Econometric Analysis of Network Formation Models

Cristina Gualdani

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Economics

University College London

I, Cristina Gualdani, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This dissertation addresses topics in the econometrics of network formation models. Chapter 1 provides a review of the literature. Statistical models focus on the specification of the probability distribution of the network. Examples include models in which nodes are born sequentially and meet existing vertices according to random meetings and network-based meetings. Within this group of models, special attention is reserved to the milestone work by Jackson and Rogers (2007): after having discussed and replicated the main results of the paper, an extension of the original model is examined and fitted to a dataset of Google Plus users.

Even if statistical models can reproduce relatively well the main characteristics of real networks, they usually lack of microfundation, essential for counterfactual analysis. The chapter hence moves to considering the econometrics of economic models of network formation, where agents form links in order to maximise a payoff function. Within this framework, Chapter 2 studies identification of the parameters governing agents' preferences in a static game of network formation, where links represent asymmetric relations between players. After having shown existence of an equilibrium, partial identification arguments are provided without restrictions on equilibrium selection. The usual computational difficulties are attenuated by restricting the attention to some local games of the network formation game and giving up on sharpness.

Chapter 3 applies the methodology developed in Chapter 2 to empirically investigate which preferences are behind firms' decisions to appoint competitors' directors as executives. Using data on Italian companies, it is found that a firm i prefers its executives sitting on the board of a rival j when executives of other competitors are hosted too, possibly because it enables *i* to engage with them in “cheap talk” communications, besides having the opportunity to learn about *j*’s decision making process.

Acknowledgements

The writing of this thesis was supported by scholarships from the Einaudi Institute for Economics and Finance and University College London (Ricardo Scholarship). I am very grateful to my supervisors Adam Rosen and Aureo de Paula for their suggestions, guidance and encouragement. I also thank all professors at the Department of Economics of University College London. Finally, I would like to thank my family and my partner Tommaso Di Chio for endless support.

Contents

List of Figures

List of Tables

General introduction

There is a successful literature showing that the structure of relations (network of links) between individuals or organizations affects several outcomes of economic interest, such as academic achievement (Sacerdote, 2001, 2011; Zimmerman, 2003; Calvó-Armengol, Patacchini and Zenou, 2009), job search (Topa, 2001; Calvó-Armengol and Jackson, 2004; Bayer, Ross and Topa, 2008), smoking decisions (Gaviria and Raphael, 2001; Nakajima, 2007; Bisin, Moro and Topa, 2011), criminal actions (Calvó-Armengol and Zenou, 2004), technology adoption (Bandiera and Rasul, 2006; Conley and Udry, 2010), executives' compensation (Hallock, 1997; Core, Holthausen and Larcker, 1999; Patnam, 2013; Gayle, Golan and Miller, 2015), and firms' performance (Uzzi, 1996; McDonald and Westphal, 2003). This recognition has fuelled a growing attention for estimating models explaining how the networks themselves arise (network formation models). Indeed, understanding network genesis is crucial to evaluate policies that might improve some relevant network features, or to investigate which agents to target in order to affect, in turn, behaviours on networks. On the other hand, estimating network formation models is challenging because agents' decisions for links are often interdependent, with resulting identification and computational issues when attempting to conduct inference on underlying incentives.

Chapter 1 provides a summary of the relevant literature on the econometrics of network formation models. It starts with discussing statistical models of network formation, which are characterised by a focus on the probability distribution of the network as the direct object of interest. Examples include the Erdös-Rényi model, the Poisson random graph model, and the class of exponential random graph models. As exponential random graph models have been extensively used because they can easily combine several types of interdependencies among links, available techniques to estimate them are analysed. Moreover, when fitting the Poisson random graph model to data, one may fail to reproduce several features of real networks. This has led researchers to develop statistical models of network formation where nodes are born sequentially and create connections more or less randomly, with the purpose of matching the main characteristics of observed networks. In this context, special attention is reserved to the milestone work by Jackson and Rogers (2007): after having discussed and replicated the main results of the paper, an extension of the original model is examined and fitted to a dataset of Google Plus users.

Even if statistical models can reproduce relatively well the main characteristics of real networks, they usually lack of microfundation, essential for counterfactual analysis. The chapter hence moves to considering the econometrics of economic models of network formation, where heterogenous agents create links according to specific rules, an explicit equilibrium concept, and payoffs depending on players' characteristics, and, possibly, actions (also called externalities).

Chapter 2 studies identification of the parameters governing agents' preferences in a static game of network formation, where links represent asymmetric relations between players, e.g., the sharing of directors across firms (firm with an executive sitting on the board of another company vs such a company), trading connections (buyer vs seller), and advice ties (advisor vs advisee). Agents have complete information and play pure strategy Nash equilibrium if link creation can be unilaterally established, or pure strategy pairwise Nash equilibrium in the bilateral case. Payoffs are non-transferable. Link decisions are interdependent, as the payoff that player i receives from linking to player j is affected by the number of additional players doing the same. For example, when examining firms' decisions to have their executives sitting on the board of other companies, the number of additional firms having an executive appointed on the board of company j may encourage firm i to join j's board too for exploiting "cheap talk"[1] opportunities with them. Likewise,

[1]Costless, informal, and unverifiable discussions about prior or planned choices (Farrell and Rabin, 1996).

in the analysis of trading connections, the number of additional agents buying from agent j may negatively affect agent i's power to bargain with j. Similarly, when considering advice ties, e.g., individuals nominating who they would ask for advice on the adoption of a technology, the number of additional people designating person j as adviser might proxy, in the eyes of individual i, j's time availability to offer proper explanations, or, if agents benefit from coordinating their subsequent advised choices, j's recommendations sharing level.

In order to show existence of an equilibrium, the network formation game is decomposed in some local games[2], which are similar in structure to entry games, and are such that the network formation game has an equilibrium if and only if each local game has an equilibrium. In turn, existence of an equilibrium in each local game is proved by combining Tarski's fixed point theorem with the constructive proof that Berry (1992) designs to verify existence of an equilibrium in an entry game with substitution effects.

The network formation game admits multiple equilibria. Thus, assuming that the researcher observes a large sample of equilibrium networks, partial identification arguments for the parameters of the model are developed avoiding restrictions on equilibrium selection mechanism. After having represented the sharp identified set by bounding the empirical probability distribution of the entire network, it is noticed that, when there are four or more players, such a characterisation of the sharp identified set comprises a prohibitively enormous quantity of moment inequalities.

To attenuate the computational difficulties, it is proposed to restrict the attention to the local games mentioned earlier and derive moment inequalities by bounding the empirical probability distribution of the outcomes of the local games, rather than of the network formation game. After having derived some sufficient conditions under which focusing on the local games preserves sharpness, it is noticed that, despite the procedure notably diminishes the number of moment inequalities to consider, the reduction is not enough and chasing sharpness remains unfeasible when there are ten or more players. It is then suggested to use a specific computationally

[2]A local game of the network formation game is intended as a game whose sets of players and strategy profiles are subsets of the network formation game's.

convenient sub-collection of the original list of moment inequalities involving the outcomes of the local games.

When estimating the characterised outer set, one gets computational gains from two sources, under general assumptions. Firstly, the list of moment inequalities to consider is substantially shorten. Secondly, checking the violation in the data of those moment inequalities is easy. Specifically, when obtaining by simulation the bounding terms, one can avoid verifying whether each of all possible outcomes of a local game is an equilibrium for every drawn value of preference shocks, an extremely demanding routine even for a moderate number of players. Indeed, by applying Tarski's fixed point theorem and reinterpreting for the local games the result from Berry (1992) on the fixed number of entrants in an entry game with substitution effects, the amount of outcomes which can strive for being an equilibrium is considerably reduced. Overall, Monte Carlo exercises reveal that conducting inference on the suggested outer set is computationally manageable with relatively limited computational resources when there are up to twenty players.

In Chapter 3 an empirical illustration shows that the methodology developed in Chapter 2 can deliver economically meaningful estimates. Specifically, the procedure is used to investigate firms' incentives for having executives sitting on the board of competitors (also called primary horizontal board interlocks[3]).

Most organisations are governed by a board of directors composed of executives and non-executives. The former lead the decision making process, the latter are involved in the supervision and advising of executives. Primary horizontal board interlocks are a common arrangement of firms' governance structure in several European countries[4]. Deeply analysed by corporate governance experts, they also draw the attention of economists because they may help firms to exchange information, and, in turn, reduce strategic uncertainty, transmit tacit knowledge,

[3]The adjective "horizontal" denotes the fact that firms are competitors (Carrington, 1981); the adjective "primary" denotes the fact firms share directors with executive roles (Stokman, Van Der Knoop and Wasseur, 1988).

[4]Legislation on primary horizontal board interlocks is not uniform across countries. For example, in the U.S., primary horizontal board interlocks are illegal under the Clayton Act of 1914 and subsequent ancillary legislation. By contrast, European countries do not impose any clear and general prohibition.

increase transparency, or encourage coordination. In such a scenario, interdependence among companies' decisions for forming primary horizontal board interlocks becomes crucial because it allows them to expand and radiate the flow of information. Indeed, firm i could find extremely attractive to have one of its executives sitting on the board of rival j when executives of other competitors are hosted too, as it would enable i to engage with them in "cheap talking" about past of future choices, besides having the opportunity to learn about j's decision making process.

At the same time, such interdependence causes endogeneity, and, thus, prevents the possibility of using standard econometric techniques to conduct inference on firms' preferences behind primary horizontal board interlocks. Applying the methodology illustrated in Chapter 2 represents an alternative. In particular, a 95% confidence region for the suggested outer set is constructed using Italian data. In line with the intuition above, results reveal that firms prefer to have their executives sitting on the board of a rival when executives of other competitors are appointed too.

Chapter 1

A review of the literature

1.1 Introduction

The chapter summarises the relevant literature on the econometric analysis of network formation models. Detailed reviews are in Graham (2015), Chandrasekhar (2016), and de Paula (2016). After a sketch of the main tools from network theory that will be used throughout the dissertation (Section 1.2), Section 1.3 considers statistical models of network formation. These models are characterised by a focus on the probability distribution of the network as the direct object of interest. Examples include the Erdös-Rényi model, the Poisson random graph model, and and the class of exponential random graph models (hereafter ERGMs). As ERGMs have been extensively used because they can easily combine several types of interdependencies among links, available techniques to estimate them are discussed. Moreover, when fitting the Poisson random graph model to data, one may fail to reproduce several features of real networks. This has led researchers to develop statistical models of network formation where nodes are born sequentially and create connections more or less randomly, with the purpose of matching the main characteristics of real networks. In this context, special attention is reserved to the milestone work by Jackson and Rogers (2007) (hereafter JR): after having discussed and replicated the main results of the paper, an extension of the original model is examined and fitted to a dataset of Google Plus users.

Statistical models of network formation are usually lacking micro-fundation.

Conversely, in economic models of network formation, heterogenous agents create links according to specific rules (decisions can be made simultaneously or sequentially, unilaterally or bilaterally; information can be complete or incomplete), an explicit equilibrium concept (e.g., Nash equilibrium, Nash stability, pairwise stability, pure strategy pairwise Nash equilibrium, Bayesian Nash equilibrium), and payoffs (transferable or nontransferable) depending on players' actions and characteristics. Estimating economic models of network formation can help to analyse the drivers of link creation and, thus, which policies might improve the network features of interest. Empirically understanding the forces and incentives determining the shape of real networks is also crucial to investigate which agents to target in order to affect, in turn, their behaviours on networks. Lastly, accounting for the endogeneity of link formation may be important to pin down peer effects spreading through ties. Available strategies to identify and estimate economic models of network formation are illustrated in Section 1.4, with a particular focus on the econometric challenges encountered in static games of network formation displaying multiple equilibria.

1.2 Useful tools from network theory

A network of size N can be graphically represented by a collection of N nodes (or vertices) controlled by agents, some of them connected by links (or edges, ties). Nodes are labelled by the integers in $\mathcal{N} := \{1,2,...,N\}$. The link from node i to node j is denominated link ij. The set of links is denoted by $\mathcal{E}$ and it has cardinality E. Alternatively, a network of size N can be represented by an $N \times N$ matrix $\boldsymbol{G}$ with ijth component

$$G_{ij} := \begin{cases} 1 & \text{if the link } ij \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

A network is undirected when links do not have directions (links are undirected), i.e., $G_{ij} = G_{ji}\ \forall i, j \in \mathcal{N}$. Undirected networks are used to represent symmetric relations between agents, such as friendships[5], coauthorships, risk sharing, and spatial

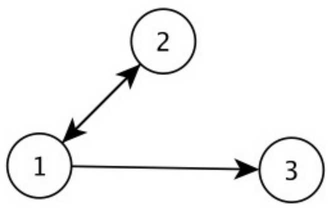
[5]The interpretation of friendship relations is controversial in the empirical literature on networks: some papers (e.g., Christakis, et al., 2010; de Paula, Richards-Shubik and Tamer, 2016; Miyauchi, 2016; Sheng, 2016) represent friendship networks as undirected networks, hence, con-

proximity. Viceversa, a network is directed when links have direction (links are directed), i.e., G_{ij} is allowed to be different from G_{ji} for any $i, j \in \mathcal{N}$. Directed networks are appropriate to handle asymmetric relations between agents, such as supply chain links, firms having their executives sitting on other companies' boards[6], advice patterns, trust connections, citations, hyperlinks, and communication ties[7]. As an example, Figure 1.1 reports a directed network of size 3.

When a network is directed, node i's direct neighbourhood is the set $\{j \in \mathcal{N} | G_{ij} = 1\}$. Links from node i to other nodes are i's outgoing links. Vice versa, links from any other node to node i are i's incoming links. Node i's in-degree (out-degree) is the number of i's incoming (outgoing) links and is denoted by d_i ($\tilde{d_i}$).

A network can be described using summary statistics, such as the maximum distance between any pair of nodes (diameter), the tendency of linked nodes to have common neighbours (clustering coefficient), the ratio of the number of links to the number of possible links (density), the average degree of nodes, the percentage of isolated nodes, and the total number of links[8].

Figure 1.1: Example of a directed network of size 3, with $\boldsymbol{G} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.



sidering the formation of links as a bilateral process. Other papers (e.g., JR; Badev, 2014; Boucher, 2016; Mele, 2017) interpret friendships "*as a mean to transmit influence, i.e.,* [...] *as role models.*" (Badev, 2014, p.9) and picture friendship networks as directed networks formed through an unilateral process, also motivated by the fact that a considerable portion of friendship nominations in the AddHealth survey is not reciprocal (see e.g., Calvó-Armengol, Patacchini, Zenou, 2009 for more details).

[6]More details are in Chapter 3.

[7]Also some online social networks can be represented as undirected networks (e.g., Facebook) or directed networks (e.g., Twitter and GooglePlus).

[8]More on summary statistics is e.g., in Jackson (2009).

1.3 Statistical models of network formation

Statistical models of network formation are characterised by a focus on the probability distribution of the network as the direct object of interest. Unless differently specified, this section considers undirected networks.

The Erdós-Rényi model The Erdós-Rényi model imposes a uniform probability on the class of networks with a given number of nodes and edges (Erdós and Rényi, 1959; 1960). Hence, each network with N nodes and E edges has probability

$$\frac{1}{\sum_{k=1}^{\frac{1}{2}N(N-1)} \binom{\frac{1}{2}N(N-1)}{k} \binom{k}{E}}$$

to arise.

The Poisson random graph model The Poisson random graph model assumes independent and identical probability $p \in (0,1)$ of link formation for each pair of nodes (Solomonoff and Rapoport, 1951; Gilbert, 1959; Erdós and Rényi, 1960). Given N nodes, it follows that node i's probability to have degree d_i is $\binom{N-1}{d_i} p^{d_i}(1-p)^{N-1-d_i}$. Moreover, its expected degree is

$$\sum_{d=0}^{N-1} d \binom{N-1}{d} p^d (1-p)^{N-1-d} = (N-1)p$$

In fact, by the binomial theorem,

$$(p+q)^{N-1} = \sum_{d=0}^{N-1} \binom{N-1}{d} p^d q^{N-1-d}$$

Then, differentiating both sides with respect to p, one gets

$$(N-1)(p+q)^{N-2} = \sum_{d=0}^{N-1} \binom{N-1}{d} d \frac{p^d}{p} q^{N-1-d}$$

and, imposing $q = 1-p$, the result is established. Lastly, keeping the expected degree constant as $N \to \infty$, the probability distribution of node i's degree can be approximated in large networks by a Poisson distribution with parameter $\gamma = (N-$

$1)p$, hence the name Poisson random graph model[9].

