic Normality)**.** *Assume the endogenous covariates satisfy*

$$\sum_{i,j\neq i}\gamma_{ij}(X,G_{-i})=\sum_{i,j\neq i}\gamma_{ij}(X,\hat{\sigma}^{G_{-i}}). \tag{21}$$

Let γ_{ij}^0 denote the output of $\gamma_{ij}(X,\sigma^G)$ and γ_0 denote the set of γ_{ij}^0 for all i,j. Then,

$$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}-\theta_0)\xrightarrow{d}N(0,[V(\gamma_0,\theta_0)]^{-1}\Psi(\gamma_0,\theta_0,G)[V(\gamma_0,\theta_0)]^{-1}) \tag{22}$$

where V and Ψ are defined as in Equations 67 and 86 in the Appendix.

As mentioned above, proposition 4 relies on the endogenous covariates satisfying condition 21. Lemma 1 proves this property for endogenous covariates of the form $\frac{1}{n-1}\sum_{k\neq i}G_{jk}\cdot\mu(X_k)$, where $\mu(X_k)$ represents some weighting function of agent k's attributes. $\mu(\cdot)$ captures any sort of observed attributes that agents might care about in their indirect friends. For instance, when deciding to form a link with someone, they may care not only about the amount of this potential partner's friends but also about their wealth. The illustration of Section 5 makes use of covariates of this form.

Lemma 1. *Let $\gamma_{ij}(X,G_{-i})\equiv\frac{1}{n-1}\sum_{k\neq i}G_{jk}\cdot\mu(X_k)$, where $\mu(X_k)$ is some weighting function of the attributes of agent k, then, for any G_{-i}, condition 21 holds.*

4 Simulations

We now describe the simulation exercise we designed to evaluate the asymptotic performance of the estimator for medium to large networks. First we describe the data generating process, then the estimation results.

4.1 Data Generating Process

For a given number of agents n with a one-dimensional attribute vector X_i, we posit a data generating process of the form:

$$X_i\sim U\{0,1,2,3,4\},\quad\forall i \tag{23}$$

$$\epsilon_{ij} \sim N(0,1), \quad \forall i,j, \quad i \neq j \tag{24}$$

$$v_{ij} = \theta_1 + \theta_2 X_i + \theta_3 |X_i - X_j| + \theta_4 \frac{1}{n-1} \sum_{k \neq i} G_{jk}, \quad \forall i,j, \quad i \neq j \tag{25}$$

$$\theta_0 = [-1.6, 0.5, -0.1, 1]' \tag{26}$$

where $\frac{1}{n-1}\sum_{k\neq i} G_{jk}$ represents the average number of indirect friends that j grants access to, as in the example of Section 2.3. θ_0 is set so that the utility function is not dominated by its deterministic component, i.e. so that proposal decisions are sensitive to ϵ_{ij}.

The data generating process consists of three steps: first we draw the attribute X_i for all i. Second we find a corresponding symmetric equilibrium σ^G. We use an algorithm that starts from a randomly drawn belief matrix, computes the corresponding linking probabilities, and updates beliefs accordingly until convergence is achieved. Algorithm 1 describes the process in more detail.[16]

Algorithm 1: Search Algorithm

1 Generate a random belief matrix σ^G
2 Calculate the matrix of linking probabilities L, given σ^G, X and θ_0:
3 $\quad L_{ij} = L_{ji} = \Phi(\mathbb{E}[v_{ij}(X, \sigma^G, \theta_0)]) \cdot \Phi(\mathbb{E}[v_{ji}(X, \sigma^G, \theta_0)])$
4 If $\sigma^G \not\approx L$:
5 $\quad$ Re-assign $\sigma^G = L$ and go back to line 2
6 Else:
7 $\quad$ Return σ^G

As a third step we draw the ϵ_{ij} values and construct a network realization G according to the following rule: a link in G exists if and only if the realization of ϵ_{ij} and ϵ_{ji} are such that $v_{ij}(X, \sigma^G, \theta_0) + \epsilon_{ij} \geq 0$ and $v_{ji}(X, \sigma^G, \theta_0) + \epsilon_{ji} \geq 0$.

For each $n \in \{100, 500, 900\}$ we generate 500 networks according to the procedure above. The networks that result from this process exhibit many commonly observed characteristics of real-world networks: the average geodesic distance between connected agents is low ($\approx$ 2.1); the clustering coefficient is high compared to the linking probability of a comparable Poisson random network (≈ 0.25 vs. ≈ 0.1); and the degree distribution is positively skewed. The average degrees are approximately 10, 53 and 95 for $n \in \{100, 500, 900\}$ respectively.

[16] For further details see Rabinovich et al. (2013).

4.2 Simulation Results

In the estimation step, for each simulation draw we use the realized network G and the agents attributes X (but not the error terms and beliefs) to estimate σ^G (as explained in Section 3.2). Then we maximize Equation (19) by replacing σ^G with $\hat{\sigma}^G$ to obtain $\hat{\theta}$.

Table 2 presents histograms of the obtained $\hat{\theta}$ values. The values of the true coefficients are depicted by the vertical lines at the center of each sub-figure. As n increases the distributions of the estimated values become increasingly tight around the true values. This shows that the estimators are consistent.

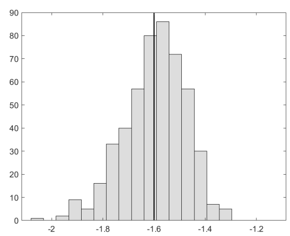
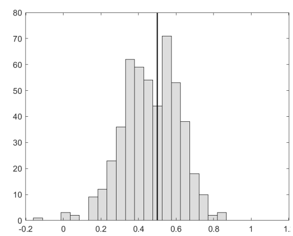
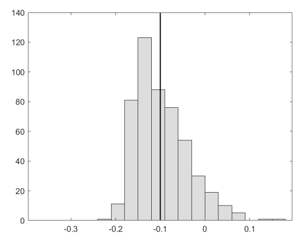
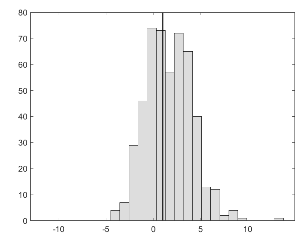
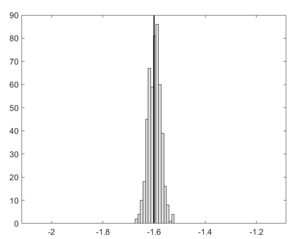
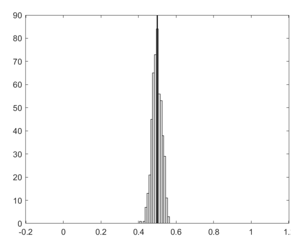
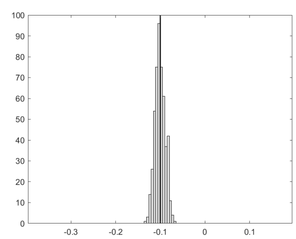
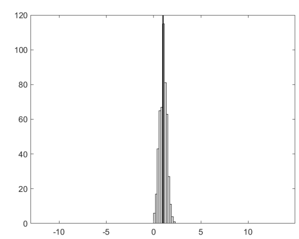
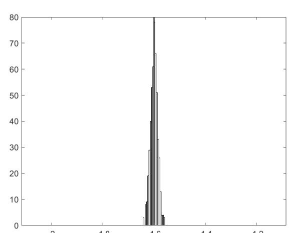
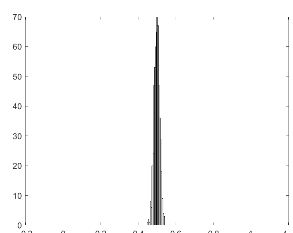
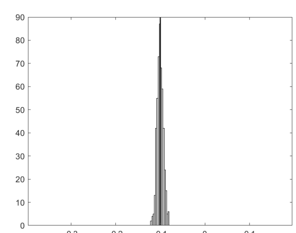
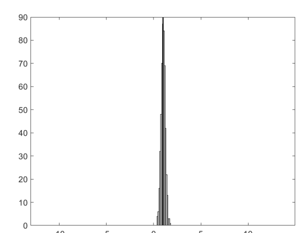
n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
100				
500				
900				

Table 2: Consistency

Note: The table reports histograms of estimated coefficients. The true values of the coefficients are depicted by the vertical line at the center of each sub-figure.

Table 3 presents the fitted Kernel distributions of $\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}-\theta_0)$ over all 500 iterations (in dashed lines) as well as true normal distributions with mean zero and variance $V^{-1}\Psi V^{-1}$ (in full lines). As n increases, the dashed lines converge to the full lines. This shows that the estimators are asymptotically normal.

n	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_1 - \theta_1)$	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_2 - \theta_2)$	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_3 - \theta_3)$	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_4 - \theta_4)$
100				
500				
900				

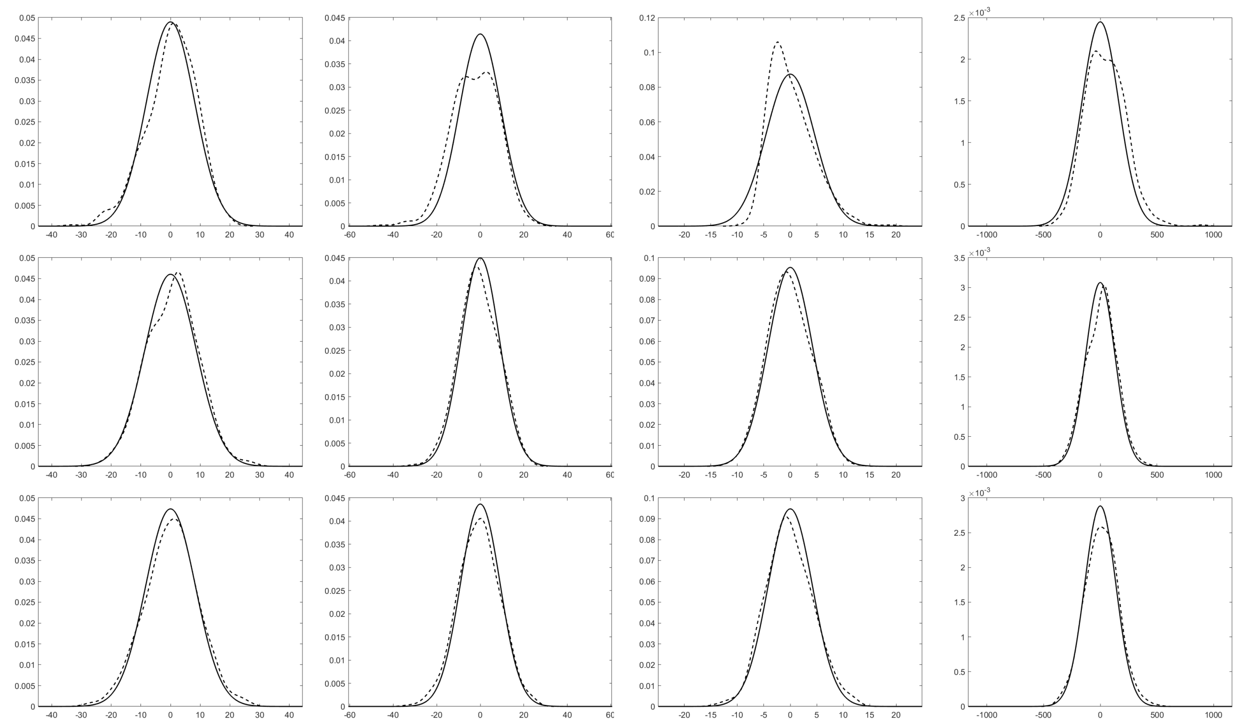


Table 3: Asymptotic normality

Note: The dashed lines depict the fitted Kernel distributions of $\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0)$. The full lines depict true normal distributions with mean 0 and variance $V^{-1}\Psi V^{-1}$.

5 Empirical Illustration

5.1 Data Description

We use data on the risk sharing network of Nyakatoke, a small village in the Buboka rural district of Tanzania. Rural villages are an appropriate setting to study network formation, because the population can be entirely surveyed and several confounding effects (such as spatial and informational barriers) can be reasonably ruled out. The village of Nyakatoke consists of 119 households which have been interviewed in five regular intervals from February to December 2000. The data contains information on households' demographics, wealth, income sources and income shocks, transfers and risk-sharing links. At the time of the study, the village of Nyakatoke is isolated (the few unpaved roads leading to the village are hardly passable during rains), densely inhabited (90% of households live within a distance of 1 kilometer from each other) and relatively poor (consumption for adult equivalent is less than 2 US$ a week, and average food share in consumption is 77%). Households earn most of their income from agricultural activities, especially the cultivation of coffee and banana; other sources of income are rare and off-farming activities are mostly considered supplementary to farming.

During the first survey round all respondents were asked *'Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?'*.[17] This question was specifically intended to capture undirected, mutual flows of assistance between respondents.[18] Thus, it is naturally fit to estimate bilateral link formation model, in line with most theoretical work on the voluntary nature of risk-sharing arrangements (Bloch et al., 2008; Jackson et al., 2012). However, this approach leaves open the issue of non-reciprocated links, i.e. situations where i mentions j as a partner but j does not mention i in return. Mis-reporting is a very common occurrence in network data depicting supposedly mutual relationships such as risk-sharing, goods exchanges or friendship, and Nyakatyoke is not an exception (Comola and Fafchamps, 2017). Since the problem of mis-reporting is beyond the scope of the current paper, we assume that a bilateral link exists whenever it is declared by at least one of the households involved. This is equivalent to assuming that all observed discordances are due to under-reporting, which is the most common stand taken by the network literature in this context (Fafchamps and Lund, 2003).[19] The resulting risk-sharing network of Nyakatoke, which consists of 490 links among $(119 \cdot 118)/2 = 7021$ undirected household dyads, is depicted in Figure 4. This network displays a mean geodesic distance of 2.5 steps and a maximum geodesic distance of 5 steps. No household is isolated, and the average degree is 8.2. The network exhibits all the empirical regularities of large social networks.[20]

[17] Respondents also mention partners who live outside the village (34% of all partners), but we are obliged to omit them from the analysis due to data limitation. No effort to track them was made in the initial data collection strategy, thus no information about them is available.

[18] This interpretation is consistent with the phrasing of the question and the way survey respondents have understood it. This question was first piloted in the Philippines (Fafchamps and Lund, 2003) and subsequently adopted in the Nyakatoke survey, because respondents understand it and are willing to answer. Other survey questions on directed flows were tried during the pilots, for instance drawing a distinction between people which respondents would help and people which respondents would seek help from. But respondents were confused by this distinction, which they perceived as non-existent, and complained they are asked the same question twice.

[19] To test the robustness of our results, we have also replicated the analysis of Table 4 under the alternative assumption of over-reporting. Results (available upon request) turn out qualitatively similar.

[20] The Nyakatoke network has a unique component covering the entire population, the diameter is in the order of $\log(n)$ and the clustering coefficient (which measures the tendency of linked nodes to have a friend in common) is 7 times larger than in a randomly generated Poisson network with similar characteristics.

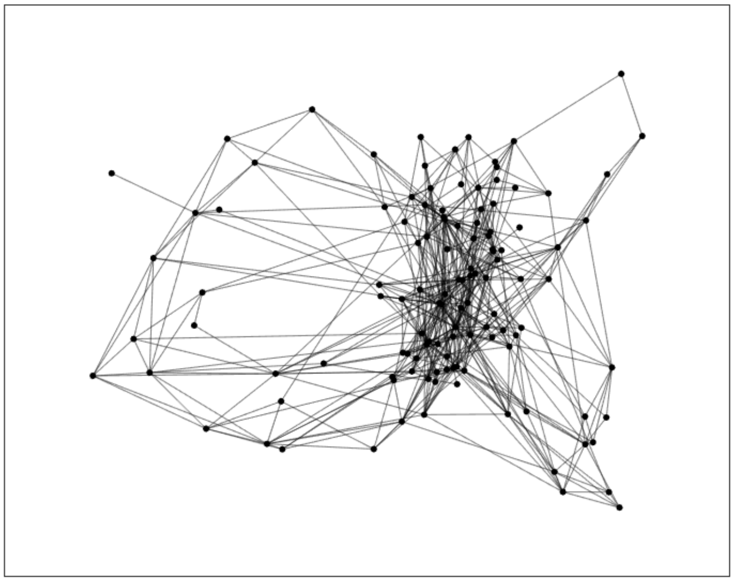


Figure 4: The risk-sharing network of Nyakatoke

5.2 Estimation Results

We now illustrate the estimation procedure described in Section 3 using the Nyakatoke data. We take the household as a unit of observation ($n = 119$) and we include as covariates: a constant, the geographical distance between households (in meters), the wealth of j,[21] three types of homophily regressors, and two types of endogenous regressors. The homophily regressors are binary variables that take the value 1 if i and j belong to the same family,[22] same clan[23] or same religion[24] respectively. The endogenous regressors are the number of j's friends ($\sum_{k\neq i} G_{jk}$) and the total wealth of j's friends ($\sum_{k\neq i} G_{jk} \cdot \text{Wealth}_k$).[25]

We run the first stage using the individual attributes that are used in the second stage (Wealth_j), as well as those implied by the relational attributes in the second stage (Family_i,

[21] The wealth of a household is defined as the total monetary value of its land and livestock assets (1 unit = 100,000 Tanzanian shillings). Data on land were originally in acres and were transformed in monetary equivalent with a conversion rate of 300,000 *tzs* for 1 acre which reflects average local prices in 2000. For international comparisons, the exchange rate in 2000 was 1 US dollar for 800 *tzs*.