ERGMs The class of the ERGMs (Frank and Strauss, 1986; Frank, 1991; Wasserman and Pattison, 1996) is a generalisation of the Poisson random graph model. In more details, given N nodes, let the network $\boldsymbol{G}$ be a random matrix with support $\mathcal{G} := \{0,1\}^{\frac{N(N-1)}{2}}$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $R_1(\boldsymbol{G}), ..., R_K(\boldsymbol{G})$ denotes a collection of observable features of the network $\boldsymbol{G}$, such as number of edges, degree sequence, and diameter. In ERGMs one assumes that

$$\mathbb{P}(\boldsymbol{G} = \boldsymbol{g}) = \frac{e^{\sum_{k=1}^{K} \theta_k R_k(\boldsymbol{g})}}{Z(\theta_1, ..., \theta_K)} \tag{1.1}$$

for any $\boldsymbol{g} \in \mathcal{G}$, where $Z(\theta_1, ..., \theta_K)$ is a normalisation constant obtained from imposing

$$\sum_{\boldsymbol{g} \in \mathcal{G}} \mathbb{P}(\boldsymbol{G} = \boldsymbol{g}) = \sum_{\boldsymbol{g} \in \mathcal{G}} \frac{e^{\sum_{k=1}^{K} \theta_k R_k(\boldsymbol{g})}}{Z(\theta_1, ..., \theta_K)} = 1$$

and $(\theta_1, ..., \theta_K) \in \Theta \subseteq \mathbb{R}^K$. The probability distribution (1.1) belongs to the exponential family and the vector $(R_1(\boldsymbol{G}), ..., R_K(\boldsymbol{G}))$ is a sufficient statistic for $(\theta_1, ..., \theta_K)$[10].

It can be shown that the Poisson random graph model with parameter p is the ERGM where the number of edges $E(\boldsymbol{G})$ is the only feature considered and $\theta := \log\left(\frac{p}{1-p}\right)$. Indeed, in that case

$$\begin{aligned} Z(\theta) := \sum_{\boldsymbol{g} \in \mathcal{G}} e^{\theta E(\boldsymbol{g})} &= \sum_{\boldsymbol{g} \in \mathcal{G}} e^{\theta \sum_{i=1}^{N} \sum_{j=i+1}^{N} g_{ij}} = \sum_{\boldsymbol{g} \in \mathcal{G}} \prod_{i=1}^{N} \prod_{j=i+1}^{N} e^{\theta g_{ij}} = \\ &= \prod_{i=1}^{N} \prod_{j=i+1}^{N} \sum_{g_{ij} \in \{0,1\}} e^{\theta g_{ij}} = \prod_{i=1}^{N} \prod_{j=i+1}^{N} (1 + e^{\theta}) = (1 + e^{\theta})^{\binom{N}{2}} \end{aligned}$$

[9]Extensions of the Poisson random graph model are proposed e.g., by Zheng, Salganik and Gelman (2006) and Hong and Xu (2015).

[10]Extensions of ERGMs are proposed e.g., by Pattison and Wasserman (1999), Robins, Pattison and Wasserman (1999), and Chandrasekhar and Jackson (2014). Chandrasekhar and Jackson (2014) also show that their model is micro-funded.

and the expected value of $E(\boldsymbol{G})$ is

$$\mathbb{E}_\theta(E(\boldsymbol{G})) := \sum_{\boldsymbol{g}\in\mathcal{G}} E(\boldsymbol{g})\mathbb{P}(\boldsymbol{G}=\boldsymbol{g}) = \frac{1}{Z(\theta)}\sum_{\boldsymbol{g}\in\mathcal{G}} E(\boldsymbol{g})e^{\theta E(\boldsymbol{g})} = \frac{1}{Z(\theta)}\frac{\partial}{\partial\theta}\sum_{\boldsymbol{g}\in\mathcal{G}} e^{\theta E(\boldsymbol{g})} = \\ = \frac{1}{Z(\theta)}\frac{\partial Z(\theta)}{\partial\theta} = \binom{N}{2}\frac{e^\theta}{1+e^\theta} \tag{1.2}$$

Consider now the Poisson random graph model with parameter p, where

$$\mathbb{E}_p(E(\boldsymbol{G})) = \binom{N}{2}p \tag{1.3}$$

Then, (1.3) is equivalent to (1.2) when

$$p := \frac{e^\theta}{1+e^\theta} \Leftrightarrow \theta := \log\left(\frac{p}{1-p}\right)$$

Estimating ERGMs ERGMs have been extensively used because they can easily combine several types of interdependencies among links. Moreover, recent papers providing their microfundation, e.g., Mele (2017)[11], have made these models even more attractive. However, estimating their parameters may be difficult. In fact, assuming that one large network sampled from (1.1) is observed and some regularity conditions are satisfied, one may attempt to compute the maximum likelihood estimator of $(\theta_1,...,\theta_K)$ as

$$(\hat{\theta}_1^{ML},...,\hat{\theta}_K^{ML}) := argmax_{(\tilde{\theta}_1,...,\tilde{\theta}_K)\in\Theta} \sum_{k=1}^{K}\tilde{\theta}_k R_k(\boldsymbol{G}) - \log(Z(\tilde{\theta}_1,...,\tilde{\theta}_K))$$

with corresponding first order conditions

$$\mathbb{E}_{(\hat{\theta}_1^{ML},...,\hat{\theta}_K^{ML})}(R_k(\boldsymbol{G})) = R_k(\boldsymbol{G}) \qquad\qquad k=1,...,K$$

as it can be shown that

$$\frac{\partial\log(Z(\tilde{\theta}_1,...,\tilde{\theta}_K))}{\tilde{\theta}_k} = \mathbb{E}_{(\tilde{\theta}_1,...,\tilde{\theta}_K)}(R_k(\boldsymbol{G})) \qquad\qquad k=1,...,K$$

[11] See the discussion of Mele (2017) in Section 1.4.

by following standard derivations for the exponential family of probability distributions. Nonetheless, in most of the cases, the normalisation constant $Z(\tilde{\theta}_1, ..., \tilde{\theta}_K)$ can not be analytically computed and should be obtained by summing over all $2^{\frac{N(N-1)}{2}}$ possible networks, a task rarely possible even for moderate N, which makes the standard likelihood-based approach infeasible.

Three alternative procedures have been developed in the literature to overcome this issue. The most popular consists in Markov Chain Monte Carlo (hereafter MCMC) estimation methods (Snijders, 2002; Kolaczyk, 2009) which are based on the simulation of the probability distribution (1.1) through a sampling algorithm to avoid the direct computation of $Z(\tilde{\theta}_1, ..., \tilde{\theta}_K)$. However, as argued by Snijders (2002), designing a sampling algorithm that converges to the target distribution is not an obvious task and requires careful investigation when there are parameter values for which the function (1.1) has a multimodal shape. Additionally "*Given the huge set of networks to sample, any MCMC procedure can visit only an infinitesimal portion of the set* [...]. *Unfortunately, important recent papers have shown that for broad classes of ERGMs standard MCMC procedures will take exponential time to mix unless the links in the network are approximately independent* [...]. *Of course, if links are approximately independent then there is no real need for an ERGM specification to begin with, and so in cases where ERGMs are really needed they cannot be accurately estimated by such MCMC techniques.*" (Chandrasekhar and Jackson, 2014, p.2-3)[12].

The second strategy is to maximise the pseudo-likelihood, obtained by considering the probability that each link is formed conditional on the remaining section of the network (Besag, 1975; Strauss and Ikeda, 1990; Frank, 1991; Wasserman and Pattison, 1996). The properties of the resulting estimator are unclear and simulation-based comparisons with the classical maximum likelihood estimators are provided e.g., by Robins, Snijders, Wang, Handcock, and Pattison (2007). When links are assumed to be independent among each other, the pseudo-likelihood coincides with the original likelihood. As the likelihood for this last case can be de-

[12] Similar considerations are made e.g., by Rinaldo, Fienberg, and Zhou (2009) and Mele (2017).

rived from the equilibrium conditions in an economic model of network formation (dyadic models), its discussion is postponed to Section 1.4.

The third avenue is the use of variational principles which allow to represent $Z(\tilde{\theta}_1, ..., \tilde{\theta}_K)$ as the solution of an optimisation problem. More details are e.g., in Jordan (2004), Jordan and Wainwright (2008), Braun and McAuliffe (2010), and Chatterjee and Diaconis (2013).

Finally, even if the estimation of ERGMs is made possible, identification of parameters and consistency of estimates are not guaranteed. As pointed out by Chandrasekhar and Jackson (2014) at p.3: "*Given that data in many settings consist of a single network or a handful of networks, we are interested in asymptotics where the number of nodes in a network grows. However, it may be the case that increasing the number of nodes does not increase the information* [...]. *With non-trivial interdependencies between links, standard asymptotic results do not apply. This does not mean that consistency is precluded, (just as it is not precluded in time series or spatial settings) as there is still a lot of information that can be discerned from the observation of a single large network. Nonetheless, it does mean that asymptotic analyses must account for potentially complex interdependencies in link formation.*" .

Sequential models When fitting the Poisson random graph model to data, one may fail to reproduce the following features of real networks: (i) the degree probability distribution has fat tails, revealing the presence of more nodes with high and low degree and fewer nodes with medium degree than in Poisson random graph models; (ii) small diameter; (iii) high clustering coefficient; (iv) inverse relation between the degree of a node and its neighbourhood's clustering coefficient[13]; (v) positive assortativity[14] (JR). Hence, researchers have begun to develop statistical models of network formation where nodes are born sequentially and create connections more or less randomly, with the purpose of matching the main characteristics of observed networks.

[13]The neighbours of a higher degree node are less likely to be linked to each other compared to the neighbours of a lower degree node.

[14]High degree nodes are more likely to be linked to other high degree nodes, and low degree nodes are more likely to be linked to other low degree nodes.

In this direction, Barabási and Albert (1999) are the firsts to show that, by introducing the so called preferential attachment, i.e., new born nodes are more likely to link with existing vertices featuring high degrees, the resulting degree probability distribution has fat tails, and, more specifically, is power law. Pennock, et al (2002) observe that the degree probability distribution in several real networks is not purely power law but shows more complicated shapes, which the authors replicate by adding random link formation to the framework of Barabási and Albert (1999). JR design a sequential model of network formation with homogeneous nodes and directed links that combines random meetings, i.e., new born nodes meet at random some of the existing nodes, and network-based meetings, i.e., new born nodes meet at random some of the neighbours of the found incumbents. Network-based meetings give to the process a preferential attachment flavour, as neighbours with a high in-degree are more likely to be found. Additionally, the simultaneous combination with random meetings allows to reproduce all of the five features listed above.

More recently, the literature has focused on extending the model in JR. For example, to understand homophily[15] patterns over time, Bramoullé and Rogers (2010) and Bramoullé, et al. (2012), distinguish nodes among different types. Atalay (2013) gives vertices a different rate of expectation of attracting links, drawn from a certain parametric distribution for each age group, which allows to reproduce the empirically observed weak correlation between nodes' age and in-degree. Chaney (2014) proposes a geographical reinterpretation to explain the heterogeneous ability of firms to access foreign markets.

Given the importance of the work by JR, the next section discusses and replicates some of the main results. Moreover, an extension of the original model is examined and fitted to a dataset of Google Plus users.

1.3.1 Jackson and Rogers (2007)

Consider a discrete time process with periods $\{1,...,T\}$. For any period $t \in \{1,...,T\}$, the network $\boldsymbol{G}_t$ is derived from the network $\boldsymbol{G}_{t-1}$ as follows. In period t a new node joins the set of existing nodes $\mathcal{N}_{t-1}$. The new born node meets m_r

[15] The term homophily refers to the tendency of linked agents to have similar characteristics.

nodes, called parents, uniformly randomly drawn without replacement from $\mathcal{N}_{t-1}$ (random meetings). A link pointing to a parent is formed by the new born node if the marginal utility that the new born node receives from the relation is positive, which happens with probability p_r independent of $\boldsymbol{G}_{t-1}$. The new born node meets also m_n nodes uniformly randomly drawn without replacement from the unified set of parents' direct neighbourhoods (network-based meetings). A link pointing to one of these nodes is formed if the marginal utility that the new born node receives from the relation is positive, which happens with probability p_n independent of $\boldsymbol{G}_{t-1}$.

The parameters of the model are m_r, p_r, m_n, p_n. $m = p_r m_r + p_n m_n$ is the expected number of outgoing links formed by a new born node[16]. $r = \frac{p_r m_r}{p_n m_n}$ is the ratio between the expected number of outgoing links formed by a new born node through random meeting and those formed through network-based meeting, and it represents the randomness of the meeting process.

The model implies that the probability for node i to attract a new link in period t is roughly[17]

$$p_r\Big(\frac{m_r}{t-1}\Big) + p_n\Big(\frac{d_{i,t-1}m_r}{t-1}\frac{m_n}{m_r(m_r p_r + m_n p_n)}\Big) = \frac{p_r m_r}{t-1} + \frac{p_n m_n d_{i,t-1}}{(t-1)m}$$

(equation 2 in JR) and coincides with the expected number of additional incoming links formed by i. The first component of the sum is probability that i is found by a new born node through random meetings and linked to. The second component of the sum is the probability that i is found by a new born node through network-based meetings and linked to. The second component reveals the preferential attachment flavour of the model in JR: the higher i's in-degree, the more i is likely to be found through network-based meetings and linked to.

To derive the stationary cumulative distribution function of nodes' in-degree,

[16] The expected number of a node's outgoing links is fixed across nodes. The number of a node's outgoing links does not vary over time. The expected number of incoming links differs across nodes depending on their age because of the network-based meetings. The number of a node's incoming links increases over time.

[17] Sampling of nodes happens without replacement. Moreover, some parents can be in other parents' neighbourhoods and the union of parents' neighbourhoods can have cardinality less than m_n. However, such approximation is close to the truth when t is large and m_n is small.

the authors use a mean-field approximation assuming that (a) any node i deterministically gains $\frac{p_r m_r}{t-1} + \frac{p_n m_n d_{i,t-1}}{(t-1)m}$ incoming links in each period $t \in \{1, ..., T\}$, and (b) the evolution of nodes' in-degree is a continuous function of time and in-degree[18]. Specifically, Theorem 1 in JR states that the stationary cumulative distribution function of nodes' in-degree in each period $t \in \{1, ..., T\}$ is

$$F_t(d) = 1 - \left(\frac{d_0 + rm}{d + rm}\right)^{(1+r)} \tag{1.4}$$

when $p_n m_n \neq 0$, and

$$F_t(d) = 1 - \exp^{(d_0 - d)/(p_r m_r)} \tag{1.5}$$

when $p_n m_n = 0$, where $d_0 \geq 0$ is the initial in-degree of a new born node[19] and $d \geq d_0$. By taking the logarithm of both sides, one gets

$$\log(1 - F_t(d)) = (1+r)[\log(d_0 + rm) - \log(d + rm)] \tag{1.6}$$

when $p_n m_n \neq 0$, and

$$\log(1 - F_t(d)) = \frac{d_0 - d}{p_r m_r} \tag{1.7}$$

when $p_n m_n = 0$.

Notice that when $p_n = 1$ and $p_r = 0$, i.e., all links are network-based[20], it follows that $r = 0$ and $m = m_n$. Hence,

$$1 - F_t(d) = d_0 d^{-1}$$

which leads to a power law probability distribution. Viceversa, when $p_n = 0$ and $p_r = 1$[21], i.e, all links are network-based, it follows that $r = \infty$ and $m = m_r$. Hence,

$$1 - F_t(d) = \exp^{(d_0 - d)/m}$$

[18] The exact in-degree probability distribution is calculated by Atalay (2013) using a result from Dorogovtsek, Mendes, and Samukhin (2000).