[22] Two households i and j are said to belong to the same family if there is some blood relation between at least one of the members of i and at least one of the members of j.

[23] There are 26 clans in Nyakatoke. 10 of them have only one household.

[24] There are three religions in Nyakatoke: Roman Catholic (49 households), Lutheran (46 households) and Muslim (24 households).

[25] For presentation purposes we do not re-scale these variables in the results of Table 4. In fact, the normalization is only needed to facilitate the asymptotic case where n approaches infinity.

Clan$_i$, Religion$_i$). Since the relational attribute "Distance$_{ij}$" does not imply a unique individual geographic location, we treat the entire vector of distances between i and the rest of the households as i's individual attribute.[26] The categorical variables (family, clan, religion) and continuous variables (distance, wealth) are combined as described in Appendix A (in particular, Equation (32)), with $\lambda = 0.1$ and h set according to the "normal reference rule-of-thumb". Figure 5 presents a histogram of the resulting estimated beliefs.

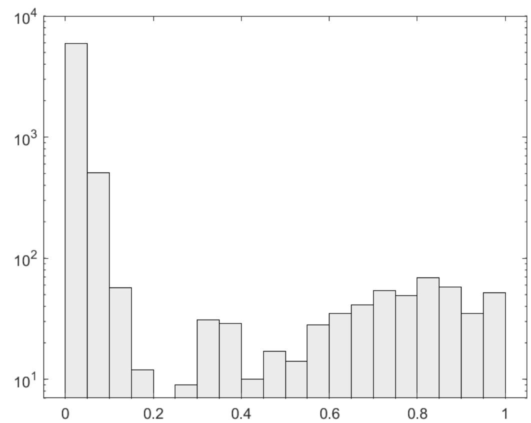


Figure 5: Histogram of the estimated beliefs in the Nyakatoke network

Note: the y-axis is on a logarithmic scale.

The results of the second stage are reported in Table 4. Columns 1-4 present different specifications of the endogenous regressors. Column 5 presents the marginal effects that correspond to the most complete specification of column 4. Standard errors are computed according the expression given in Proposition 4 with the true parameters replaced by their estimates.

As for the endogenous regressors, the coefficient of the number of j's friends can be positive or negative depending on whether households prefer potential partners to have many or few other partners. In principle, both types of externalities are conceivable in the context of risk-sharing arrangements: if j has many friends she may have a rather limited amount of resources to devote to i, implying a negative coefficient. If j has many friends she is likely to be well-positioned to provide i with financial support in case of need, and is also less likely to rely heavily on i in case she herself is in need, implying a positive coefficient.

[26] Consider a three-agent network in which agents 1 and 2 have the same geographic distances from (2,3) and (1,3), respectively. These distance profiles can be obtained by assuming various individual locations for agents 1 and 2, e.g. all location configurations in which all agents are located on a line and agents 1 and 2 are located symmetrically around agent 3.

The sum of wealth of j's friends is expected to be positive, as this grants j access to greater wealth which may indirectly benefit i.

The significance of the endogenous regressors' coefficients in Table 4 provides evidence for the existence of network externalities. Concentrating on the full specification in column 4, the positive sign of the coefficient of the number of j's friends suggests that the benefits from having a partner with many other partners (greater financial resilience) outweigh the costs (dilution of attention and/or resources). For the average pair i and j, an increment of one unit in the expected number of j's friends ($\approx 12\%$ of the average expected number of j's friends) is associated with an increase of roughly 0.016 in the probability of a proposal ($\approx 9\%$ of the average predicted proposal probability).

The signs of the other coefficients conform to our expectations. The constant appears negative, reflecting the idea that maintaining links is costly. The coefficient of the geographical distance between households is also negative, as distance is likely to render links harder to maintain. The coefficient of wealth is positive, as the wealthier a potential partner is the more helpful she could be in case of a negative income shock. The coefficients of the homophily regressors are all positive, in line with the large evidence that similarity between agents renders them more desirable to each other.

6 Concluding remarks

While network theory has long advocated the importance of network architecture in predicting link formation, the empirical contributions in this direction are still rare. Models of link formation with network externalities are at the frontier of the econometric research, facing difficulties related to dimensionality and multiplicity of equilibria (Graham, 2015; Chandrasekhar, 2016; De Paula, 2017). Data on network interactions were previously scarce but are now becoming more available. The current enthusiasm for network data from digital interaction platforms (Vosoughi et al., 2018; Blumenstock, 2018) has refueled the research interest about how non-digital links are formed, and how they respond to strategic incentives.

In this paper we propose a method to estimate network externalities in a simultaneous game of bilateral link formation with incomplete information. We know from theory that unilateral and bilateral link formation rules result in fundamentally different network structures, which in turn has profound implications on the aggregate outcome that can be achieved (Bala and Goyal, 2000; Charness and Jackson, 2007). We provide existence, consistency and asymptotic normality results for the proposed estimator, and we test its asymptotic perfor-

	G				ME
	(1)	(2)	(3)	(4)	(5)
Same family	0.8436***	0.8496***	0.8556***	0.8493***	0.2934***
	(0.0627)	(0.0643)	(0.0642)	(0.0644)	(0.0256)
Same clan	0.1661***	0.1483**	0.1487**	0.1485**	0.0415**
	(0.0579)	(0.0601)	(0.0605)	(0.0602)	(0.0177)
Same religion	0.1649***	0.1752***	0.1735***	0.1751***	0.0495***
	(0.0401)	(0.041)	(0.0411)	(0.041)	(0.0118)
Distance_{ij}	-0.0009***	-0.0009***	-0.0009***	-0.0009***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0)
Wealth_j	0.0586***	0.0358***	-0.021	0.0376**	0.0098**
	(0.0069)	(0.008)	(0.0192)	(0.0155)	(0.004)
Number of j's friends		0.0598***		0.0607***	0.0159***
		(0.0086)		(0.0113)	(0.003)
Wealth of j's friends			0.0075***	-0.0002	0
			(0.0017)	(0.0013)	(0.0003)
Constant	-0.6482***	-1.0888***	-0.6243***	-1.0967***	
	(0.0563)	(0.0863)	(0.0579)	(0.1063)	
# observations	7021	7021	7021	7021	

Table 4: Estimated coefficients.

Notes: Column 5 reports the marginal effects for the specification of column 4. Standard errors in parentheses. * Significance at 10% level. ** Significance at 5% level. *** Significance at 1% level.

mance through a simulation exercise. Our procedure allows to make inference about some aspects of agents' preferences over network topology when data on a single (and possibly large) network are available. It provides a simpler alternative to methods exploiting pairwise stability under complete information (de Paula et al., 2018; Sheng, 2016).

We illustrate the method using data on risk-sharing in a Tanzanian village. Risk-sharing links are commonly assumed to be mutual and provide an intriguing case for the importance of externalities from indirect connections. Results confirm that the network architecture has an explanatory value: households seem to take into consideration the number of indirect friends they stand to gain when making linking decisions. Our estimates suggest that an additional two-steps-away connection is associated with an average increase of roughly 9% in the predicted proposal probability.

References

Ambrus, A. and Elliott, M. (2018). Consumption risk-sharing in social networks. *mimeo Duke University and University of Cambridge.*

Ambrus, A., Mobius, M., and Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review*, 104(1):149–82.

Badev, A. (2018). Nash equilibria on (un)stable networks. *mimeo.*

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. econometrica 68, 1181q1230. the role of social structure in facilitating delinquent behavior. *International Economic Review*, 45.

Bjorn, P. A. and Vuong, Q. H. (1984). Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation.

Bloch, F., Genicot, G., and Ray, D. (2008). Informal insurance in social networks. *Journal of Economic Theory*, 143(1):36–58.

Blumenstock, J. (2018). Dont forget people in the use of big data for development.

Boucher, V. and Mourifie, I. (2017). My friend far, far away: a random field approach to exponential random graph models. *The Econometrics Journal*, 20(3):S14–S46.

Bramoullé, Y. and Kranton, R. (2007). Risk-sharing networks. *Journal of Economic Behavior & Organization*, 64(3-4):275–294.

Breschi, S. and Lissoni, F. (2005). " cross-firm" inventors and social networks: Localized knowledge spillovers revisited. *Annales d'Economie et de Statistique*, pages 189–209.

Bresnahan, T. F. and Reiss, P. C. (1991). Empirical models of discrete games. *Journal of Econometrics*, 48(1-2):57–81.

Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford Handbook of the Economics of Networks*, pages 303–357.

Chandrasekhar, A. G. and Jackson, M. O. (2016). A network formation model based on subgraphs. *Available at SSRN 2660381.*

Charness, G. and Jackson, M. O. (2007). Group play in games and the role of consent in network formation. *Journal of Economic Theory*, 136(1):417–445.

Coate, S. and Ravallion, M. (1993). Reciprocity without commitment: Characterization and performance of informal insurance arrangements. *Journal of development Economics*, 40(1):1–24.

Comola, M. and Fafchamps, M. (2014). Testing Unilateral and Bilateral Link Formation. *Economic Journal*, 124(579):954–976.

Comola, M. and Fafchamps, M. (2017). The missing transfers: Estimating misreporting in dyadic data. *Economic Development and Cultural Change*, 65(3):549–582.

De Paula, A. (2017). Econometrics of network models. In *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, pages 268–323. Cambridge University Press Cambridge.

de Paula, A., Richards-Shubik, S., and Tamer, E. T. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86:263 – 288.

De Weerdt, J. and Dercon, S. (2006). Risk-sharing networks and insurance against illness. *Journal of development Economics*, 81(2):337–356.

De Weerdt, J. and Fafchamps, M. (2011). Social identity and the formation of health insurance networks. *Journal of Development Studies*, 47(8):1152–1177.

Fafchamps, M. and Lund, S. (2003). Risk-sharing networks in rural philippines. *Journal of development Economics*, 71(2):261–287.

Genicot, G. and Ray, D. (2003). Group formation in risk-sharing arrangements. *The Review of Economic Studies*, 70(1):87–113.

Gilleskie, D. and Zhang, Y. (2009). Friendship formation and smoking initiation among teens. unpublished.

Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., and Yariv, L. (2010). The 1/d law of giving. *American Economic Journal: Microeconomics*, 2(1):183–203.

Goyal, S., Van der Leij, M., and MoragaGonzlez, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–412.

Graham, B. S. (2015). Methods of identification in social networks. *Annu. Rev. Econ.*, 7(1):465–485.

Hoshino, T. (2019). Two-step estimation of incomplete information social interaction models with sample selection. *Journal of Business Economic Statistics*, 37(4):598–612.

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

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1857–97.

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

Kimball, M. S. (1988). Farmers' cooperatives as behavior toward risk. *The American Economic Review*, 78(1):224–232.

Kocherlakota, N. R. (1996). Implications of efficient risk sharing without commitment. *The Review of Economic Studies*, 63(4):595–609.

König, M., Liu, X., and Zenou, Y. (2014). R&d networks: Theory, empirics and policy implications.

König, M. D. (2016). The formation of networks with local spillovers and limited observability. *Theoretical Economics*, 11(3):813–863.

Krishnan, P. and Sciubba, E. (2009). Links and architecture in village networks. *The Economic Journal*, 119(537):917–949.

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

Lewbel, A. and Tang, X. (2015). Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics*, 189(1):229–244.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.

Menzel, K. (2016). Inference for games with many players. *Review of Economic Studies*, 83(1):306–337.

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

Myerson, R. (1991). Game theory: Analysis of conflict harvard univ. *Press, Cambridge*.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Poirier, D. J. (1980). Partial Observability in Bivariate Probit Models. 12(7802).

Rabinovich, Z., Naroditskiy, V., Gerding, E. H., and Jennings, N. R. (2013). Computing pure bayesian-nash equilibria in games with finite actions and continuous types. *Artificial Intelligence*, 195:106–139.

Sheng, S. (2016). A structural econometric analysis of network formation games. *mimeo*.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, 51(5):756–770.

Song, Y. and van der Schaar, M. (2015). Dynamic network formation with incomplete information. *Economic Theory*, 59(2):301–331.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165.

Townsend, R. M. (1994). Risk and Insurance in Village India. *Econometrica*, 62(3):539–591.

Udry, C. (1994). Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *The Review of Economic Studies*, 61(3):495–526.

Vandenbossche, J. and Demuynck, T. (2013). Network formation with heterogeneous agents and absolute friction. *Computational Economics*, 42(1):23–45.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Appendices

A Extensions

A.1 Continuous Attributes

The restriction that the selected equilibrium is symmetric requires that identical pairs of agents have identical ex-ante linking probabilities. In the case of continuous attributes no two pairs of agents are identical. The symmetric equilibrium condition is therefore non-restrictive - it is trivially fulfilled for any equilibrium. When attributes are continuous we substitute the symmetry condition with a continuity condition, requiring that *similar* pairs of agents have *similar* ex-ante linking probabilities. Formally, an equilibrium σ^G is continuous if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $ij \neq kl \in N$:

$$\begin{gathered} (\|X_i - X_k\| < \delta \textit{ and } \|X_j - X_l\| < \delta) \textit{ or } (\|X_i - X_l\| < \delta \textit{ and } \|X_j - X_k\| < \delta) \\ \Downarrow \\ |\sigma_{ij} - \sigma_{kl}| < \varepsilon \end{gathered} \tag{27}$$

The following proposition establishes the existence of a continuous equilibrium.

Proposition 5 (Existence)**.** *For any continuous X and for any θ_0, there exists a continuous equilibrium.*

Under the assumption that the selected equilibrium is continuous, we can estimate σ^G_{ij} using Kernel methods. Letting $d(X_i, X_j, X_k, X_l)$ denote the vector of distances in attributes between the two unordered pairs,[27] K denote a standard product kernel function, and h denote the bandwidth selection, the estimator is:

$$\hat{\sigma}^G_{ij} \equiv \frac{\sum_{l,k>l} G_{kl} \cdot K\left(\frac{d(X_i, X_j, X_k, X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i, X_j, X_k, X_l)}{h}\right)} \tag{28}$$

Proposition 6. *When X is continuous and the selected equilibrium is continuous, $\hat{\sigma}^G_{ij}$ is consistent for σ^G_{ij} for all $i, j \in N$ such that $i \neq j$.*

[27] $d(X_i, X_j, X_k, X_l) = [X_i - X_k, X_j - X_l]$ if $\|[X_i - X_k, X_j - X_l]\| \leq \|[X_i - X_l, X_j - X_k]\|$ and $[X_i - X_l, X_j - X_k]$ otherwise.

While the estimator in equation (20) applies only to discrete attributes, the estimator in equation (28) applies only to continuous attributes. In various applications, however, X may contain a mix of both discrete and continuous attributes. One approach to deal with these cases is to weigh the discrete variables of an observation according to equation (20), the continuous variables according to equation (28), and define the weight of the observation as the product of the two. Formally, letting X_i^d be i's discrete attributes, X_i^c her continuous ones and $X_i = [X_i^d, X_i^c]$, this approach yields the following estimator:

$$\hat{\sigma}_{ij}^G \equiv \frac{\sum_{l,k>l} G_{kl} \mathbb{1}\left\{(X_i^d = X_k^d \wedge X_j^d = X_l^d) \vee (X_i^d = X_l^d \wedge X_j^d = X_k^d)\right\} K\left(\frac{d(X_i^c, X_j^c, X_k^c, X_l^c)}{h}\right)}{\sum_{l,k>l} \mathbb{1}\left\{(X_i^d = X_k^d \wedge X_j^d = X_l^d) \vee (X_i^d = X_l^d \wedge X_j^d = X_k^d)\right\} K\left(\frac{d(X_i^c, X_j^c, X_k^c, X_l^c)}{h}\right)} \quad (29)$$

A.2 Smoothing

A practical concern that may arise with respect to both the "mixed attributes" estimator and the "only discrete" estimator is that in a finite sample the number of observations with identical discrete attributes may be too small to allow for a meaningful estimation. This happens in particular when the sample size is small, the number of discrete variables is high and their support is large. Li and Racine (2007) suggest overcoming this problem by smoothing the discrete variables. Let $X_{i,s}^d$ be the sth component of the X_i^d vector and define

$$t(X_{i,s}^d, X_{j,s}^d, X_{k,s}^d, X_{l,s}^d, \lambda) \equiv \begin{cases} 1 & \text{if } (X_{i,s}^d = X_{k,s}^d \wedge X_{j,s}^d = X_{l,s}^d) \vee (X_{i,s}^d = X_{l,s}^d \wedge X_{j,s}^d = X_{k,s}^d) \\ \lambda & \text{otherwise} \end{cases} \quad (30)$$

and

$$T(X_i^d, X_j^d, X_k^d, X_l^d, \lambda) \equiv \prod_s \lambda^{t(X_{i,s}^d, X_{j,s}^d, X_{k,s}^d, X_{l,s}^d, \lambda)} \quad (31)$$