[19] $d_0 > 0$ when $p_r = 0$ otherwise no links are formed.

[20] Also called pure preferential attachment case.

[21] Also called uniformly random network.

which leads to an exponential probability distribution.

To check the goodness of the mean-field approximation, the authors simulate the model and compare the empirical cumulative distribution function of nodes' in-degree with the predicted one. Figure 1.2 replicates Figure 1 in JR . The three panels are obtained by setting $T = 25,000$, $m_r = m_n = 10$, $d_0 = 1$, and an arbitrary initial number of nodes $m_0 \geq m_r + m_n + 1$, as indicated in the paper. The blue line is the prediction from (1.6) or (1.7), the magenta line is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree in the simulated network. Panel (a) is obtained by imposing $p_r = 0$ and $p_n = 1$. Panel (b) is obtained by imposing $p_r = p_n = 1$. Panel (c) is obtained by imposing $p_r = 1$ and $p_n = 0$. In all cases, it can be noticed that the predicted in-degree probability distribution well approximates the empirical one. Moreover, as expected, the decay accelerates across panels.

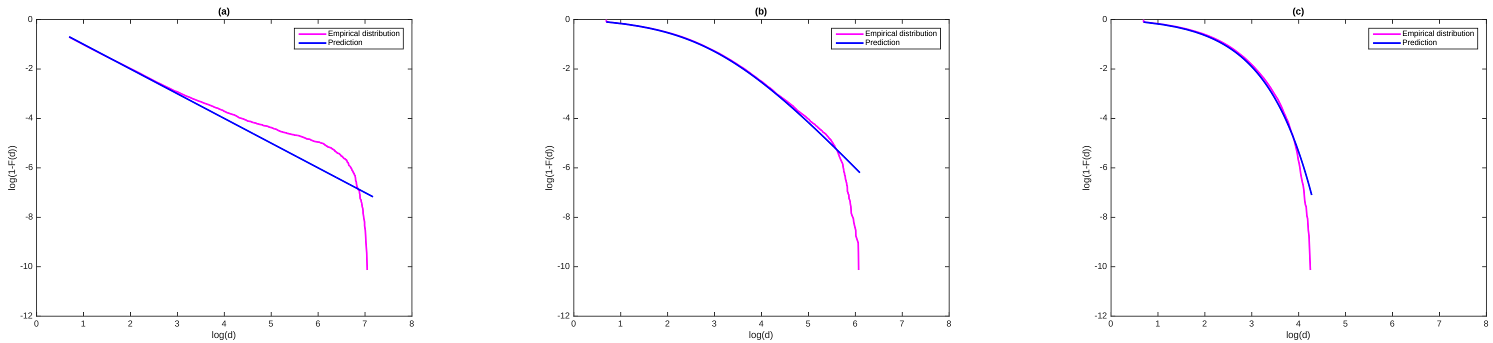


Figure 1.2: This figure replicates Figure 1 in JR. The three panels are obtained by setting $T = 25,000$, $m_r = m_n = 10$, $d_0 = 1$, and an arbitrary initial number of nodes $m_0 \geq m_r + m_n + 1$, as indicated in the paper. The blue line is the prediction from (1.6) or (1.7), the magenta line is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree in the simulated network. Panel (a) is obtained by imposing $p_r = 0$ and $p_n = 1$. Panel (b) is obtained by imposing $p_r = p_n = 1$. Panel (c) is obtained by imposing $p_r = 1$ and $p_n = 0$.

In addition to the study of the stationary cumulative distribution function of nodes' in-degree, the authors derive the closed form solution for the clustering coefficient (Theorem 2) and the diameter (Theorem 3) of the network under the mean-field approximation. Specifically, with respect to the clustering coefficient, they show how its approximated value tends to zero when $p_n = 0$ or $p_r = 0$. This contrasts with the empirical evidence suggesting, instead, that large decentralised networks have positive clustering. Viceversa, the hybrid case with $p_n > 0$ and $p_r > 0$ is able to generate a positive clustering coefficient. Positive assortativity is produced by the temporal development of the model, that causes a direct relation between the age and the in-degree of a vertex (Theorem 4). The negative clustering-degree relationship is a direct consequence of the combination between network-based meetings and a fixed number of outgoing links per node (Theorem 5).

To show its flexibility and capacity of matching the main features of real networks, the authors fit the model to six datasets: "*the links among Web sites at Notre Dame University, the network of coauthorship relations among economists publishing in journals listed by EconLit in the 1990s, a citation network of research articles stemming from Milgram's 1960 paper* [...]*, a friendship network among 67 prison inmates in the 1950s, a network of ham radio calls during a one-month period, and, finally, a network of romantic relationships among high school students.*" (JR, p.900). The results obtained for the citation network are now replicated[22]. In the network each node represents a research article that contains the words "small world" or a reference to Milgram (1967), for a total of 396 nodes. The link ij exists when the author of paper j cites article i. The dataset is organised in two columns. For each row, it reports an in-degree value (in ascending order from zero to 147) and the number of nodes characterised by that in-degree value.

Following the procedure outlined in JR, the expected out-degree m is obtained by taking the average of nodes' in-degree values. After having imposed $d_0 = 0$, r is estimated from (1.6) through an iterative least square procedure. Specifically, let r_0 be an initial guess for r. A linear regression of $\log(1 - F_t(d))$ on 1 and $\log(d + r_0 m)$

[22]The dataset can be found at `http://www.stanford.edu/~jacksonm/Data.html`.

gives the parameter value $\hat{\beta}^1$. r is updated by setting $r_1 = -1 - \hat{\beta}^1$. The procedure is repeated until convergence. Bounds for the diameter are derived using the estimates of r and m. After having imposed $p_r = p_n = p$, the average clustering coefficient C^{avg} is computed as a function of p via the mean-field approximation from Theorem 2[23]. Lastly, p is obtained by setting C^{avg} equal to the empirical average clustering coefficient[24].

The results are reported in Table 1.1 and coincide with those in JR. Specifically, the average in-degree m is equal to 5.02. The estimate of r is 0.63 and tells that the network is formed mainly through network-based meetings. As expected, this is due to the fact that often references are found because mentioned in other previously discovered articles. The estimate of p is 0.33 and reveals that the fraction of formed links tends to less than one half of the meetings.

Table 1.1: Parameter estimates of the citation network in JR.

Number of nodes	396
m	5.02
r from fit	0.63
p from fit	0.33

Figure 1.3 replicates the panel in Figure 2 of JR referred to the citation network. The red curve is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve is the prediction from (1.6). It can be noticed that the predicted in-degree probability distribution well approximates the empirical one. Moreover, r close to zero produces a roughly linear shape, as anticipated by panel (a) of Figure 1.2.

[23] As observed by the authors, the parameter p is not identified from the cumulative distribution function of nodes' in-degree.

[24] The empirical average clustering coefficient can not be computed from the dataset as nodes' labels are unknown. Its value is taken from Garfield (2006).

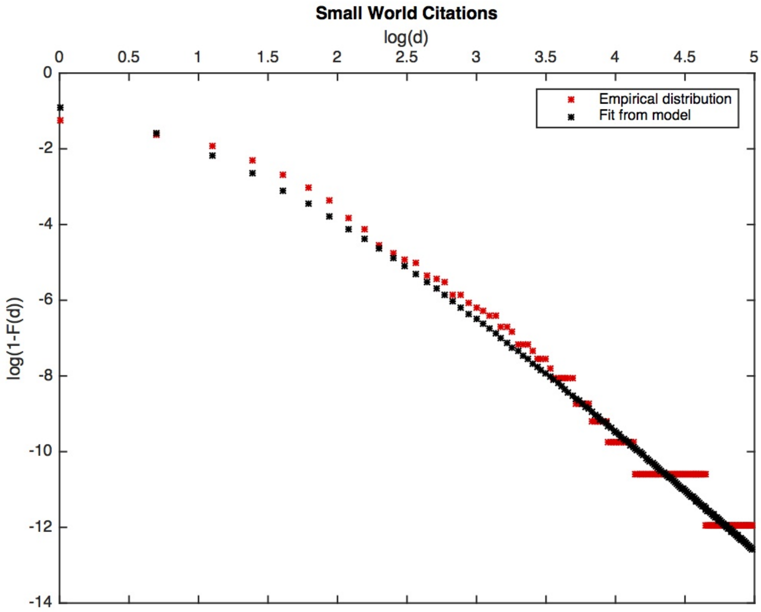


Figure 1.3: This figure replicates the panel of Figure 2 in JR referred to the citation network. The red curve is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve is the prediction from (1.6).

To test further the capacity of the model in JR to reproduce the cumulative distribution function of nodes' in-degree in modern social networks, the next part of this section fits it also to a dataset of Google Plus users[25]. More precisely, a Google Plus user profile is a public visible account of a Google user. It includes basic social networking services like a profile photo, previous work and education history, interests, and placed lived. Through the circle function, a Google Plus user can decide to share her private content with other users. This enables to build a network where each node represents a Google Plus user and the link ij exists when the Google Plus user j shares her private content with Google Plus user i. The dataset is organised in two columns. For each row, it reports the labels of two nodes connected by a link from the first to the second. The resulting network is composed of $107,596$ nodes and $60,989,732$ links.

Following the procedure outlined by JR, the expected out-degree m is computed by taking the average of nodes' in-degree values and it is equal to 283.42. Then, after having imposed $d_0 = 0$, r is estimated from (1.6) using the iterative least

[25]The dataset can be found at `https://snap.stanford.edu/data/egonets-Gplus.html`.

square procedure discussed above. The resulting r is 1.53[26].

Figure 1.4 compares the empirical probability distribution of nodes' in-degree with the predicted one. The red curve is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve is the prediction from (1.6). It can be noticed that the predicted probability distribution does not well approximate the empirical one. In particular, it seems to underestimate the preferential attachment flavour of the real process. This becomes clearer in the detail reported by Figure 1.5, where the red curve, representing a section of the complementary empirical cumulative distribution function of nodes' in-degree, shows fatter tails than the black curve, representing the same section as predicted by (1.4).

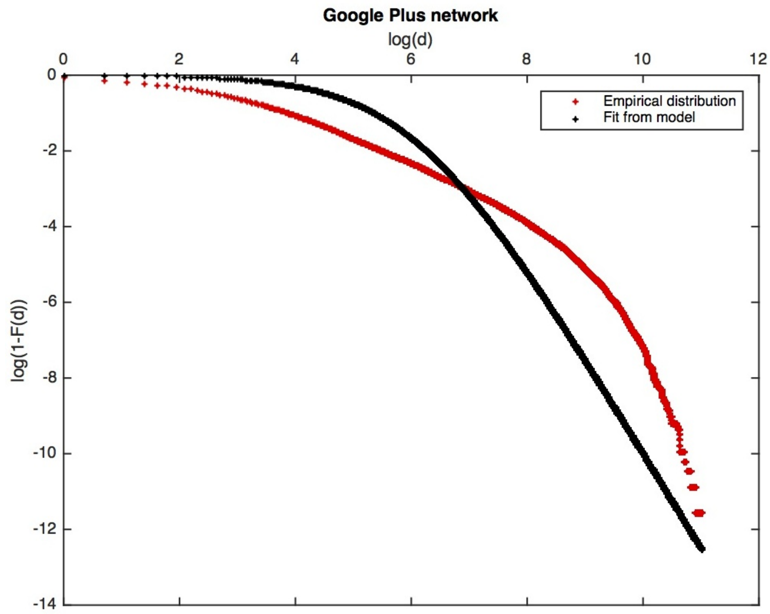


Figure 1.4: This figure reports the results obtained by fitting the model in JR to the dataset of Google Plus users described in Section 1.3. The red curve is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve is the prediction from (1.6).

[26]The empirical average clustering coefficient is 0.49 as reported at `https://snap.stanford.edu/data/egonets-Gplus.html`. When equalising C^{avg} to the empirical average clustering coefficient to derive $p_r = p_n = p$, as suggested in JR, no value of p between 0 and 1 seems to represent a solution.

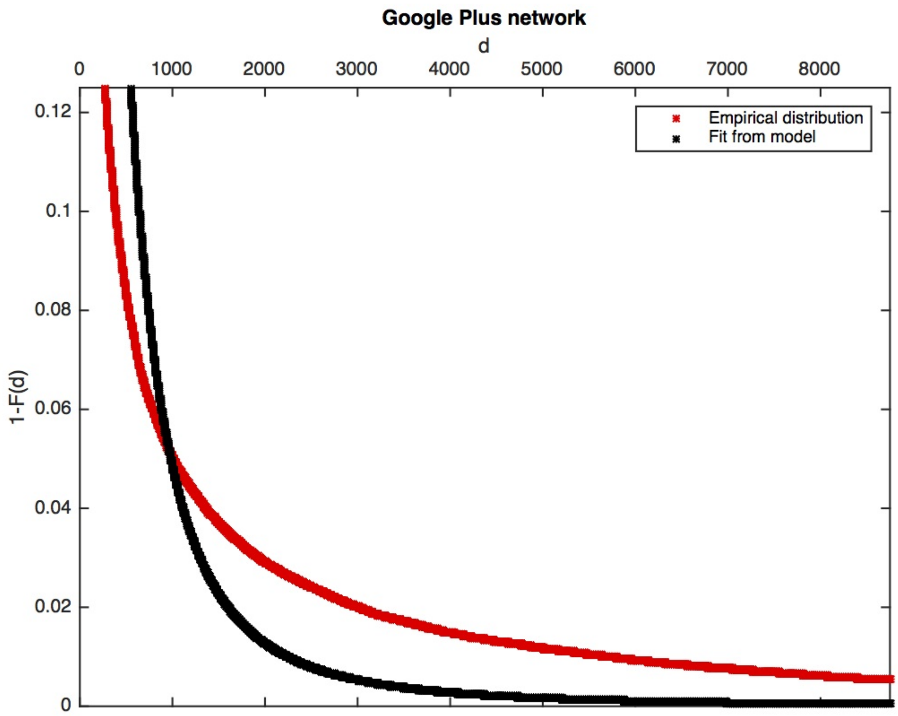


Figure 1.5: The red curve represents a detail of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve represents the same detail as predicted by (1.4).

Given these considerations, a way to improve fitting precision could be to magnify the preferential attachment inclination of the original model in JR, by assuming that the probability for new born node i to form a link pointing to node j, found through network-based meetings, is increasing with j's in-degree. Specifically, following an extension suggested by the authors at the end of the paper, let the marginal utility that new born node i gets from linking to node j found through network-based meetings be $U_{ij} := \tilde{U}_{ij}d_j - c$ where $\tilde{U}_{ij} \sim U([0,u])$ and $0 < c < u$ is a cost parameter. Hence that the probability for new born node i to form the link ij is

$$Pr(U_{ij} \geq 0) = Pr(\tilde{U}_{ij}d_j - c \geq 0) = Pr(\tilde{U}_{ij} \geq \frac{c}{d_j}) = 1 - P(\tilde{U}_{ij} \leq \frac{c}{d_j}) = 1 - \frac{c}{d_j u}$$

which is increasing with d_j. Notice that $d_j \geq 1$ since node j is found through network-based meetings. It follows that the probability for node i to attract a new link in period t is roughly

$$p_r\left(\frac{m_r}{t-1}\right)+\left(1-\frac{c}{d_{i,t-1}u}\right)\left(\frac{d_{i,t-1}m_r}{t-1}\frac{m_n}{m_r m_t}\right)=\frac{p_r m_r}{t-1}+\left(1-\frac{c}{d_{i,t-1}u}\right)\frac{m_n d_{i,t-1}}{(t-1)m_{t-1}}$$

where m_{t-1} is the expected number of outgoing links formed by a node j having a link pointing to node i and, therefore, it depends on node i's in-degree at the time when node j entered the network.