Using $T(\cdot)$ as the product kernel function for the discrete variables, the mixed attributes estimator (29) becomes:

$$\hat{\sigma}_{ij}^G \equiv \frac{\sum_{l,k>l} G_{kl} \cdot T(X_i^d, X_j^d, X_k^d, X_l^d, \lambda) \cdot K\left(\frac{d(X_i^c, X_j^c, X_k^c, X_l^c)}{h}\right)}{\sum_{l,k>l} T(X_i^d, X_j^d, X_k^d, X_l^d, \lambda) \cdot K\left(\frac{d(X_i^c, X_j^c, X_k^c, X_l^c)}{h}\right)} \quad (32)$$

Note that when $\lambda = 0$, $T(X_i^d, X_j^d, X_k^d, X_l^d, 0)$ takes the value 1 if ij and kl are identical in their discrete attributes and 0 otherwise. (32) therefore reduces to (29) and no smoothing occurs. On the other extreme, when $\lambda = 1$, $T(X_i^d, X_j^d, X_k^d, X_l^d, 1) = 1$ for all ij and kl. The discrete attributes are therefore completely smoothed out. $\lambda \in (0, 1)$ corresponds to different levels of smoothing of the discrete variables.

B Proofs

B.1 Proposition 1

Proof. Denote by Σ the set of all σ^G matrices such that:

1. $\forall i, j \in N,\ \sigma_{ij}^G \in [0, 1]$
2. $\forall i \in N,\ \sigma_{ii}^G = 0$
3. $\forall i, j \in N,\ \sigma_{ij}^G = \sigma_{ji}^G$
4. $\forall i, j, k, l \in N,\ (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k) \implies \sigma_{ij}^G = \sigma_{kl}^G$

Denote by $\Gamma(\cdot)$ the function that maps belief matrices to linking probabilities:

$$\Gamma_{ij}(\sigma^G) \equiv \begin{cases} \Phi(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0) \mid \sigma^{G_{-i}}]) \cdot \Phi(\mathbb{E}[v_{ji}(X, G_{-j}; \theta_0) \mid \sigma^{G_{-j}}]) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{33}$$

By 9, an equilibrium is a fixed point of $\Gamma(\cdot)$, i.e. a σ^G such that for all $i, j \in N$:

$$\Gamma_{ij}(\sigma^G) = \sigma_{ij}^G \tag{34}$$

To prove that such σ^G exists we verify the conditions of Brouwer's fixed point theorem.

$\Gamma(\cdot)$ **maps from Σ to Σ.** First, since Γ_{ij} is either the product of two probabilities or 0, $\Gamma_{ij} \in [0, 1]$ for all i, j. Second, by definition $\Gamma_{ii} = 0$ for all i. Third, since Γ_{ij} depends symmetrically on the expected utility of i from a link with j and of j from a link with i, $\Gamma_{ij} = \Gamma_{ji}$ for all i, j. Fourth, for any two agents i and k such that $X_i = X_k$, condition 4 above implies that for any third agent $j \neq i, k$ the input matrix must satisfy $\sigma_{ij}^G = \sigma_{kj}^G$. By conditions 2 and 3, $\sigma_{ii}^G = \sigma_{kk}^G$ and $\sigma_{ik}^G = \sigma_{ki}^G$. The ith and kth rows and columns in σ^G therefore contain the same elements, implying that $\sigma^{G_{-i}}$ and $\sigma^{G_{-k}}$ are identical up to a permutation of labels. Anonymity of $v_{ij}(\cdot)$ thus implies that $\Gamma_{ij} = \Gamma_{kj}$ for all i, j, k. Applying the same logic for an agent l such that $X_j = X_l$, we obtain also that $\Gamma_{ij} = \Gamma_{kl}$ for all i, j, k, l.

Γ is continuous in σ^G. The expected utilities are continuous in σ^G, and $\Phi(\cdot)$ is a continuous function. Therefore Γ is continuous in σ^G.

Σ is a convex subset of $[0,1]^{n\times n}$. Since any linear combination of any two matrices in Σ yields a matrix in Σ, it is a convex set.

Σ is a compact subset of $[0,1]^{n\times n}$. The sets of values that each entry in the matrices in Σ can obtain are bounded (by 0 and 1) and closed (for off-diagonal elements because the boundaries 0 and 1 are included and for diagonal elements because they are singletons). The Cartesian product of bounded and closed sets is also bounded and closed, so Σ is bounded and closed. By the Heine-Borel theorem it follows that Σ is compact.

The existence of a symmetric Bayesian equilibrium thus follows from Brouwer's fixed point theorem. □

B.2 Proposition 2

Proof. First, note that the linking statuses of all pairs of agents which are observationally equivalent are independent and have the same expected value (due to symmetry). Thus, for any i and j we can apply a law of large numbers:

$$\hat{\sigma}^G_{ij} \equiv \frac{\sum\limits_{l,k>l}\Big[G_{kl}\cdot \mathbb{1}\big\{(X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k)\big\}\Big]}{\sum\limits_{l,k>l}\Big[\mathbb{1}\big\{(X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k)\big\}\Big]} \tag{35}$$

$$= \frac{\sum\limits_{l,k>l\,:\,(X_i=X_k\wedge X_j=X_l)\vee(X_i=X_l\wedge X_j=X_k)} G_{kl}}{\sum\limits_{l,k>l}\Big[\mathbb{1}\big\{(X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k)\big\}\Big]} \tag{36}$$

$$\xrightarrow{p} \mathbb{E}[G_{kl} \mid (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k), X, \sigma^G] \tag{37}$$

In addition:

$$\begin{aligned}
\mathbb{E}[G_{kl} \mid (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k), X, \sigma^G] &= \mathbb{E}(G_{ij} \mid X, \sigma^G) && (38)\\
&= P(G_{ij} = 1 \mid X, \sigma^G) && (39)\\
&= \sigma^G_{ij} && (40)
\end{aligned}$$

Line 38 is true because the probabilities of observationally equivalent pairs to be linked are equal (due to symmetry), and line 40 is true because in equilibrium beliefs are correct. □

B.3 Proposition 3

Proof. To prove that $\hat{\theta}$ is consistent for θ we verify the conditions of Theorem 2.1 in Newey and McFadden (1994).

$\mathbb{E}[l(\theta, \sigma^G)]$ **is uniquely maximized at** θ_0**.** $\mathbb{E}[l(\theta, \sigma^G)]$ is uniquely maximized at θ_0 if for all $\theta \neq \theta_0$, $\mathbb{E}[l(\theta, \sigma^G)] - \mathbb{E}[l(\theta_0, \sigma^G)] < 0$. We now show that this is true.

$$\begin{aligned}
\mathbb{E}[l(\theta, \sigma^G)] - \mathbb{E}[l(\theta_0, \sigma^G)] &= \mathbb{E}\left[\frac{\log(L(\theta, \sigma^G))}{\frac{1}{2}n(n-1)} - \frac{\log(L(\theta_0, \sigma^G))}{\frac{1}{2}n(n-1)}\right] && (41)\\
&= \frac{1}{\frac{1}{2}n(n-1)}\mathbb{E}\left[\log(L(\theta, \sigma^G) - \log(L(\theta_0, \sigma^G))\right] && (42)\\
&= \frac{1}{\frac{1}{2}n(n-1)}\mathbb{E}\left[\log\left(\frac{L(\theta, \sigma^G)}{L(\theta_0, \sigma^G)}\right)\right] && (43)\\
&= \frac{1}{\frac{1}{2}n(n-1)}\sum_{i,j>i}^{n}\Bigg[\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\big)\Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\big)\times\\
&\quad \log\left(\frac{\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta\big)\Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta\big)}{\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\big)\Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\big)}\right)+ && (44)\\
&\quad \big(1-\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\big)\Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\big)\big)\times\\
&\quad \log\left(\frac{1-\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta\big)\Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta\big)}{1-\Phi\big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\big)\Phi\big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\big)}\right)\Bigg]\\
&\leq \frac{1}{\frac{1}{2}n(n-1)}\sum_{i,j>i}^{n}\Big[\log(1)\Big] && (45)\\
&= 0 && (46)
\end{aligned}$$

where line 45 is obtained by applying Jensen's inequality.
This establishes that θ_0 maximizes $\mathbb{E}[l(\theta, \sigma^G)]$. It remains to show that it is its *unique* maximizer. Consider line 44. Since $\Phi(\cdot)$ is strictly positive, the only way 44 would equal 0 is if the fractions inside the logs evaluate to 1, but this only happens when $\theta = \theta_0$. Thus, θ_0 is the unique maximizer of $\mathbb{E}[l(\theta, \sigma^G)]$.

Θ is compact. True by assumption.

$\mathbb{E}[l(\theta, \sigma^G)]$ is continuous in θ and $l(\theta, \hat{\sigma}^G)$ converges uniformly in probability to $\mathbb{E}[l(\theta, \sigma^G)]$. We show that this is true by verifying the conditions of Lemma 2.4 in Newey and McFadden (1994). First, Θ is compact, by assumption. Second, $l(\theta, \sigma^G)$ is continuous in θ because $\Phi(\cdot)$ and $\log(\cdot)$ are continuous. Third, we need to show that there exists a function $d(G, \delta, \hat{\gamma})$ such that $|l(\theta, \hat{\sigma}^G)| \leq d(G, \delta, \hat{\gamma})$ and $\mathbb{E}[d(G, \delta, \hat{\gamma})] < \infty$. We start by considering the absolute value of the first part of the log-likelihood function:

$$
\begin{aligned}
& |\log\left(\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\theta)\Phi([\delta_{ji}, \hat{\gamma}_{ji}]\theta)\right)| && (47)\\
&= |\log(\Phi(0)\Phi(0)) + \frac{\phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}([\delta_{ij}, \hat{\gamma}_{ij}]\theta - 0) + \frac{\phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}([\delta_{ij}, \hat{\gamma}_{ij}]\theta - 0)| && (48)\\
&\leq 2 + \frac{\phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}|[\delta_{ij}, \hat{\gamma}_{ij}]\theta| + \frac{\phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}|[\delta_{ij}, \hat{\gamma}_{ij}]\theta| && (49)\\
&\leq 2 + (1 + |[\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta}|)|[\delta_{ij}, \hat{\gamma}_{ij}]\theta| + (1 + |[\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta}|)|[\delta_{ij}, \hat{\gamma}_{ij}]\theta| && (50)\\
&\leq 2 + (1 + [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \tilde{\theta}) \cdot [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \theta + (1 + [\delta_{ji}, \hat{\gamma}_{ji}] \cdot \tilde{\theta}) \cdot [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \theta && (51)
\end{aligned}
$$

Where line 48 is true by the mean value theorem (recall that the derivative of $\log(\Phi(v)\Phi(u))$ w.r.t v is $\frac{\phi(v)}{\Phi(v)}$ and w.r.t u is $\frac{\phi(u)}{\Phi(u)}$), line 49 is true by the triangular inequality, line 50 is true because $\frac{\phi(v)}{\Phi(v)} \leq 1 + |v|$ for all v, and line 51 is true by the Cauchy-Schwartz inequality.
Consider now the absolute value of the second part of the log-likelihood function:

$$
\begin{aligned}
& |\log\left(1 - \Phi([\delta_{ij}, \hat{\gamma}_{ij}]\theta)\Phi([\delta_{ji}, \hat{\gamma}_{ji}]\theta)\right)| && (52)\\
&= |\log(1 - \Phi(0)\Phi(0)) + \frac{-\phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})}{1 - \Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})}([\delta_{ij}, \hat{\gamma}_{ij}]\theta - 0) \\
&\quad + \frac{-\phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}{1 - \Phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}([\delta_{ji}, \hat{\gamma}_{ji}]\theta - 0)| && (53)
\end{aligned}
$$

$$
\begin{aligned}
&\leq 2 + \frac{\phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})}{1 - \Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})} \cdot |[\delta_{ij}, \hat{\gamma}_{ij}]\theta| \\
&\quad + \frac{\phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})}{1 - \Phi([\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta})} \cdot |[\delta_{ji}, \hat{\gamma}_{ji}]\theta| \qquad (54) \\
&\leq 2 + (1 + |[\delta_{ij}, \hat{\gamma}_{ij}]\tilde{\theta}|) \cdot |[\delta_{ij}, \hat{\gamma}_{ij}]\theta| + (1 + |[\delta_{ji}, \hat{\gamma}_{ji}]\tilde{\theta}|) \cdot |[\delta_{ji}, \hat{\gamma}_{ji}]\theta| \qquad (55) \\
&\leq 2 + (1 + [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \tilde{\theta}) \cdot [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \theta + (1 + [\delta_{ji}, \hat{\gamma}_{ji}] \cdot \tilde{\theta}) \cdot [\delta_{ji}, \hat{\gamma}_{ji}] \cdot \theta \qquad (56)
\end{aligned}
$$

Where line 55 is true because $\frac{\phi(v)\Phi(u)}{1-\Phi(v)\Phi(u)} \leq \frac{\phi(v)}{1-\Phi(v)} = \frac{\phi(v)}{\Phi(-v)} \leq 1 + |v|$.
Letting $\theta_m = \sup_{\theta \in \Theta} \theta$, 51 and 56 imply that $|l(\theta, \hat{\sigma}^G)|$ is bounded from above by:

$$
2 + (1 + [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \theta_m) \cdot [\delta_{ij}, \hat{\gamma}_{ij}] \cdot \theta_m + (1 + [\delta_{ji}, \hat{\gamma}_{ji}] \cdot \theta_m) \cdot [\delta_{ji}, \hat{\gamma}_{ji}] \cdot \theta_m \qquad (57)
$$

And a sufficient condition for the expected value of this function to be finite is that $\mathbb{E}\left[[\delta_{ij}, \hat{\gamma}_{ij}][\delta_{ij}, \hat{\gamma}_{ij}]'\right]$ and $\mathbb{E}\left[[\delta_{ji}, \hat{\gamma}_{ji}][\delta_{ji}, \hat{\gamma}_{ji}]'\right]$ exist and are finite. □

B.4 Proposition 4

Proof. Denote the score of the log-likelihood by S and its ijth summand by S_{ij}:

$$
\begin{aligned}
S(\gamma, \theta) &\equiv \nabla_\theta l(\theta) && (58) \\
&= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} S_{ij}(\gamma_{ij}, \gamma_{ji}, \theta) && (59) \\
&= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \frac{q_{ij}(G_{ij} - m_{ij})}{m_{ij}(1 - m_{ij})} && (60)
\end{aligned}
$$

Where:

$$
\begin{aligned}
m_{ij} &\equiv \Phi([\delta_{ij}, \gamma_{ij}]\theta) \cdot \Phi([\delta_{ji}, \gamma_{ji}]\theta) && (61) \\
q_{ij} &\equiv \phi([\delta_{ij}, \gamma_{ij}]\theta) \cdot [\delta_{ij}, \gamma_{ij}] \cdot \Phi([\delta_{ji}, \gamma_{ji}]\theta) + \phi([\delta_{ji}, \gamma_{ji}]\theta) \cdot [\delta_{ji}, \gamma_{ji}] \cdot \Phi([\delta_{ij}, \gamma_{ij}]\theta) && (62)
\end{aligned}
$$

Let $\hat{\gamma}_{ij}$ denote the output of $\gamma_{ij}(X, \hat{\sigma}^G)$ and $\hat{\gamma}$ denote the set of $\hat{\gamma}_{ij}$ for all i, j. By first order condition:

$$
S(\hat{\gamma}, \hat{\theta}) = 0 \qquad (63)
$$

By the mean value theorem, there exists a θ^* between $\hat{\theta}$ and θ_0 such that:

$$S(\hat{\gamma}, \hat{\theta}) = S(\hat{\gamma}, \theta_0) + \nabla_\theta S(\hat{\gamma}, \theta^*)(\hat{\theta} - \theta_0) \tag{64}$$

Combining 63 and 64, and solving for $(\hat{\theta} - \theta_0)$ gives:

$$\hat{\theta} - \theta_0 = -\left(\nabla_\theta S(\hat{\gamma}, \theta^*)\right)^{-1} S(\hat{\gamma}, \theta_0) \tag{65}$$

Since $\hat{\gamma}$ and $\hat{\theta}$ are consistent, and given that θ^* is "trapped" between $\hat{\theta}$ and θ_0 (which makes it also consistent):

$$\nabla_\theta S(\hat{\gamma}, \theta^*) - \mathbb{E}[\nabla_\theta S(\gamma_0, \theta_0) \mid X, \sigma^G] \xrightarrow{p} 0 \tag{66}$$