Let m be the limit of m_t for t large. By applying Lemma 1 in JR, the stationary cumulative distribution function of nodes' in-degree in each period $t \in \{1,...,T\}$ is

$$F_t(d)=1-\left(\frac{d_0+p_r\frac{m_r}{m_n}m-\frac{c}{u}}{d+p_r\frac{m_r}{m_n}m-\frac{c}{u}}\right)^{(m/m_n)} \tag{1.8}$$

when $m_n \neq 0$, and

$$F_t(d)=1-\exp^{(d_0-d)/(p_r m_r)} \tag{1.9}$$

when $m_n = 0$. By taking the logarithm of both sides, one gets

$$\log(1-F_t(d))=\frac{m}{m_n}\left[\log\left(d_0+p_r\frac{m_r}{m_n}m-\frac{c}{u}\right)-\log\left(d+p_r\frac{m_r}{m_n}m-\frac{c}{u}\right)\right] \tag{1.10}$$

when $m_n \neq 0$, and

$$\log(1-F_t(d))=\frac{d_0-d}{p_r m_r} \tag{1.11}$$

when $m_n = 0$.

Results from simulations of the extended model are reported in Figure 1.6. The blue line is the prediction from (1.10), the magenta line is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree in the simulated network. Both panels are obtained by setting $T = 300,000$, $m_r = m_n = 10$, $u = 3$, $c = 2$, $d_0 = 1$, and an arbitrary initial number of nodes $m_0 \geq m_r + m_n + 1$. Panel (a) of Figure 1.6 is obtained by imposing $p_r = 0$. Panel (b) of Figure 1.6 is obtained by imposing $p_r = 1$. In all cases, the predicted in-degree probability distribution well approximates the empirical one. Moreover, the decay accelerates across the two panels. Lastly, by comparing panels (a) and (b) respectively with panels (a) and (b) of Figure 1.2, it can be noticed that the degree distribution for the extended model shows a lower decay, hence resulting in fatter tails.

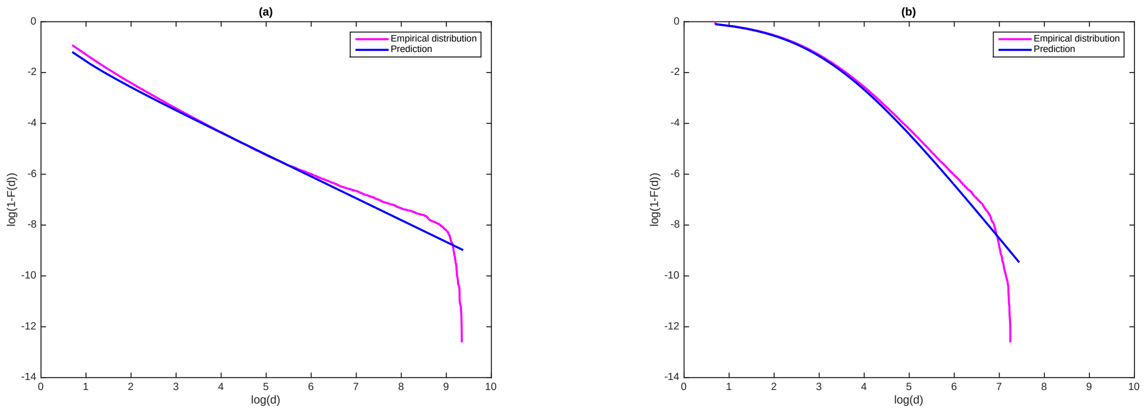


Figure 1.6: Both panels are obtained by setting $T = 300,000$, $m_r = m_n = 10$, $u = 3$, $c = 2$, $d_0 = 1$, and an arbitrary initial number of nodes $m_0 \geq m_r + m_n + 1$. The blue line is the prediction from (1.10), the magenta line is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree in the simulated network. Panel (a) is obtained by imposing $p_r = 0$. Panel (b) is obtained by imposing $p_r = 1$.

It is also made a first rough attempt to fit the extended model to the same dataset of Google Plus users. For simplicity, let $u = 1$, $c = 0$, $d_0 = 0$. Let $\frac{m_0}{m_{n,0}}$ and $p_{r,0}m_{r,0}$ be initial guesses respectively for $\frac{m}{m_n}$ and $p_r m_r$. A linear regression of $\log(1 - F_t(d))$ on 1 and $\log(d + p_{r_0}m_{r,0}\frac{m_0}{m_{n,0}})$ gives the parameter value $\hat{\beta}^1$. $\frac{m}{m_n}$ is updated by setting $\frac{m_1}{m_{n,1}} = -\hat{\beta}^1$. Then $m_{n,1}$ is isolated by using the average of nodes' in-degree as an estimate for m_1. Lastly, $p_{r,1}m_{r,1}$ is derived from $m_1 = p_{r,1}m_{r,1} + m_{n,1}$[27]. The procedure is repeated until convergence. Figure 1.7 compares the empirical probability distribution of nodes' in-degree with the predicted one. The red curve is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve is the prediction from (1.10). As expected, it can be noticed that the predicted in-degree probability distribution approximates the empirical one better than in Figure 1.4 because it shows fatter tails. Further investigations are left to future research.

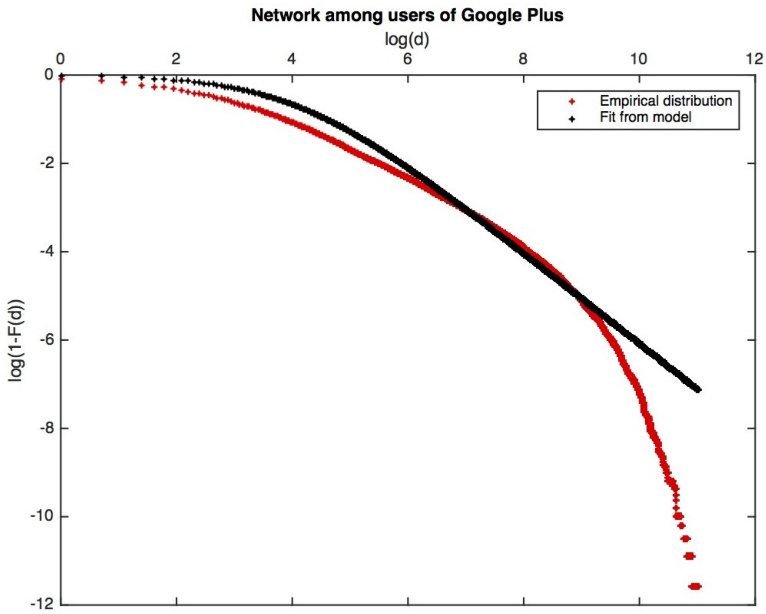


Figure 1.7: This figure reports the results obtained by fitting the extended model in JR to the dataset of Google Plus users described in Section 1.3. The red curve is the logarithm of the complementary empirical cumulative distribution function of nodes' in-degree; the black curve is the prediction from (1.10).

[27] Simulations reveal that $m \approx p_r m_r + m_n$.

1.4 Econometric analysis of economic models of network formation

In economic models of network formation agents create links according to specific rules (decisions can be made simultaneously or sequentially, unilaterally or bilaterally; information can be complete or incomplete), an explicit equilibrium concept (e.g., Nash equilibrium, Bayesian Nash Equilibrium, Nash stability[28], pairwise stability[29], pure strategy pairwise Nash equilibrium[30]), and a payoff (transferable or non-transferable) depending on players' actions and characteristics.

1.4.1 Dyadic models

Dyadic models focus separately on each pair of nodes (dyad) and the link between them. For example, Fafchamps and Gubert (2007) consider the creation of a risk sharing network among N households residing in four villages of Northern Philippines. Each node represents a household and there is a link from node i to node j if household i cites household j as source of assistance in case of need (directed network). The formation process is as follows: households simultaneously announce the desired outgoing links[31] under complete information and each household i gets as payoff

$$U_i := \sum_{j=1}^{N} G_{ij} \times \left[\theta' X_{ij} + \varepsilon_{ij}\right]$$

where X_{ij} is an $L \times 1$ vector with the lth component representing a measure of distance between households i and j (possibly asymmetric), ε_{ij} is a scalar collecting the residual variables affecting the net benefit that household i receives from the formation of the link ij which are unobserved by the researcher (also called preference shock), and $\theta \in \mathbb{R}^L$. Hence, the network $\boldsymbol{G}$ is a pure strategy Nash equilibrium if it

[28] Myerson (1991).

[29] Jackson and Wolinski (1996).

[30] Jackson and Wolinski (1996); Calvó-Armengol (2004); Bloch and Jackson (2006); Goyal and Joshi (2006); Calvó-Armengol and Ilkiliç (2009).

[31] Household i unilaterally decides about the formation of the link ij.

solves the system of $N(N-1)$ equations

$$G_{ij} = \mathbb{1}\{\theta' X_{ij} + \varepsilon_{ij} \geq 0\} \qquad \forall i, j \in \mathcal{N}, i \neq j \tag{1.12}$$

Assuming that a sample of N households is randomly drawn from the population with N large[32] and all links connecting them are observed together with $\{X_{ij}\}_{i,j \in \mathcal{N}, i \neq j}$, a consistent maximum likelihood estimator for θ based on Logit procedures can be obtained, after having imposed a specific structure of the error correlations and exogeneity of observed covariates[33]. Specifically, Fafchamps and Gubert (2007) allow $\mathbb{E}(\varepsilon_{ij}, \varepsilon_{ik})$, $\mathbb{E}(\varepsilon_{ij}, \varepsilon_{kj})$, $\mathbb{E}(\varepsilon_{ij}, \varepsilon_{jk})$, and $\mathbb{E}(\varepsilon_{ij}, \varepsilon_{ki})$ to be different from zero, $\forall i, j, k \in \mathcal{N}$ with $i \neq j \neq k$, and build on Conley (1999) to derive a a robust expression for the covariance matrix. Comola and Fafchamps (2014) extend the methodology to undirected networks for which discordant responses on link existence may be reported by households.

A more recent literature attempts to include agents' fixed effects. For example, Graham (2016) focuses on undirected networks and assumes that

$$G_{ij} = \left\{ \theta' X_{ij} + \varepsilon_{ij} + A_i + A_j \geq 0 \right\} \qquad \forall i, j \in \mathcal{N}, i \neq j \tag{1.13}$$

where A_i and A_j represent components varying with the unobserved agent-level attributes and are possibly correlated with X_{ij}. Equations in (1.13) can be interpreted as pairwise stable conditions when utility transfers are allowed and payoff functions are additively separable over links. After having imposed that $\{\varepsilon_{ij}\}_{i,j \in \mathcal{N}, i \neq j}$ are standard logistic random variables i.i.d. across ij conditional on $\{X_{ij}\}_{i,j \in \mathcal{N}, i \neq j}$, the author provides two estimators and shows their consistency and asymptotic normality. In particular, the first estimator, called tetrad logit estimator, conditions on a sufficient statistic for $\{A_i\}_{i=1}^N$ which allows to partial out the fixed effects and bypass the incidental parameter problem. The second estimator, called joint maximum likelihood estimator, considers instead $\{A_i\}_{i=1}^N$ as parameters to be estimated, hence

[32] One of the beauties of dyadic models is the non-necessity of observing the entire network.

[33] Otherwise, covariates that are suspected to be endogenous should be instrumented, as done by Fafchamps and Gubert (2007) in some robustness checks.

facing a non-standard asymptotic behaviour due to a parameter space growing with N.

Other papers allowing for fixed effects are e.g., Dzemski (2014), Candelaria (2016), Jochmans and Weidner (2016), and Jochmans (2017).

1.4.2 Models with externalities

Dyadic models do not take into account that, in several settings of economic interest, link decisions are interdependent, i.e., the presence of a link in one section of the network may affect the payoffs from link formation in other parts of the network (also called externalities). Indeed, "*A key defining feature of strategic network formation models is some form of externality that goes beyond direct links, such as the idea that friends of friends matter.*" (de Paula, Richards-Shubik and Tamer, 2016, p.1).

Among models with externalities, sequential models assume that at each iteration of a meeting protocol, a pair of perfectly informed agents is drawn at random and determines the formation, maintenance or dissolution of a link according to payoffs depending on the current structure of the network (myopic behaviour). For example, Mele (2017) focuses on directed networks and considers a process in which the payoff that agent i gets from forming the link ij[34] at iteration t is

$$\begin{aligned} U_{ij,t} := u(X_i, X_j; \theta_u) + G_{ji,t-1} \times m(X_i, X_j; \theta_m) + \sum_{k \neq i}^{N} G_{kj,t-1} \times v(X_i, X_k; \theta_v) + \\ + \sum_{k \neq j}^{N} G_{ki,t-1} w(X_k, X_j; \theta_w) + \varepsilon_{ij} \end{aligned}$$

where X_i is a $K \times 1$ vector of agent i's characteristics, $u(\cdot;\theta_u)$, $m(\cdot;\theta_m)$, $v(\cdot;\theta_v)$, $w(\cdot;\theta_w)$ are bounded and real-valued functions know by the researcher up to the finite dimensional parameters $\theta_u, \theta_m, \theta_v, \theta_w$, and $\{\varepsilon_{ij}\}_{i,j \in \mathcal{N}, i \neq j}$ are extreme value random variables i.i.d. across ij. $u(\cdot;\theta_u)$ captures the direct payoff from the link ij, $m(\cdot;\theta_m)$ represents the payoff received if the connection with agent j is mutual, $v(\cdot;\theta_v)$ denotes the payoff from agent j's links, and $w(\cdot;\theta_w)$ corresponds to a

[34] Agent i unilaterally decides about the formation, maintenance or dissolution of the link ij.

popularity effect. Mele (2017) shows that the process converges to an unique stationary distribution that belongs to the class of ERGMs (Theorem 1) and provides a MCMC estimation methods that takes into account the issues described in Section 1.3. Additionally, the author provides some identification results when nodes are homogeneous.

Badev (2014) considers a framework similar to the one in Mele (2017) but simultaneously models behavioural choices made by the agents conditional on the endogenous network structure. The author then applies the methodology to data on friendship networks and smoking decisions from the AddHealth survey.

Other papers estimating sequential models with externalities are e.g., Currarini, Jackson and Pin (2009) and Christakis, et al. (2010). Lastly, there are also works providing econometrics analysis of sequential models with forward-looking behaviour, such as Lee and Fong (2013) who study bilateral oligopoly and buyer-seller networks.

Conversely, in static models agents announce desired links simultaneously. Among the papers presuming complete information, Sheng (2016) focuses on undirected networks and specifies the payoff that agent i gets from participating to the network formation game as

$$U_i := \sum_{j=1}^{N} G_{ij} \times \left[u(X_i, X_j; \beta) + \frac{\gamma_1}{N-2} \sum_{k \neq i}^{N} G_{jk} + \frac{\gamma_2}{N-2} \sum_{k \neq i, j}^{N} G_{ik} G_{jk} + \varepsilon_{ij}^{i} \right]$$

where, for any potential friend j, the parameters γ_1 and γ_2 capture the impact on i's utility, respectively, from the number of j's links, and from the number of common links between i and j[35].

The equilibrium concept used is pairwise stability: the network $\boldsymbol{G}$ is pairwise stable if it is robust to unilateral one-link deletion and bilateral one-link formation,

[35]The normalisation by $N-2$ implies that the sum terms converge as $N \to \infty$, as shown by Sheng (2016) in Corollary 4.2.

i.e., with non-transferable payoffs,

$$G_{ij} = \mathbb{1}\{u(X_i,X_j;\beta) + \frac{\gamma_1}{N-2}\sum_{k\neq i}^{N} G_{jk} + \frac{\gamma_2}{N-2}\sum_{k\neq i,j}^{N} G_{ik}G_{jk}) + \varepsilon^i_{ij} \geq 0,$$
$$u(X_j,X_i;\beta) + \frac{\gamma_1}{N-2}\sum_{k\neq j}^{N} G_{ik} + \frac{\gamma_2}{N-2}\sum_{k\neq i,j}^{N} G_{jk}G_{ik} + \varepsilon^j_{ij} \geq 0\} \qquad \forall i,j \in \mathcal{N}, i \neq j \tag{1.14}$$

and, with transferable payoffs,

$$G_{ij} = \mathbb{1}\{u(X_i,X_j;\beta) + u(X_j,X_i;\beta) + \frac{\gamma_1}{N-2}(\sum_{k\neq i}^{N} G_{jk} + \sum_{k\neq j}^{N} G_{ik}) + \frac{2\gamma_2}{N-2}\sum_{k\neq i,j}^{N} G_{ik}G_{jk} +$$
$$+ \varepsilon^i_{ij} + \varepsilon^j_{ij}) \geq 0\} \quad \forall i,j \in \mathcal{N}, i \neq j \tag{1.15}$$

When payoffs are transferable, Sheng (2016) shows existence of a pairwise stable network for any $u(\cdot;\beta)$, γ_1, and γ_2. Viceversa, when payoffs are non-transferable, Sheng (2016) shows existence of a pairwise stable network for any $u(\cdot;\beta)$, $\gamma_1 \geq 0$, and $\gamma_2 \geq 0$ (Propositions 2.1 and 2.2).