Denote the expected value of the hessian, by V and its ijth summand by V_{ij}:

$$V(\gamma, \theta) \equiv \mathbb{E}[\nabla_\theta S(\gamma, \theta) \mid X, \sigma^G] \tag{67}$$

$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} V_{ij}(\gamma_{ij}, \gamma_{ji}, \theta) \tag{68}$$

$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \frac{-q_{ij}q'_{ij}}{m_{ij}(1-m_{ij})} \tag{69}$$

We can thus rewrite 65 as:

$$\hat{\theta} - \theta_0 = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} S(\hat{\gamma}, \theta_0) \tag{70}$$

By adding and subtracting $\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]$ we obtain:

$$\hat{\theta} - \theta_0 = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} \left(\underbrace{S(\hat{\gamma}, \theta_0) - \mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]}_{a} + \underbrace{\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]}_{b} \right) \tag{71}$$

By a second order Taylor expansion of $S(\hat{\gamma}, \theta_0)$:

$$\begin{aligned} S(\hat{\gamma}, \theta_0) = S(\gamma_0, \theta_0) + \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} [\nabla_{\gamma_{ij}} S(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ij} - \gamma^0_{ij}) \\ + \nabla_{\gamma_{ji}} S(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ji} - \gamma^0_{ji})] + o_p(1) \end{aligned} \tag{72}$$

By a second-order Taylor expansion of $\mathbb{E}[S(\hat{\gamma},\theta_0) \mid X, \sigma^G]$:

$$\mathbb{E}[S(\hat{\gamma},\theta_0) \mid X,\sigma^G] = \underbrace{\mathbb{E}[S(\gamma_0,\theta_0) \mid X,\sigma^G]}_{0} + \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \mathbb{E}\big[\nabla_{\gamma_{ij}} S(\gamma^0_{ij},\gamma^0_{ji},\theta_0) \cdot (\hat{\gamma}_{ij}-\gamma^0_{ij}) + \nabla_{\gamma_{ji}} S(\gamma^0_{ij},\gamma^0_{ji},\theta_0)\cdot(\hat{\gamma}_{ji}-\gamma^0_{ji}) \mid X,\sigma^G\big] + o_p(1) \tag{73}$$

Since, by the law of large numbers, the middle part of 72 converges to the middle part of 73:

$$a = S(\gamma_0,\theta_0) + o_p(1) \tag{74}$$

Denote the expected value of $\nabla_{\gamma_{ij}} S(\gamma,\theta)$ by M and its ijth summand by M_{ij}:

$$M(\gamma,\theta) \equiv \mathbb{E}[\nabla_{\gamma_{ij}} S(\gamma,\theta) \mid X,\sigma^G] \tag{75}$$
$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} M_{ij}(\gamma_{ij},\gamma_{ji},\theta) \tag{76}$$
$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \frac{-q_{ij}p'_{ij}}{m_{ij}(1-m_{ij})} \tag{77}$$

Where:

$$p_{ij} \equiv \phi([\delta_{ij},\gamma_{ij}]\theta)\cdot\theta^\gamma\cdot\Phi([\delta_{ji},\gamma_{ji}]\theta) \tag{78}$$

and θ^γ denotes the elements in θ which correspond to the endogenous regressors γ.

Using this notation we can rewrite 73 as:

$$\mathbb{E}[S(\hat{\gamma},\theta_0) \mid X,\sigma^G] = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \big[M_{ij}(\gamma^0_{ij},\gamma^0_{ji},\theta_0)\cdot(\hat{\gamma}_{ij}-\gamma^0_{ij}) + M_{ji}(\gamma^0_{ij},\gamma^0_{ji},\theta_0)\cdot(\hat{\gamma}_{ji}-\gamma^0_{ji})\big] + o_p(1) \tag{79}$$
$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j\neq i} \big[M_{ij}(\gamma^0_{ij},\gamma^0_{ji},\theta_0)\cdot(\hat{\gamma}_{ij}-\gamma^0_{ij})\big] + o_p(1) \tag{80}$$

Since we assume that $\sum_{i,j\neq i}\gamma_{ij}(X,G_{-i}) = \sum_{i,j\neq i}\gamma_{ij}(X,\hat{\sigma}^{G_{-i}})$, we can replace $\gamma_{ij}(\hat{\sigma}^G)$ by

$\gamma_{ij}(G)$, which we denote here by α_{ij}:

$$\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G] = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \big[M_{ij}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \cdot (\alpha_{ij} - \gamma_{ij}^0) \\ + M_{ji}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \cdot (\alpha_{ji} - \gamma_{ji}^0)\big] + o_p(1) \tag{81}$$

We now rewrite 71 using our replacements for a and b:

$$\hat{\theta} - \theta_0 = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \big[S_{ij}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \\ + M_{ij}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \cdot (\alpha_{ij} - \gamma_{ij}^0) + M_{ji}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \cdot (\alpha_{ji} - \gamma_{ji}^0) + o_p(1)\big] \tag{82}$$

Denote the ijth summand in this equation by W_{ij}:

$$W_{ij} \equiv S_{ij}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) + M_{ij}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \cdot (\alpha_{ij} - \gamma_{ij}^0) + M_{ji}(\gamma_{ij}^0, \gamma_{ji}^0, \theta_0) \cdot (\alpha_{ji} - \gamma_{ji}^0) \tag{83}$$

Using this notation and multiplying through by $\sqrt{\frac{1}{2}n(n-1)}$:

$$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0) = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} \\ \cdot \sqrt{\frac{1}{2}n(n-1)} \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \big[W_{ij} + o_p(1)\big] \tag{84}$$

Since the summands under the summation sign are conditionally independent (because conditional on X and σ^G, the variation in G_{ij} comes only from ϵ_{ij} and ϵ_{ji}, which are all assumed to be i.i.d.), we can now apply a central limit theorem:

$$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0) \sim N(0, V^{-1}\Psi V^{-1}) \tag{85}$$

Where:

$$\Psi \equiv \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} W_{ij} W_{ij}' \tag{86}$$

□

B.5 Lemma 1

Proof. Plugging in the definition of $\gamma_{ij}(\cdot)$ into condition 21, we obtain:

$$\sum_i \sum_{j \neq i} \frac{1}{n-1} \sum_{k \neq i,j} G_{jk} \cdot \mu(X_k) = \sum_i \sum_{j \neq i} \frac{1}{n-1} \sum_{k \neq i,j} \hat{\sigma}^G_{jk} \cdot \mu(X_k) \tag{87}$$

Below, we show that this statement is true if and only if $\sum_{i,j>i} G_{ij} \cdot \mu(X_j) = \sum_{i,j>i} \hat{\sigma}^G_{ij} \cdot \mu(X_j)$. We then proceed to show that the latter is indeed true. From 87:

$$\sum_i \sum_{j \neq i} \sum_{k \neq j} G_{jk}\mu(X_k) - \sum_i \sum_{j \neq i} G_{ji}\mu(X_i) = \sum_i \sum_{j \neq i} \sum_{k \neq j} \hat{\sigma}^G_{jk}\mu(X_k) - \sum_i \sum_{j \neq i} \hat{\sigma}^G_{ji}\mu(X_j) \tag{88}$$

$$(n-1) \cdot \sum_{i;j \neq i} G_{ij}\mu(X_j) - \sum_{i;j \neq i} G_{ij}\mu(X_j) = (n-1) \cdot \sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) - \sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) \tag{89}$$

$$(n-2) \cdot \sum_{i;j \neq i} G_{ij}\mu(X_j) = (n-2) \cdot \sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) \tag{90}$$

$$\sum_{i;j \neq i} G_{ij}\mu(X_j) = \sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) \tag{91}$$

We now show that this is true:

$$\sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) = \sum_{X_A, X_B \in X} \sum_{i;j>i} \mathbb{1}\big\{(X_i = X_A \wedge X_j = X_B) \vee (X_i = X_B \wedge X_j = X_A)\big\} \times \hat{\sigma}^G_{ij} \cdot \mu(X_j) \tag{92}$$

$$= \sum_{X_A, X_B \in X} \sum_{i;j \neq i} \mathbb{1}\big\{(X_i = X_A \wedge X_j = X_B) \vee (X_i = X_B \wedge X_j = X_A)\big\} \times \frac{\sum_{k,l \neq k} \Big[G_{kl} \cdot \mathbb{1}\big\{(X_k = X_A \wedge X_l = X_B) \vee (X_k = X_B \wedge X_l = X_A)\big\}\Big]}{\sum_{k,l \neq k} \mathbb{1}\big\{(X_k = X_A \wedge X_l = X_B) \vee (X_k = X_B \wedge X_l = X_A)\big\}} \times \mu(X_l) \tag{93}$$

$$= \sum_{X_A, X_B \in X} \sum_{k;l \neq k} \mathbb{1}\big\{(X_k = X_A \wedge X_l = X_B) \vee (X_k = X_B \wedge X_l = X_A)\big\} \times G_{kl} \cdot \mu(X_l) \tag{94}$$

$$= \sum_{X_A, X_B \in X} \sum_{i;j \neq i} \mathbb{1}\big\{(X_i = X_A \wedge X_j = X_B) \vee (X_j = X_B \wedge X_i = X_A)\big\} \times G_{ij} \cdot \mu(X_j) \tag{95}$$

$$= \sum_{i;j\neq i} G_{ij}\mu(X_j) \tag{96}$$

Intuitively, this result comes from the fact that when calculating $\hat{\sigma}^G$ we essentially partition the agents into mutually exclusive groups of observationally equivalent pairs, and for each group "redistribute" the total number of links within it among its pairs (uniformly). $\square$

B.6 Proposition 5

Proof. The proof is identical to that of Proposition 1, only that condition 4 has to be replaced by the definition of a continuous equilibrium and the claim that $\Gamma(\cdot)$ **maps from** Σ **to** Σ has to be reestablished.

Denote by Σ the set of all σ^G matrices such that:

1. $\forall i, j \in N,\ \sigma^G_{ij} \in [0,1]$
2. $\forall i \in N,\ \sigma^G_{ii} = 0$
3. $\forall i, j \in N,\ \sigma^G_{ij} = \sigma^G_{ji}$
4. $\forall \varepsilon > 0\ \exists \delta > 0$ such that $\forall i, j \neq k, l \in N$:

$$(\|X_i - X_k\| < \delta \textit{ and } \|X_j - X_l\| < \delta) \textit{ or } (\|X_i - X_l\| < \delta \textit{ and } \|X_j - X_k\| < \delta)$$
$$\Downarrow$$
$$|\sigma^G_{ij} - \sigma^G_{kl}| < \varepsilon$$

We need to show that $\Gamma(\cdot)$ (defined in 33) maps from Σ to Σ.

First, since Γ_{ij} is either the product of two probabilities or 0, $\Gamma_{ij} \in [0,1]$ for all i, j. Second, by definition $\Gamma_{ii} = 0$ for all i. Third, since Γ_{ij} depends symmetrically on the expected utility of i from a link with j and of j from a link with i, $\Gamma_{ij} = \Gamma_{ji}$ for all i, j.

It remains to show that $\Gamma(\cdot)$ maps into matrices that satisfy the 4th condition above, that is, that by choosing a small δ we can make $|\Gamma_{ij}(\sigma^G) - \Gamma_{kj}(\sigma^G)|$ arbitrarily small for all i, k such that $\|X_i - X_k\| < \delta$ and $j \neq i, k$ (by taking another agent $l \neq j$ such that $\|X_j -$

$X_l\| < \delta$ it then follows that we can also make $|\Gamma_{ij}(\sigma^G) - \Gamma_{kl}(\sigma^G)|$ arbitrarily small). Since $|\Gamma_{ij}(\sigma^G) - \Gamma_{kj}(\sigma^G)| = |\Phi(\mathbb{E}[v_{ij}|X, \sigma^G])\Phi(\mathbb{E}[v_{ji}]) - \Phi(\mathbb{E}[v_{kj}])\Phi(\mathbb{E}[v_{jk}])|$ and $\Phi(\cdot)$ is continuous, it is sufficient to show that $|\mathbb{E}[v_{ij}] - \mathbb{E}[v_{kj}]|$ and $|\mathbb{E}[v_{ji}] - \mathbb{E}[v_{jk}]|$ can be made arbitrarily small. For $|\mathbb{E}[v_{ij}] - \mathbb{E}[v_{kj}]|$, by the triangle inequality:

$$\begin{aligned}\left|\mathbb{E}[v_{ij}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-k}}]\right| &= |\mathbb{E}[v_{ij}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}] \\ &\quad + \mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-k}}]| \qquad (97) \\ &\leq |\mathbb{E}[v_{ij}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}]| \\ &\quad + |\mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-k}}]| \qquad (98)\end{aligned}$$

The first part of 98 can be made arbitrarily small by choosing a small δ because the expected value of $v(\cdot)$ is continuous in X. The second part can be made arbitrarily small because by condition 4 the closer X_i and X_k are the closer $\sigma^{G_{-i}}$ and $\sigma^{G_{-k}}$ must be, and the expected value of $v(\cdot)$ is continuous in beliefs. By a similar argument, $|\mathbb{E}[v_{ji}] - \mathbb{E}[v_{jk}]|$ can also be made arbitrarily small: the closer X_i and X_k are the closer the exogenous variables of v_{ji} and v_{jk}, and, by condition 4 so are the ith and kth rows (and columns) of σ^G, and hence so are the endogenous variables of v_{ji} and v_{jk} (in expectancy). Therefore, $\Gamma(\cdot)$ maps from Σ to Σ and the existence of a continuous equilibrium follows from Brouwer's fixed point theorem. □

B.7 Proposition 6

Proof. We show that $\left|\frac{\sum_{l,k>l} G_{kl} \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} - \mathbb{E}[G_{ij}|X, \sigma^G]\right| \xrightarrow{p} 0.$

By adding and subtracting $\frac{\sum_{l,k>l} \mathbb{E}[G_{kl}|X,\sigma^G] \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}$ to the left hand side and applying the triangle inequality we obtain that the left hand side is at most:

$$\left|\underbrace{\frac{\sum_{l,k>l}(G_{kl}-\mathbb{E}[G_{kl}|X,\sigma^G])\cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l}K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}}_{a}\right| + \left|\underbrace{\frac{\sum_{l,k>l}\mathbb{E}[G_{kl}|X,\sigma^G]\cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l}K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}-\mathbb{E}[G_{ij}|X,\sigma^G]}_{b}\right| \tag{99}$$

We deal with a and b separately and show that as h goes to zero and nh^q goes to infinity each of them converges in probability to zero. Starting with a, note that it can be written as the sample average of the random variable $(G_{kl}-\mathbb{E}[G_{kl}|X,\sigma^G])\cdot w_{kl}$, with $w_{kl}=\frac{K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l}K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}\cdot\frac{1}{2}n(n-1)$:

$$a=\frac{1}{\frac{1}{2}n(n-1)}\sum_{l,k>l}(G_{kl}-\mathbb{E}[G_{kl}|X,\sigma^G])\cdot w_{kl} \tag{100}$$

By the law of large numbers this average converges to the expectation of $(G_{kl}-\mathbb{E}[G_{kl}|X,\sigma^G])\cdot w_{kl}$, which, by the law of iterated expectations is zero. $|a|$ therefore converges to zero.

For b, note that conditional on σ^G the expected value of G_{ij} is a function of X. We can thus write $\mathbb{E}[G_{ij}|X,\sigma^G]=\rho(X_i,X_j,X_{-ij})$. Similarly, $\mathbb{E}[G_{kl}|X,\sigma^G]=\rho(X_k,X_l,X_{-kl})$. Because of the undirected nature of the network, $\rho(\cdot)$ is invariant to the order of its first two arguments. In addition, by anonymity, $\rho(\cdot)$ is invariant to permutations of the components of its third argument. The only relevant difference between the inputs in the two cases above is therefore the difference in attributes of the unordered pairs ij and kl. Applying a mean value theorem, we therefore obtain:

$$\mathbb{E}[G_{kl}|X,\sigma^G]=\rho(X_k,X_l,X_{-kl})=\underbrace{\rho(X_i,X_j,X_{-ij})}_{=\mathbb{E}[G_{ij}|X,\sigma^G]}+D\rho(C)\cdot q(X_k,X_l,X_i,X_j) \tag{101}$$

where $D\rho(\cdot)$ denotes the derivative of $\rho(\cdot)$ with respect to its first two arguments and C lies in between (X_k,X_l,X_{-kl}) and (X_i,X_j,X_{-ij}).

By plugging 101 in b:

$$b = \left| \frac{\sum_{l,k>l} D\rho(C) \cdot d(X_k, X_l, X_i, X_j) \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} \right| \tag{102}$$

$$= |D\rho(C)| \cdot \left| \frac{\sum_{l,k>l} d(X_k, X_l, X_i, X_j) \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} \right| \tag{103}$$

The first term is a constant. The second term converges to zero because as h approaches zero, the larger the difference in attributes between ij and kl the smaller the weight ascribed to it. b therefore also converges to zero.

□