Moreover, Sheng's model admits multiple pairwise stable networks for some values of payoff-relevant variables and parameters. When data are composed of a large number of networks, it is possible to avoid assumptions regarding how players select the outcome observed by the researcher from the equilibrium set predicted by the model (hereafter equilibrium selection mechanism)[36] and construct bounds for the empirical probability distribution of the network (moment inequalities), mimicking the most recent empirical literature on entry games (Tamer, 2003; Ciliberto and Tamer, 2009 -hereafter CT-; de Paula, 2013; Aradillas-Lopez and Rosen, 2016). Consequently, the parameters of the model may be only partially identified and the estimation of the set of all admissible parameter values (sharp identified set) can be performed using different techniques: if unconditional moment inequalities, Cher-

[36]Sequentiality is an example of restriction on the equilibrium selection mechanism: "[...] *pairwise stable networks would be rest points for link formation sequences produced via a meeting protocol if the payoff structure does not change at each new meeting. In that case, the realised sequence of meetings could be seen as a selection among the possible stable networks which the approaches described below try to be agnostic about. The articles described above circumvent this issue by introducing noise in the meeting process (as unobservables are drawn as new at each meeting opportunity).*" (de Paula, 2016, p.37).

nozhukov, Hong, and Tamer (2007); Beresteanu and Molinari (2008); Romano and Shaikh (2008; 2010); Rosen (2008); Stoye (2009); Andrews and Soares (2010) -hereafter AS-; Bugni (2010); Canay (2010); Romano, Shaikh, and Wolf (2014); Pakes, et al. (2015); Bugni, et al. (2016); Chernozhukov, Chetverikov and Kato (2016); Kaido, Molinari and Stoye (2016); Bugni, Canay and Shi (2017); if conditional moment inequalities: Andrews and Shi (2013); Chernozhukov, Lee, and Rosen (2013); Lee, Song, and Whang (2013; 2014); Amstrong (2014); Chetverikov (2017).

However, inference with multiple equilibria becomes computationally very intensive as the number of players increases because of the exponential spreading of possible networks ($2^{\frac{1}{2}N(N-1)}$) and, hence, moment inequalities. These difficulties can be attenuated by (a) restricting the attention to some local games of the network formation game, i.e., games whose sets of players and strategy profiles are subsets of the network formation game ones, and (b) constructing bounds for the empirical probability distribution of the outcomes of such local games, rather than of the network formation game. Indeed, "thinking locally" may entail a significant reduction in the number of moment inequalities to consider, thanks to the fewer support points of the bounded probability distribution. However, at the same time, ignoring "the whole picture" may cause a loss of information about players' preferences for links, thus leading the researcher to conduct inference on a set of parameter values larger than, and containing, the sharp identified set (outer set).

Different local games can be examined. For example, Sheng (2016) focuses on the local games underlying the formation of all the subnetworks of size equal to or smaller than a certain α[37], set by the researcher according to the available computational resources, with $2 \leq \alpha \leq N$.

In a similar spirit but with a focus on directed networks and non-transferable payoffs, Chapter 2 considers the N local games underlying the formation of the *section* 1, *section* 2, ..., *section N* of a network (respectively *section* 1 *game*, *section* 2 *game*,..., *section N game*), where, for any $j \in \mathcal{N}$, the *section* j of a network is

[37] A subnetwork of size α is defined by a subset of α vertices, $\tilde{\mathcal{N}} \subseteq \mathcal{N}$, and a subset of links, $\tilde{\mathcal{E}} \subseteq \mathcal{E}$, such that $\tilde{\mathcal{E}}$ contains all links in $\mathcal{E}$ connecting any two nodes in $\tilde{\mathcal{N}}$.

intended as the network portion collecting all nodes and the links pointing to node j.

Alternatively, Miyauchi (2016) restricts to positive externalities. The resulting super-modularity of the game guarantees the existence of a greatest and a lowest equilibrium by Tarski's fixed point theorem, which, in turn, allows the author to provide a feasible framework for inference. The model is estimated using data on friendship networks from the AddHealth survey. Similarly (but for directed networks), Boucher (2016) analyses a situation in which agents get positive spillovers from joining cliques[38]. However, differently from Miyauchi (2016), the author assumes that agents always play the greatest equilibrium and performs inference based on the observation of one large network.

Another paper developing econometric arguments based on the observation of one or few large networks is the work by de Paula, Richards-Shubik and Tamer (2016). Specifically, the authors characterise an outer set of parameter values (sharp under some conditions) and propose a quadratic programming algorithm to compute it, assuming that preference shocks depend on the characteristics of potential friends and not on their identities, agents can create a limited number of links, and, crucially, only connections up to a certain distance affect payoffs. Other papers with asymptotics depending on N are e.g., Menzel (2016) and Leung (2015), the latter imposing incomplete information.

Finally, identification of parameters in static game of network formation with externalities and fixed effects is an open question in the literature. For example, from Mele (2017) at p.4: "*I abstract from unobserved heterogeneity, which can be included in our model with substantial additional computational effort. However, it is not clear whether it is possible to separately identify unobserved heterogeneity from externalities using a single observation of the network* [...].". Further insights are in Graham (2016).

[38] A clique in a network is a groups of individuals, each of whom is linked to all the others.

Chapter 2

An econometric analysis of a network formation game

2.1 Introduction

The chapter studies identification of the parameters governing agents' preferences in a static game of network formation, where links represent asymmetric relations between players, e.g., the sharing of directors across firms (firm with an executive sitting on the board of another company vs such a company), trading connections (buyer vs seller), and advice ties (advisor vs advisee). Agents have complete information and play pure strategy Nash equilibrium (hereafter PSNE) if link creation can be unilaterally established, or pure strategy pairwise Nash equilibrium (hereafter PSPNE) in the bilateral case. Payoffs are non-transferable. Link decisions are interdependent, as the payoff that player i receives from linking to player j is affected by the number of additional players doing the same. For example, when examining firms' decisions to have their executives sitting on the board of other companies, the number of additional firms having an executive appointed on the board of company j may encourage firm i to join j's board too for exploiting "cheap talk" opportunities with them. Likewise, in the analysis of trading connections, the number of additional agents buying from agent j may negatively affect agent i's power to bargain with j. Similarly, when considering advice ties, e.g., individuals nominating who they would ask for advice on the adoption of a technology, the number of additional people designating person j as adviser might proxy, in the eyes of individual i, j's time availability to offer proper explanations, or, if agents benefit from coordinating their subsequent advised choices, j's recommendations sharing level.

In order to show existence of an equilibrium, the network formation game is decomposed into some local games, which are similar for structure to entry games, and are such that the network formation game has an equilibrium if and only if each local game has an equilibrium. In turn, existence of an equilibrium in each local game is proved by combining Tarski's fixed point theorem with the constructive proof that Berry (1992) designs to verify existence of an equilibrium in an entry game with substitution effects.

The network formation game admits multiple equilibria. Thus, assuming that the researcher observes a large sample of equilibrium networks, partial identification arguments for the parameters of the model are developed without restrictions on equilibrium selection, as seen in the most recent empirical literature on entry games (Tamer, 2003; Ciliberto and Tamer, 2009 - hereafter CT; Beresteanu, Molchanov and Molinari, 2011 - hereafter BMM; Aradillas-Lopez and Rosen, 2016). After having represented the region of all admissible parameter values (sharp identified set) by bounding the empirical probability distribution of the entire network as in BMM, it is noticed that the sharp identified set is characterised by a prohibitively enormous quantity of moment inequalities that makes inference on it impractical: $2^{2^{N(N-1)}} - 2$ moment inequalities for each value of parameters and exogenous observables, where N is the number of players.

Even if some moment inequalities can be shown to be redundant, constructing an algorithm to check the violation in the data of the relevant ones only - or even count them - seems unworkable to the best of the author's knowledge. Hence, to attenuate the computational difficulties more effectively, it is proposed to restrict the attention to the local games mentioned earlier and consider the region of parameter values cropped by bounding the empirical probability distribution of the outcomes of the local games, rather than of the network formation game. Indeed, "thinking locally" may entail a significant reduction in the number of moment inequalities to deal with, thanks to the fewer mass points of the bounded probability distribution. However, at the same time, ignoring "the whole picture" may cause a loss of information about players' preferences for links, hence leading the researcher to conduct inference on a set of parameter values larger than, and containing, the sharp identified set (outer set). After having derived some sufficient conditions under which focusing on the local games preserves sharpness, it is noticed that, despite the local approach notably diminishes the quantity of moment inequalities to handle, chasing sharpness remains unmanageable when $N \geq 10$. It is then suggested to give up on sharpness and use a specific computationally convenient sub-collection of the original list of moment inequalities involving the outcomes of the local games.

When estimating the characterised outer set, one gets computational gains from two sources, under general assumptions. Firstly, the number of moment inequalities to exploit is substantially shortened: from $2^{2^{N(N-1)}} - 2$ to $2N \times 2^{N-1}$ moment inequalities for each value of parameters and exogenous observables. Secondly, checking the violation in the data of those moment inequalities is easy. Specifically, when obtaining by simulation the bounding terms, one can avoid verifying whether each of all possible 2^{N-1} outcomes of the local games is an equilibrium for every drawn value of preference shocks, an extremely demanding routine even for moderate N. Indeed, by applying Tarski's fixed point theorem and reinterpreting a result from Berry (1992) on the number of entrants in an entry game with substitution effects, the amount of outcomes of the local games which can strive for being equilibria is remarkably reduced. Overall, Monte Carlo exercises show that conducting inference on the proposed outer set is computationally feasible using relatively limited computational resources with up to 20 players. Lastly, some advantages in terms of computational tractability and width, over the outer set that one would obtain by applying Tamer (2003) and CT are provided.

Literature review The chapter is related to the empirical literature on entry games with multiple equilibria and no restrictions on equilibrium selection (Tamer, 2003; CT; BMM; de Paula, 2013; Aradillas-Lopez and Rosen, 2016). However, the techniques laid out by those works can not be directly applied to the setting considered here due to the huge number of possible networks even for small N. Thus, the present study aims to contribute by developing an identification methodology that allows to conduct inference on players' preferences for links using relatively limited computational resources up to $N = 20$, under general assumptions. As two side contributions, the chapter also offers some refinements in the characterisation of the sharp identified set with respect to BMM, and a proof for the existence of an equilibrium in a static game of network formation where links represent asymmetric relations between agents, never provided before in the literature. Some insights could be further extended to many-to-many two-sided matching models with externalities and non-transferable utilities.

Additionally, the chapter is related to the literature on the econometrics of economic models of network formation with interdependent link decisions (Currarini, Jackson and Pin, 2009; Christakis, et al., 2010; Badev, 2014; Leung, 2015; 2016; Boucher, 2016; de Paula, Richards-Shubik and Tamer, 2016; Menzel, 2016; Miyauchi, 2016; Sheng, 2016; Mele, 2017). Sheng (2016) proposes a close analysis, but designed for a static game of network formation where links represent symmetric relations between agents, payoffs are transferable or, if payoffs are non-transferable, then the game is assumed super-modular[39] in order to achieve the desired computational advantages. The present work considers instead a setting where links describe asymmetric relations between players, payoffs are non-transferable, and neither super-modularity nor sub-modularity is imposed.

The rest of the chapter is organised as follows: Section 2.2 illustrates the model; Section 2.3 discusses identification; Section 2.4 describes how to conduct inference; Section 2.5 reports some Monte Carlo experiments; Section 2.6 provides conclusions and directions for future research. All the tools from network theory used throughout the chapter are defined in Section 1.2. In terms of notation: bold case letters denote matrices; capital letters represent random variables/vector/matrices, small case letters indicate their realisations; the symbol $|\cdot|$

[39]I.e., the parameters governing the impact on a player's payoff of other players' actions are restricted to positive sign.

stands for the cardinality of a set; given a set $\mathcal{R}$, $\mathcal{K}_{\mathcal{R}}$ indicates the family of its non-empty compact subsets.

2.2 A network formation game

There are $N \in [3,\infty)$ players, labelled with the integers in $\mathcal{N} := \{1,2,...,N\}$, simultaneously deciding who to link with. A link between players i and j represents an asymmetric relation, i.e., i and j's payoffs from linking have different functional forms. PHBIs are an example of asymmetric relations between organizations. Indeed, if firm i has one of its executives sitting on the board of rival j with a non-executive role, i has the right to know about j's decision making process, but the converse is not true. Viceversa, i has advice duties towards j. Other examples of asymmetric relations are trading connections (buyer vs seller), and advice ties (advisor vs advisee).

The output of the game can be displayed as a directed network of size N with matrix $\boldsymbol{G}$.

Strategies Up to Section 2.3.8, it is assumed that the formation and deletion of the link ij requires the consent of player i only, as for advice ties, e.g., individuals nominating who they would ask for advice on the adoption of a technology. Section 2.3.8 then explains how the econometric methodology can be extended to events in which the consent of both players i and j is necessary for the formation of the link ij while deletion can be done unilaterally and payoffs are non-transferable, as for PHBIs and trading connections.

Unilateralism of the network formation process implies that players decide on outgoing links only. Hence, for any $i \in \mathcal{N}$, a pure strategy vector of player i is $G_{i\cdot} \in \{0,1\}^{N-1}$ collecting G_{ij} $\forall j \neq i \in \mathcal{N}$. A pure strategy profile of the game is $\boldsymbol{G} \in \mathcal{G} := \{0,1\}^{N(N-1)}$[40].

Preferences and information Players have complete information on preferences for links. This restriction, together with simultaneity of actions, is based on the idea that observed networks are realisations of a long-run equilibrium. Complete in-

[40] $|\mathcal{G}| = 2^{N(N-1)}$.

formation in a network formation model is imposed also e.g., by Badev (2014), de Paula, Richards-Shubik and Tamer (2016), Boucher (2016), Menzel (2016), Miyauchi (2016), Sheng (2016), and Mele (2017).

Players' preferences for links depend on agents' characteristics. Moreover, link decisions are interdependent, as the payoff that player i receives from linking to player j is affected by the number of additional players doing the same.

Specifically, for any $i \in \mathcal{N}$, let X_i denote a $K \times 1$ vector of player i's observed (to the researcher) characteristics. For any $i \in \mathcal{N}, j \in \mathcal{N}$ with $i \neq j$, let ε_{ij} be a scalar collecting the residual variables affecting the payoff that player i receives from the formation of the link ij which are unobserved by the researcher (also called preference shock). Lastly, let $\boldsymbol{X}$ be an $N \times K$ matrix listing X_i $\forall i \in \mathcal{N}$, and ε be an $N(N-1) \times 1$ vector collecting ε_{ij} $\forall i \in \mathcal{N}, \forall j \in \mathcal{N}$ with $i \neq j$. Each player $i \in \mathcal{N}$ gets as payoff

$$U_i(\boldsymbol{G}, \boldsymbol{X}, \varepsilon; \theta_u) := \sum_{j=1}^{N} G_{ij} \times \left[z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{N} G_{kj}; \delta) + \varepsilon_{ij} \right] \tag{2.1}$$

where $z(\cdot;\beta)$ is any function of X_i and X_j known by the researcher up to a vector of parameters β, $v(\cdot;\delta)$ is any function (weakly)[41] monotone in $\sum_{k \neq i}^{N} G_{kj}$ and known by the researcher up to a vector of parameters δ, and $\theta_u := (\beta, \delta) \in \Theta_u \subseteq \mathbb{R}^{d_\beta + d_\delta}$, with d_β and d_δ denoting the dimensions of β and δ.

Remark 1. (Discussion on the payoff function) $U_i(\cdot;\theta_u)$ is additively separable over player i's outgoing links. The same assumption is in Leung (2015), Badev (2014), and Mele (2017), and it helps to make the econometric analysis tractable. Moreover, as the network formation process is unilateral, the net benefit that player i possibly receives from her incoming links is ignored.

The payoff that player i gets from forming the link ij, $z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{N} G_{kj}; \delta) + \varepsilon_{ij}$, is assumed to depend on the number of additional agents connecting to player j, through the function $v(\sum_{k \neq i}^{N} G_{kj}; \delta)$. Such interdependence among link decisions captures a sort of "volume effect" and arises in several set-

[41] Monotonicity is weakly intended throughout the chapter.

tings of economic interest. For example, when examining PHBIs, the number of additional rivals having an executive appointed on the board of competitor j may encourage company i to join j's board too for exploiting "cheap talk" opportunities with them. Likewise, in the analysis of trading connections, the number of additional agents buying from family j may negatively affect agent i's power to bargain with j. Similarly, when considering advice ties, e.g., individuals nominating who they would ask for advice on the adoption of a technology, the number of additional people designating person j as adviser might proxy, in the eyes of individual i, j's time availability to offer proper explanations, or, if agents benefit from coordinating their subsequent advised choices, j's recommendations sharing level.

Moreover, the function $v(\cdot;\delta)$ is assumed to be monotone. The direction of the monotonicity is left unrestricted because economic theory provides no clear guidance on it in many empirical applications. The monotonicity requirement is used below to show existence of an equilibrium and reduce the computational burden of inference.

The results of the chapter also hold if one specifies the payoff that player i gets from forming the link ij as a function of the label-specific components $z_{ij}(\cdot;\beta_{ij})$ and $v_j(\cdot;\delta_j)$, provided that the direction of the monotonicity of the terms $\{v_j(\cdot;\delta_j)\}_{\forall j\in\mathcal{N}}$ is restricted to the same sign across j. However, it seems natural to proceed with functions and parameters independent of players' labels, as the data generating process (hereafter DGP) illustrated by Assumption 1 below postulates that players' identities can vary across networks and, hence, labels are assigned arbitrarily, as often is the case with data on networks, e.g., those used for the empirical illustration in Chapter 3.

Lastly, (2.1) pretends that, when player i decides about the link ij, she does not care about the identity, characteristics[42] or links of agent k with a link pointing to player j. Despite these effects might be attractive and reasonable in some empirical

[42]However, notice that if one is willing to assume that $v(\sum_{k\neq i}^{N} G_{kj};\delta)$ is monotone increasing, then the identification results of the chapter also hold if $v(\sum_{k\neq i}^{N} G_{kj};\delta)$ is replaced by $v(\sum_{k\neq i}^{N} G_{kj},(X_k)_{\forall k\neq i\in\mathcal{N}};\delta)$ provided that the random matrix $\boldsymbol{X}$ has positive support.

settings[43], the technical complexity of the problem urges to be less ambitious and leave them for future research. Indeed, adding them would break some of the results discussed below - specifically, existence of an equilibrium and computational burden reduction[44]. $\triangle$

Equilibrium Agents play PSNE. A network $\boldsymbol{G}$ is a PSNE if it is robust to multilink deviations by each player. More formally, let $\boldsymbol{G}_{-\{i\cdot\}}$ be the matrix $\boldsymbol{G}$ with ith row deleted.

Definition 1. (PSNE of the network formation game) $\boldsymbol{G} \in \mathcal{G}$ is a PSNE of the network formation game if

$$U_i(G_{i\cdot}, \boldsymbol{G}_{-\{i\cdot\}}, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\tilde{G}_{i\cdot}, \boldsymbol{G}_{-\{i\cdot\}}, \boldsymbol{X}, \varepsilon; \theta_u)$$

$\forall \tilde{G}_{i\cdot} \neq G_{i\cdot} \in \{0,1\}^{N-1}$ and $\forall i \in \mathcal{N}$. $\star$

2.3 Identification

2.3.1 Overview

This section examines identification of the parameters governing players' preferences for links, using data on $\boldsymbol{G}$ and $\boldsymbol{X}$[45]. From a first glance to (2.1), one may attempt to apply standard techniques developed for multivariate binary choice mod-

[43] For example, when examining PHBIs, firm i could prefer to have an executive sitting on the board of competitor j when executives of leaders in the industry have joined too, as "cheap talk" opportunities with them would be extremely valuable. Similarly, firm i could prefer to have an executive sitting on the board of competitor j when it hosts executives of rivals who are appointed, in turn, on several additional boards, because they might bring precious information about other companies operating in the sector. These effects are not considered by the present analysis.

[44] Sheng (2016) allows the payoff that player i gets from forming the link ij to depend also on the number of common ties with j, hence capturing a sort of "transitivity effect". Two remarks follow. Firstly, the analysis here is complicated by the fact that links represent asymmetric relations between players and, hence, introducing such additional component would break some of the results discussed below - specifically, existence of an equilibrium and computational burden reduction (unless one is willing to impose super-modularity, which, however, may be inappropriate in several contexts). Secondly, while the mentioned "transitivity effect" seems relevant in the formation of friendship networks (individuals with friends in common are more likely to become friends), its importance in other empirical settings, e.g., PHBIs, trading connections, advice ties, is more questionable.

[45] It is assumed that observed networks correspond to equilibria, as often imposed in the empirical literature on networks.

els. Indeed, by exploiting the additive separability of $U_i(\cdot;\theta_u)$ over player i's outgoing links, it can be shown that the inequalities in Definition 1 simplify to a system of $N(N-1)$ equations whose solution is a PSNE of the network formation game.

Lemma 1. (Characterisation of a PSNE of the network formation game) $\boldsymbol{G} \in \mathcal{G}$ is a PSNE of the network formation game if and only if

$$G_{ij} = \mathbb{1}\{z(X_i, X_j;\beta) + v(\sum_{k\neq i}^{N} G_{kj};\delta) + \varepsilon_{ij} \geq 0\} \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j \qquad (2.2)$$

$\diamond$

Unfortunately, interdependence among link decisions induced by the function $v(\cdot;\delta)$ precludes using identification results for multivariate binary choice models, as $\sum_{k\neq i}^{N} G_{kj}$ may be correlated with ε_{ij}. Moreover, it makes necessary investigating whether (2.2) has at least one solution (existence of a PSNE) or more than one solution (multiplicity of PSNE). (2.2) belongs instead to the class of simultaneous equation models, whose identification analysis starts with the seminal work by Heckman (1978) who provides results for a two-equation case. The discussion there is based on deriving the reduced form parameters so as to prove that the structural ones are identified. Mimicking that approach for (2.2) seems very hard, due to the complicated way in which the $N(N-1)$ equations are related. Therefore, an alternative identification procedure is developed in what follows.

The study opens with the description of the main assumption used throughout the section (Assumption 1). Next, in order to show existence of an equilibrium (Proposition 1), the network formation game is decomposed into some local games, which are similar for structure to entry games, and are such that the network formation game has an equilibrium if and only if each local game has an equilibrium (Lemma 2). In turn, existence of an equilibrium in each local game is proved by combining Tarski's fixed point theorem with the constructive proof that Berry (1992) designs to verify existence of an equilibrium in an entry game with substitution effects (Lemma 3).

It is then observed that the network formation game admits multiple equilibria. Thus, assuming that the researcher observes a large sample of equilibrium networks, partial identification arguments for the parameters of the model are developed under no restrictions on equilibrium selection, as seen in the most recent empirical literature on entry games (Tamer, 2003; CT; BMM; Aradillas-Lopez and Rosen, 2016)[46]. After having represented the region of all admissible parameter values (sharp identified set) by bounding the empirical probability distribution of the entire network as in BMM, it is noticed that the sharp identified set is characterised by a prohibitively enormous quantity of moment inequalities, indexed by every non-empty compact subset of the support $\mathcal{G}$ of $\boldsymbol{G}$, that makes inference on it impractical: $2^{2^{N(N-1)}}-2$ moment inequalities, for each value of parameters and exogenous observables. For example, with four players, one would need to check the violation in the data of $2^{4096}-2$ moment inequalities for each value of parameters and exogenous observables, which is a number greater than the quantity of atoms in the observed universe.

Even if some moment inequalities can be shown to be redundant by exploiting Lemma 2 (Proposition 2), constructing an algorithm to check the violation in the data of the relevant ones only - or even count them - seems unworkable to the best of the author's knowledge. Hence, to attenuate the computational difficulties more effectively, it is proposed to restrict the attention to the local games mentioned earlier and consider the region of parameter values cropped by bounding the empirical probability distribution of the outcomes of the local games, rather than of the network formation game. Indeed, "thinking locally" may entail a significant reduction in the number of moment inequalities to deal with, thanks to the fewer mass points of the bounded probability distribution. However, at the same time, ignoring "the whole picture" may cause a loss of information about players' preferences for links, hence leading the researcher to conduct inference on a set of parameter values larger than, and containing, the sharp identified set (outer set). After having derived

[46]Two comments should be made. Firstly, there may be sufficient conditions ensuring point identification even without restrictions on equilibrium selection; for example, one may attempt to extend to the case at hand the point identification results provided by Tamer (2003) for a two-player entry game, which, however, seems very hard for a generic N. Secondly, point identification may fail also if one rules out multiplicity of equilibria, depending on the functional forms assigned to $z(\cdot;\beta)$ and $\nu(\cdot;\delta)$.

some sufficient conditions (Assumption 2) under which focusing on the local games preserves sharpness (Proposition 3), it is noticed that, despite the local approach notably diminishes the number of moment inequalities to handle, chasing sharpness remains unmanageable when $N \geq 10$. It is then suggested to give up on sharpness and use a sub-collection of the original list of moment inequalities involving the outcomes of the local games. Specifically, it is proposed to conduct inference on the outer set of parameter values such that the empirical probability of each outcome of the local games is between the probability of such an outcome being the unique equilibrium of the local games, and the probability of such an outcome being a possible equilibrium of the local games, conditional on $\boldsymbol{X}$.

When estimating the characterised outer set, one gets computational gains from two sources, under Assumption 1. Firstly, the number of moment inequalities to exploit is substantially shortened: from $2^{2^{N(N-1)}} - 2$ to $2N \times 2^{N-1}$ moment inequalities for each value of parameters and exogenous observables. Secondly, checking the violation in the data of those moment inequalities is easy. In particular, when obtaining by simulation the bounding terms, one can avoid verifying whether each of all possible 2^{N-1} outcomes of the local games is an equilibrium for every drawn value of preference shocks, an extremely demanding routine even for moderate N. Indeed, by applying Tarski's fixed point theorem and reinterpreting a result from Berry (1992) on the number of entrants in an entry game with substitution effects, the amount of outcomes of the local games which can strive for being equilibria is notably reduced. Overall, Monte Carlo exercises show that conducting inference on the proposed outer set is computationally feasible using relatively limited computational resources up to $N = 20$. Lastly, Proposition 4 highlights some advantages, in terms of computational tractability and width, over the outer set that one would obtain by applying Tamer (2003) and CT.

2.3.2 Main assumption

Assumption 1. (Data generating process)

(i) The data generating process (hereafter DGP) is as follows: an integer N is drawn from $\mathbb{N} \setminus \{1,2\}$. N agents are selected from a population and labelled

$1,2,...,N$. Agents are endowed with characteristics collected in $\boldsymbol{X}_N$ and ε_N, where the subscript N highlights the dependence of matrix and vector sizes on N[47]. Agents play the network formation game described in Section 2.2 and a PSNE $\boldsymbol{G}_N$ arises.

The procedure is repeated M times and a sample of observations

$$\{n_m, \boldsymbol{x}_{n_m}, \boldsymbol{g}_{n_m}\}_{m=1}^{M}$$

is collected. The sampling scheme is designed such that the probability distribution of $\boldsymbol{G}_N$ conditional on $N, \boldsymbol{X}_N$ (hereafter empirical probability distribution of $\boldsymbol{G}_N$ conditional on $N, \boldsymbol{X}_N$) is identified[48].

In order to simplify the exposition and without loss of generality, in the remaining of Section 2.3 it is assumed that N is a degenerate random variable with support $\{n\}$, for $n \in \mathbb{N} \setminus \{1,2\}$. Moreover, the subscript N is deleted from $\boldsymbol{G}_N, \boldsymbol{X}_N, \varepsilon_N$ to clean up the notation.

(ii) ε is continuously distributed on $\mathbb{R}^{n(n-1)}$, independently of $\boldsymbol{X}$, with cdf denoted by $F(\cdot;\theta_\varepsilon)$ and known up to the vector of parameters $\theta_\varepsilon \in \Theta_\varepsilon \subseteq \mathbb{R}^{d_\varepsilon}$.

(iii) There exists $\theta_0 := (\theta'_{u,0}\ \theta'_{\varepsilon,0})' \in \Theta := (\Theta_u \cup \Theta_\varepsilon)$ generating the empirical probability distribution of $\boldsymbol{G}$ conditional on $\boldsymbol{X}$. Moreover, $\Theta := (\Theta_u \cup \Theta_\varepsilon) \ni \theta := (\theta'_u\ \theta'_\varepsilon)'$ is compact.

(iv) All random variables are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

•

Remark 2. (Discussion on Assumption 1) As highlighted by Assumption 1 (i), all the identification arguments developed in Section 2.3 are based on the observation of several equilibrium networks, as e.g., in Miyauchi (2016) and Sheng (2016). This

[47] Agents' characteristics are not required to be i.i.d.

[48] "*All that is needed is for the law of large numbers to hold.*" (CT, p.1799). Instead, Epstein, Kaido and Seo (2016) discuss inference without restricting the DGP.

requirement is fairly stringent but necessary, together with a sampling scheme compatible with a law of large number, to identify the empirical probability distribution of $\boldsymbol{G}$ conditional on $\boldsymbol{X}$. Some papers in the literature remove such a condition and offer procedures based on the observations of one or few large networks (Leung, 2015; 2016; de Paula, Richards-Shubik and Tamer, 2016; Menzel, 2016), at the expense of imposing other restrictions to control for interdependence among link decisions.

In accordance with the data used for the empirical illustration in Chapter 3, Assumption 1 (i) allows players' identities to vary across networks and, hence, labels are assigned arbitrarily. It follows that payoffs do not depend on players' labels and, consequently, neither do equilibrium sets. However, all the results in Section 2.3 also hold if players are the same across networks and, thus, labels are fixed.

Assumption 1 (ii) admits correlated preference shocks and makes the model fully parametric with the exception of the equilibrium selection mechanism. By imposing independence between ε and $\boldsymbol{X}$, it also rules out any source of unobserved heterogeneity correlated with $\boldsymbol{X}$. All the results in Section 2.3 also hold if one works with the probability distribution of ε conditional on $\boldsymbol{X}$, provided that it belongs to a known parametric family. Some papers in the literature remove such a requirement and include fixed effects in agents' payoffs (Dzemski, 2014; Candelaria, 2016; Graham, 2016; Jochmans and Weidner, 2016; Jochmans, 2017), at the expense of deleting interdependence among link decisions. Identification of parameters in a static game of network formation with interdependent link decisions and fixed effects remains an open question in the literature.

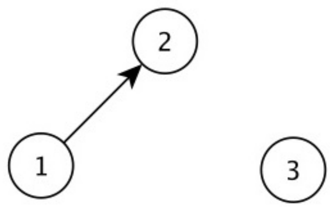
Lastly, Assumption 1 (iii) imposes that the model is correctly specified and that the parameter space is compact, and it is standard in the partial identification literature. In particular, the correct specification of the model can be tested following e.g., Andrews and Soares (2010), Andrews and Shi (2013), and Bugni, Cany and Shi (2015).

$\triangle$

2.3.3 Existence of an equilibrium

This section establishes equilibrium existence for every value of payoff-relevant variables and parameters as follows. For any $j \in \mathcal{N}$, the *section j* of a network is defined as the network portion collecting all nodes and the links pointing to node j. Figure 2.1 reports, as an example, the *section* 2 of the network in Figure 1.1. The network formation game is then decomposed into the n local games underlying the formation of the *section* 1, *section* 2, ..., *section n* (respectively called *section* 1 *game*, *section* 2 *game*,..., *section n game*). These local games are such that the network formation game has a PSNE if and only if each local game has a PSNE (Lemma 2). Existence of a PSNE in each local game is verified (Lemma 3). Hence, the network formation game has at least one PSNE (Proposition 1).

Figure 2.1: The *section 2* of the network in Figure 1.1.



More formally, for any $j \in \mathcal{N}$, in the *section j game* players other than player j simultaneously decide whether they want to link to j. For any $i \neq j \in \mathcal{N}$, a pure strategy of player i is $G_{ij} \in \{0,1\}$. A pure strategy profile of the game is $G_{\cdot j} \in \{0,1\}^{n-1}$ collecting G_{ij} $\forall i \neq j \in \mathcal{N}$[49]. Each player $i \neq j \in \mathcal{N}$ gets as payoff

$$U_i^j(G_{\cdot j}, \boldsymbol{X}, \varepsilon_{\cdot j}; \theta_u) := G_{ij} \times \left[z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon_{ij} \right] \tag{2.3}$$

where $\varepsilon_{\cdot j}$ is an $(n-1) \times 1$ vector listing ε_{ij} $\forall i \neq j \in \mathcal{N}$. Agents play PSNE.

Definition 2. (PSNE of the *section j game*) $G_{\cdot j} \in \{0,1\}^{n-1}$ is a PSNE of the *section j game* if

$$G_{ij} = \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon_{ij} \geq 0\} \quad \forall i \neq j \in \mathcal{N} \tag{2.4}$$

[49] $|\{0,1\}^{N-1}| = 2^{N-1}$.

$\star$

It can be seen that, for any $j \in \mathcal{N}$, players' payoffs within the *section j game* depend exclusively on $G_{\cdot j}$[50]. Moreover, by gathering $G_{\cdot 1}, G_{\cdot 2}, ..., G_{\cdot n}$, one uniquely obtains $\boldsymbol{G} \in \mathcal{G}$. Combining these facts,

Lemma 2. (Decomposing the network formation game) $\boldsymbol{G}$ is a PSNE of the network formation game if and only if $G_{\cdot j}$ is a PSNE of the *section j game* $\forall j \in \mathcal{N}$. $\diamond$

Lemma 2 is important beyond the existence arguments discussed here. Indeed, it helps to characterise further the set of PSNE of the network formation game, by revealing that it is the Cartesian product of the set of PSNE of the local games considered. Such a property will be exploited when defining the sharp identified set in Section 2.3.5.

Proceeding now with the existence proof, existence of a PSNE of the *section j game* when $v(\cdot;\delta)$ is monotone increasing is guaranteed by Tarski's fixed point theorem. Moreover, it can be noticed that the structure of the *section j game* when $v(\cdot;\delta)$ is monotone decreasing is similar to the structure of an entry game with substitution effects[51]. Existence of a PSNE in an entry game with substitution effects is proved in Berry (1992) by means of a constructive proof which can be easily reinterpreted for the *section j game*[52]. Merging these results,

Lemma 3. (Existence of a PSNE of the *section j game*) There exists a PSNE of the *section j game* $\forall j \in \mathcal{N}$. $\diamond$

Hence, combining Lemmas 2 and 3,

Proposition 1. (Existence of a PSNE of the network formation game) There exists a PSNE of the network formation game. $\diamond$

---

[50]This means that $\forall j \in \mathcal{N}$ players' payoffs within the *section j game* are not affected by players' choices outside the *section j game*. Nevertheless, the outcome of the *section j game* can be correlated with the outcome of the *section h game* through players' characteristics and equilibrium selection mechanisms, for any $h \in \mathcal{N}, j \in \mathcal{N}$, with $h \neq j$.

[51]Just by replacing player j with the entry market.

[52]More details are in the proof of Lemma 3 in Appendix B.

Remark 3. (Observation on existence of a PSNE) Existence of a PSNE of the network formation game does not require payoffs to display additive separability of preference shocks or a parametric specification. Indeed, it holds under the more general utility function

$$\tilde{U}_i(\boldsymbol{G},\boldsymbol{X},\varepsilon) := \sum_{j=1}^{N} G_{ij} \times \Big[r(X_i,X_j,\varepsilon_{ij}) + t(\sum_{k\neq i}^{n} G_{kj}) \Big] \tag{2.5}$$

where $r(\cdot)$ is any function of X_i, X_j and ε_{ij}, and $t(\cdot)$ is any function monotone in $\sum_{k\neq i}^{n} G_{kj}$. $\triangle$

2.3.4 Multiplicity of equilibria

By running simulations, it can be seen that the network formation game admits multiple PSNE for some values of payoff-relevant variables and parameters. This means that values of observed and unobserved exogenous variables do not uniquely pin down the value of endogenous variables, or, equivalently, there is no unique mapping from parameters and observed and unobserved exogenous variables, to endogenous variables. Moreover, simulations reveal that the equilibrium set may contain outcomes with a diametrically opposite economic meaning, such as the empty network and the fully connected network.

As the equilibrium selection mechanism is unobserved by the researcher and economic theory provides no guidance regarding its form, any assumption on it may be inappropriate and could bias estimates. Therefore, the econometric analysis proceeds by leaving the equilibrium selection mechanism totally unrestricted, as seen in the most recent empirical literature on entry games (Tamer, 2003; CT; BMM; Aradillas-Lopez and Rosen, 2016).

Without restrictions on equilibrium selection, deriving sufficient conditions ensuring point of identification of θ_0 becomes hard. Therefore, the study discusses partial identification of θ_0, i.e. it allows for the possibility that there may be more than one parameter value able to generate the empirical probability distribution of observables under the model's assumptions.

Before entering in the core of identification, few considerations are made. In principle one may want to achieve point identification of θ_0 with the help of direct or indirect assumptions on equilibrium selection[53]. For instance, one might restrict the function $v(\cdot;\delta)$ to be monotone decreasing. As explained in Section 2.3.3, the structure of the *section j game* when $v(\cdot;\delta)$ is monotone decreasing is similar to the structure of an entry game with substitution effects. Berry (1992) shows that in an entry game with substitution effects all the equilibria are characterised by the same number of firms entering the market. Reinterpreting this result for the *section j game*, it can be proved that all the PSNE of the *section j game* are characterised by the same number of players linking to player j. Consequently, θ_0 may be point identified by considering that number as the equilibrium outcome of interest $\forall j \in \mathcal{N}$, as in Berry (1992). Viceversa, one might restrict the function $v(\cdot;\delta)$ to be monotone increasing, which would guarantee existence of a greatest and a lowest PSNE by Tarski's fixed point theorem. In turn, it could be possible to derive point identification arguments after having postulated, e.g., that agents always play the greatest PSNE (Boucher, 2016). However, all these assumptions may bias estimates unless the researcher has a strong prior supporting them.

In the same spirit, another possibility might be imposing that the outcome observed by the researcher is chosen by players at random from the equilibrium set (Bjorn and Vuong, 1984; Kooreman, 1994). However, such a strategy would hardly be justifiable within this framework and may produce biased empirical results.

Alternatively, one could re-design the network formation game as a sequential model, where, at each iteration of a meeting protocol, a pair of players is drawn at random and determines the formation, maintenance or dissolution of a link (Christakis, et al., 2010; Badev, 2014; Mele, 2017). However, a potentially unattractive feature of this approach is that the realised sequence of meetings (in absence of noise in the meeting process) is contained in the set of equilibria predicted by the underlying static game, acting as an indirect restriction on the equilibrium selection mechanism that may bias estimates.

[53]However, as discussed in footnote 46, assumptions on equilibrium selection may not be sufficient to guarantee point identification of θ_0, which depends also on the functional forms assigned to $z(\cdot;\beta)$ and $v(\cdot;\delta)$.

A fourth option could be assigning a parametric form to the equilibrium selection mechanism, as in the entry game of Bajari, Hong and Ryan (2010). However, when applied to this framework, such a strategy could bias empirical results because economic theory provides no guidance on which parametric form to choose.

Given the inappropriateness of those four approaches, remaining agnostic as to equilibrium selection offers an alternative. The present work adopts this last strategy and discusses partial identification of θ_0 in what follows.

2.3.5 The sharp identified set under Assumption 1

In the language of BMM and Chesher and Rosen (2012), the set of parameter values generating the empirical probability distribution of observables under the model's assumptions is denominated the sharp identified set and indicated by $\Theta^\star$. Following Berry and Tamer (2006), $\Theta^\star$ can be equivalently defined as the set of parameter values for which one can find an equilibrium selection mechanism that, combined with the model's assumptions, delivers the empirical joint probability distribution of observables. However, this representation of $\Theta^\star$ does not facilitate inference because it involves the equilibrium selection mechanism which, as totally unrestricted, is a function representing an infinite dimensional nuisance parameter. BMM offer a powerful alternative by showing that the sharp identified set in the class of models with convex moment predictions can be expressed as the set of parameter values satisfying a collection of inequalities that do not contain the equilibrium selection mechanism. Hence, after observing that under Assumption 1 this model belongs to the class of models analysed by BMM, $\Theta^\star$ is characterised below adopting their approach[54]. Additionally, by exploiting Lemma 2 according to which the set of PSNE of the network formation game is the Cartesian product of the set of PSNE of the *section* 1 *game*,..., *section n game*, Proposition 2 shows that some of the inequalities determining $\Theta^\star$ are redundant in the present setting.

More formally, let $\mathcal{A}_\mathcal{G} \subset \mathcal{K}_\mathcal{G}$ be the collection of non-empty compact subsets of $\mathcal{G}$ obtained by taking the Cartesian product of all the possible ordered n-tuples

[54]Following BMM, $\Theta^\star$ is characterised by using random sets defined in the space of observables. One could also proceed with random sets defined in the space of unobservables (Chesher and Rosen, 2012).

with repetition from $\mathcal{K}_{\{0,1\}^{n-1}}$[55]. Given any $\theta \in \Theta$, consider the random closed set $\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon):\Omega \to \mathcal{A}_{\mathcal{G}}$ of PSNE of the network formation game. Notice that $\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)$ takes values in $\mathcal{A}_{\mathcal{G}}$ by Lemma 2. In order to characterise $\Theta^\star$, it is possible to apply Theorem D.2 in BMM, whose sufficient conditions are entirely satisfied by Assumption 1. Specifically,

$$\Theta^\star = \left\{ \theta \in \Theta | \mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \le T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K)\ \forall K \in \mathcal{K}_{\mathcal{G}}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \right\} \quad (2.6)$$

where $\mathcal{X}$ is the support of $\boldsymbol{X}$, and $T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}} : \mathcal{K}_{\mathcal{G}} \to [0,1]$ is the capacity functional of $\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)$ conditional on $\boldsymbol{x}$, prescribed by $T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K) := \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon) \cap K \neq \emptyset | \boldsymbol{X} = \boldsymbol{x})$ for any $K \in \mathcal{K}_{\mathcal{G}}$[56]. Each inequality in (2.6) is known as Artstein's inequality, for a total $2^{2^{n(n-1)}} - 2$ Artstein's inequalities $\forall \theta \in \Theta$ and $\forall \boldsymbol{x} \in \mathcal{X}$[57].

It is worth recognising that, despite $\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)$ takes values in $\mathcal{A}_{\mathcal{G}}$, Artstein's inequalities in (2.6) should be checked $\forall K \in \mathcal{K}_{\mathcal{G}}$, hence including sets that are not part of the support of $\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)$. By exploring further this point, it can be shown that some Artstein's inequalities in (2.6) are redundant.

Proposition 2. (Redundant Artstein's inequalities) Consider a set $K \in \mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1,...,|\mathcal{G}|-2\}$, where the set $C := \mathcal{G} \setminus K$ is such that $\exists$ a non-empty set $D \subset C$ with $\{\tilde{D} \cup \tilde{C}\} \notin \mathcal{A}_{\mathcal{G}}\ \forall \tilde{D} \subseteq D$ and $\forall \tilde{C} \subseteq \{C \setminus D\}$. Then, for any $\theta \in \Theta$

$$\mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \le T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K)$$

is implied by

$$\begin{cases} \mathbb{P}(\boldsymbol{G} \in \{K \cup D\} | \boldsymbol{X} = \boldsymbol{x}) \le T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(\{K \cup D\}) \\ \mathbb{P}(\boldsymbol{G} \in \{K \cup \{C \setminus D\}\} | \boldsymbol{X} = \boldsymbol{x}) \le T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(\{K \cup \{C \setminus D\}\}) \end{cases}$$

[55] $|\mathcal{A}_{\mathcal{G}}| = (2^{2^{n-1}} - 1)^n < |\mathcal{K}_{\mathcal{G}}| = 2^{2^{n(n-1)}} - 1$. Appendix B explains how to construct $\mathcal{A}_{\mathcal{G}}$.

[56] The same characterisation can be derived by applying some results in Galichon and Henry (2011).

[57] Despite the cardinality of $\mathcal{K}_{\mathcal{G}}$ is $2^{2^{n(n-1)}} - 1$, the number of Artstein's inequalities to consider is $2^{2^{n(n-1)}} - 2$ because Artstein's inequality arising when $K = \mathcal{G}$ is trivial.

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. $\diamond$

Additionally, a necessary and sufficient condition for a set $K \in \mathcal{K}_{\mathcal{G}}$ to satisfy the sufficient conditions of Proposition 2 is provided by Corollary 1.

Corollary 1. (Necessary and sufficient condition for Proposition 2) Given a set $K \in \mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1,...,|\mathcal{G}|-2\}$, the set $C := \mathcal{G} \setminus K$ is such that $\exists$ a non-empty set $D \subset C$ with $\{\tilde{D} \cup \tilde{C}\} \notin \mathcal{A}_{\mathcal{G}}$ $\forall \tilde{D} \subseteq D$ and $\forall \tilde{C} \subseteq \{C \setminus D\}$ if and only if all the pairs of matrices $\boldsymbol{g}_D \in D, \boldsymbol{g}_{\{C \setminus D\}} \in \{C \setminus D\}$ differ for at least two rows. $\diamond$

Example 1 helps to clarify Proposition 2 and Corollary 1.

Example 1. (Example on Proposition 2 and Corollary 1) Let $n = 3$ and

$$K := \mathcal{G} \setminus \{ \overbrace{\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}}^{\boldsymbol{g}_1}, \overbrace{\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}}^{\boldsymbol{g}_2}, \overbrace{\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}^{\boldsymbol{g}_3} \}$$

K satisfies the sufficient conditions of Proposition 2. Indeed, given $C := \{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$, it can be set $D := \{\boldsymbol{g}_3\}$ so that $\{\boldsymbol{g}_1\} \cup \{\boldsymbol{g}_3\} \notin \mathcal{A}_{\mathcal{G}}$, $\{\boldsymbol{g}_2\} \cup \{\boldsymbol{g}_3\} \notin \mathcal{A}_{\mathcal{G}}$, and $\{\boldsymbol{g}_1, \boldsymbol{g}_2\} \cup \{\boldsymbol{g}_3\} \notin \mathcal{A}_{\mathcal{G}}$. Furthermore, K satisfies the necessary and sufficient condition of Corollary 1. Indeed, given the choice of D above, $\boldsymbol{g}_1$ and $\boldsymbol{g}_3$ have all rows different, and $\boldsymbol{g}_2$ and $\boldsymbol{g}_3$ differ for the second and third rows. $*$

Unfortunately, providing the exact number of sets satisfying the necessary and sufficient condition of Corollary 1 for a generic n seems an open problem in combinatorics to the best of the author's knowledge. At most, one can bound such a number by making use of some graph theory results from Brouwer and Koolen (2009).

Corollary 2. (Bounds for Corollary 1) The number of sets $K \in \mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1,...,|\mathcal{G}|-2\}$ such that, given the set $C := \mathcal{G} \setminus K$, $\exists$ a non-empty set $D \subset C$ with all pairs of matrices $\boldsymbol{g}_D \in D, \boldsymbol{g}_{\{C \setminus D\}} \in \{C \setminus D\}$ differing for at least two rows is $a \in \{|\mathcal{G}|, |\mathcal{G}|+1, ..., \sum_{k=n(2^{n-1}-1)}^{|\mathcal{G}|-2} \binom{2^{n(n-1)}}{k}\}$. $\diamond$

2.3.6 The sharp identified set based on local games

Conducting inference on $\Theta^{\star}$ as characterised in (2.6) is prohibitively complex. Indeed, Artstein's inequalities determining $\Theta^{\star}$ are obtained by bounding the empirical probability that $\boldsymbol{G}$ takes value in a given compact subset of its support $\mathcal{G}$, for every possible non-empty compact subset of $\mathcal{G}$, hence, generating $2^{2^{n(n-1)}} - 2$ inequalities, for each value of parameters and exogenous observables. For example, with four players, one would need to check the violation in the data of $2^{4096} - 2$ inequalities for each value of parameters and exogenous observables, which is a number greater than the quantity of atoms in the observed universe. Moreover, even if Proposition 2 reveals that some Artstein's inequalities in (2.6) are redundant by exploiting Lemma 2, constructing an algorithm to check the violation in the data of the relevant ones only - or even count them - seems unworkable to the best of the author's knowledge.

A strategy to attenuate the computational difficulties more effectively is to restrict the attention to some local games of the network formation game, and constructing bounds for the empirical probability distribution of the outcomes of the local games, rather than of the network formation game. Indeed, "thinking locally" may entail a significant reduction in the number of inequalities to consider, thanks to the fewer mass points of the bounded probability distribution. However, at the same time, ignoring "the whole picture" may cause a loss of information about players' preferences for links, thus leading the researcher to conduct inference on an outer set of parameter values larger than and containing the sharp identified set.

In this spirit, the section proposes to focus on the *section* 1 *game*, *section* 2 *game*,..., *section n game*, described earlier. In particular, it is shown that, if one is willing to assume independence among these local games (Assumption 2), $\Theta^{\star}$ can be equivalently characterised through Artsein's inequalities obtained by bounding the empirical probability that $G_{\cdot j}$ takes value in a given compact subset of its support $\{0,1\}^{n-1}$, for every possible non-empty compact subset of $\{0,1\}^{n-1}$, $\forall j \in \mathcal{N}$, and for each value of parameters and exogenous observables (Proposition 3). This, in turn, implies that the number of Artstein's inequalities determining $\Theta^{\star}$ becomes $n(2^{2^{n-1}} - 2)$, with a notable reduction with respect to (2.6).

Arguments are articulated as follows. By construction of $\mathcal{A}_{\mathcal{G}}$, for any $K \in \mathcal{A}_{\mathcal{G}}$, there exists n sets, $K_{\cdot 1} \in \mathcal{K}_{\{0,1\}^{n-1}}$, ..., $K_{\cdot n} \in \mathcal{K}_{\{0,1\}^{n-1}}$, such that their Cartesian product delivers K. Hence, one may think that verifying Artstein's inequality for $\boldsymbol{G}$ indexed by K is equivalent to verifying Artstein's inequality for $G_{\cdot j}$ indexed by $K_{\cdot j}$ $\forall j \in \mathcal{N}$. It turns out that this is true under independence among the *section* 1 *game*, *section* 2 *game*,..., *section n game*. Indeed, combining this restriction with Lemma 2, Artstein's inequality for $\boldsymbol{G}$ indexed by K is equal to the product across $j \in \mathcal{N}$ of Artstein's inequality for $\boldsymbol{G}_{\cdot j}$ indexed by $K_{\cdot j}$. As all terms are between 0 and 1, if Artstein's inequality for $\boldsymbol{G}_{\cdot j}$ indexed by $K_{\cdot j}$ is satisfied $\forall j \in \mathcal{N}$, then, by taking the product across j, Artstein's inequality for $\boldsymbol{G}$ indexed K is verified too. Thus, $\Theta^\star$ can be characterised by fewer inequalities than (2.6) because the cardinality of $\mathcal{K}_{\{0,1\}^{n-1}}$ is smaller than the cardinality of $\mathcal{K}_{\mathcal{G}}$. A more formal discussion is now presented.

Consider the map $\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) : \Omega \to \mathcal{K}_{\{0,1\}^{n-1}}$ such that $\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}(\omega), \varepsilon_{\cdot j}(\omega))$ is the set of PSNE of the *section j game*, for any $\omega \in \Omega$. Following Proposition 3.1 in BMM, under Assumption 1, $\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j})$ is a random closed set almost surely non-empty. Additionally, the following assumption is introduced.

Assumption 2. (Independence)

(i) The vectors forming the sequence $(\varepsilon_{\cdot j})_{\forall j \in \mathcal{N}}$ are independently distributed across j.

(ii) The equilibrium selection mechanisms adopted by players in the *section* 1 *game*, *section* 2 *game*,..., *section n game* are independent of each other, i.e.

$$\mathbb{P}(G_{\cdot 1} \in K_{\cdot 1}, ..., G_{\cdot n} \in K_{\cdot n} | \mathcal{S}_{\theta_u,\cdot 1}(\boldsymbol{X}, \varepsilon_{\cdot 1}), ..., \mathcal{S}_{\theta_u,\cdot n}(\boldsymbol{X}, \varepsilon_{\cdot n})) = \Pi_{j=1}^n \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}))$$

$\forall K_{\cdot 1} \in \mathcal{K}_{\{0,1\}^{n-1}}, ..., \forall K_{\cdot n} \in \mathcal{K}_{\{0,1\}^{n-1}}$, and $\forall \theta \in \Theta$.

•

Remark 4. (Discussion on Assumption 2) Assumption 2 imposes independence among the *section* 1 *game*, *section* 2 *game*,..., *section n game*, by restricting the correlation among preference shocks and the collection of admissible equilibrium selection mechanisms for the *section* 1 *game*, ..., *section n game*.

More specifically, $\forall i \in \mathcal{N}$, Assumption 2 (i) does not allow ε_{ij} to be correlated with ε_{ik} $\forall j \in \mathcal{N}, \forall k \in \mathcal{N}$ with $i \neq j \neq k$. Assumption 2 (i) is satisfied if, $\forall i \in \mathcal{N}$ and $\forall j \in \mathcal{N}$ with $i \neq j$, $\varepsilon_{ij} = \beta_j + \xi_{ij}$, where $\Big((\beta_j)_{\forall j \in \mathcal{N}}, (\xi_{ij})_{\forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j}\Big)$ are i.i.d. (random effects across j). Instead, Assumption 2 (i) is violated if, for some $i \in \mathcal{N}$ and for some $j \in \mathcal{N}$ with $i \neq j$, $\varepsilon_{ij} = \alpha_i + \beta_j + \xi_{ij}$, where $\Big((\alpha_i)_{\forall i \in \mathcal{N}}, (\beta_j)_{\forall j \in \mathcal{N}}, (\xi_{ij})_{\forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j}\Big)$ are i.i.d. (random effects across i and j).

$\forall j \in \mathcal{N}$, Assumption 2 (ii) excludes that, in case of multiple equilibria, players coordinate on a specific outcome in the equilibrium set of the *section j game* considering equilibrium selection rules adopted in the *section h game* for any $h \neq j \in \mathcal{N}$. For example, Assumption 2 (ii) is satisfied if, $\forall j \in \mathcal{N}$, the equilibrium selection rule of the *section j game* assigns a uniform probability distribution over the outcomes in the equilibrium set of the *section j game*. Assumption 2 (ii) is also met if, $\forall j \in \mathcal{N}$, players select the outcome providing the highest total payoff from the equilibrium set of the *section j game*. Instead, Assumption 2 (ii) is violated if e.g., players choose an outcome from the equilibrium set of the *section j game* with the purpose of enhancing heterogeneity across agents creating links within the whole network.

It can also be noticed that, by Lemma 2, $\forall K_{\cdot 1} \in \mathcal{K}_{\{0,1\}^{n-1}}, ..., \forall K_{\cdot n} \in \mathcal{K}_{\{0,1\}^{n-1}}$, and $\forall \theta \in \Theta$,

$$\mathbb{P}(G_{\cdot 1} \in K_{\cdot 1}, ..., G_{\cdot n} \in K_{\cdot n} | \mathcal{S}_{\theta_u, \cdot 1}(\boldsymbol{X}, \varepsilon_{\cdot 1}), ..., \mathcal{S}_{\theta_u, \cdot n}(\boldsymbol{X}, \varepsilon_{\cdot n})) = \mathbb{P}(\boldsymbol{G} \in K | \mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon))$$

where K is obtained by taking the Cartesian product of $K_{\cdot 1}, ..., K_{\cdot n}$. This reveals that Assumption 2 (ii) restricts also the collection of admissible equilibrium selection mechanisms for the whole network formation game.

Lastly, Assumptions 1 and 2 imply the testable prediction that $G_{\cdot 1}, ..., G_{\cdot n}$ are independent conditional on $\boldsymbol{X}$, which, is, in turn, the key result used to show Proposition 3 below. Indeed, $\forall K_{\cdot 1} \in \mathcal{K}_{\{0,1\}^{n-1}}, ..., \forall K_{\cdot n} \in \mathcal{K}_{\{0,1\}^{n-1}}$, and $\forall \theta \in \Theta$, under

Assumptions 1 and 2, the model states that

$$\prod_{j=1}^{n}\mathbb{P}(G_{\cdot j}\in K_{\cdot j}|\boldsymbol{X};\theta)=\prod_{j=1}^{n}\int_{e_{\cdot j}\in\mathbb{R}^{n-1}}\mathbb{P}(G_{\cdot j}\in K_{\cdot j}|\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},e_{\cdot j}))d\tilde{F}_{\cdot j}(e_{\cdot j};\theta_\varepsilon)$$

$$=\int_{e_{\cdot 1}\in\mathbb{R}^{n-1}}\int_{e_{\cdot 2}\in\mathbb{R}^{n-1}}\cdots\int_{e_{\cdot n}\in\mathbb{R}^{n-1}}\prod_{j=1}^{n}\mathbb{P}(G_{\cdot j}\in K_{\cdot j}|\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},e_{\cdot j}))d\tilde{F}_{\cdot j}(e_{\cdot j};\theta_\varepsilon)$$

$$=\int_{e_{\cdot 1}\in\mathbb{R}^{n-1}}\int_{e_{\cdot 2}\in\mathbb{R}^{n-1}}\cdots\int_{e_{\cdot n}\in\mathbb{R}^{n-1}}\prod_{j=1}^{n}\mathbb{P}(G_{\cdot j}\in K_{\cdot j}|\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},e_{\cdot j}))\prod_{j=1}^{n}d\tilde{F}_{\cdot j}(e_{\cdot j};\theta_\varepsilon)$$

$$\underbrace{=}_{\text{Ass. 2 (i)}}\int_{e:=(e_{\cdot 1},\dots,e_{\cdot n})\in\mathbb{R}^{n(n-1)}}\prod_{j=1}^{n}\mathbb{P}(G_{\cdot j}\in K_{\cdot j}|\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},e_{\cdot j}))dF(e;\theta_\varepsilon)$$

$$\underbrace{=}_{\text{Ass 2 (ii)}}\int_{e:=(e_{\cdot 1},\dots,e_{\cdot n})\in\mathbb{R}^{n(n-1)}}\mathbb{P}(G_{\cdot 1}\in K_{\cdot 1},\dots,G_{\cdot n}\in K_{\cdot n}|\mathcal{S}_{\theta_u,\cdot 1}(\boldsymbol{X},e_{\cdot 1}),\dots,\mathcal{S}_{\theta_u,\cdot n}(\boldsymbol{x},e_{\cdot n}))dF(e;\theta_\varepsilon)$$

$$=\mathbb{P}(G_{\cdot 1}\in K_{\cdot 1},\dots,G_{\cdot n}\in K_{\cdot n}|\boldsymbol{X};\theta)$$

$\triangle$

Define the set

$$\Theta^{\star\star}:=\left\{\theta\in\Theta|\mathbb{P}(G_{\cdot j}\in K_{\cdot j}|\boldsymbol{X}=\boldsymbol{x})\leq T_{\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(K_{\cdot j})\ \forall K_{\cdot j}\in\mathcal{K}_{\{0,1\}^{n-1}},\forall j\in\mathcal{N},\forall\boldsymbol{x}\in\mathcal{X}\text{ a.s.}\right\} \tag{2.7}$$

where $T_{\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}:\mathcal{K}_{\{0,1\}^{n-1}}\to[0,1]$ is the capacity functional of $\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})$ conditioned on $\boldsymbol{x}$, prescribed by $T_{\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(K_{\cdot j}):=\mathbb{P}(\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})\cap K_{\cdot j}\neq\emptyset|\boldsymbol{X}=\boldsymbol{x})$, for any $K_{\cdot j}\in\mathcal{K}_{\{0,1\}^{n-1}}$.

Proposition 3. (Sharp identified set under Assumptions 1, 2) (i) Under Assumption 1, $\Theta^{\star\star}\supseteq\Theta^{\star}$. (ii) Under Assumptions 1 and 2, $\Theta^{\star\star}=\Theta^{\star}$. $\diamond$

2.3.7 An outer set

Section 2.3.6 shows that, by considering Artstein's inequalities for the *section* 1 *game*, *section* 2 *game*,..., *section n game*, $\Theta^{\star}$ can be characterised by fewer inequalities. Even so, when there are 10 or more players, such a reduction is not enough and conducting inference on $\Theta^{\star}$ remains prohibitively complex. For example, with 10 players and imposing Assumption 2, one would need to verify the violation in the data of $10(2^{512}-2)$ inequalities $\forall\theta\in\Theta$ and $\forall\boldsymbol{x}\in\mathcal{X}$, which is a number still greater than the quantity of atoms in the observed universe.

As anticipated earlier, a way to attenuate further the computational difficulties is giving up on sharpness and selecting only some computationally convenient Artstein's inequalities from (2.7). Specifically, this section proposes to conduct inference on the outer set of parameter values such that the empirical probability of each realisation of $G_{\cdot j}$ is between the probability of such a realisation being the unique equilibrium of the *section j game*, and the probability of such a realisation being a possible equilibrium of the *section j game*, conditional on $\boldsymbol{X}$, $\forall j\in\mathcal{N}$.

More formally, consider Artstein's inequalities for $G_{\cdot j}$ indexed by the compact sets $\{g_{\cdot j}\}\in\mathcal{K}_{\{0,1\}^{n-1}}$ and $\{0,1\}^{n-1}\setminus\{g_{\cdot j}\}\in\mathcal{K}_{\{0,1\}^{n-1}}$ $\forall g_{\cdot j}\in\{0,1\}^{n-1}$ and $\forall j\in\mathcal{N}$. The suggested outer set is hence

$$\begin{aligned}\Theta^o := \Big\{\theta\in\Theta|\mathbb{P}(G_{\cdot j}\in\{0,1\}^{n-1}\setminus\{g_{\cdot j}\}|\boldsymbol{X}=\boldsymbol{x}) &\le T_{\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(\{0,1\}^{n-1}\setminus\{g_{\cdot j}\})\\ \mathbb{P}(G_{\cdot j}=g_{\cdot j}|\boldsymbol{X}=\boldsymbol{x}) &\le T_{\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(\{g_{\cdot j}\})\ \forall g_{\cdot j}\in\{0,1\}^{n-1},\forall j\in\mathcal{N},\forall\boldsymbol{x}\in\mathcal{X}\text{ a.s.}\Big\}\end{aligned} \tag{2.8}$$

It can be noticed that, under Assumption 1, $\Theta^o\supseteq\Theta^\star$. In fact, $\Theta^o\supseteq\Theta^{\star\star}$ by construction, and $\Theta^{\star\star}\supseteq\Theta^\star$ by Proposition 3.

Computational gains When conducting inference on a region of parameters (sharp or not sharp), one has to check the violation in the data of the inequalities defining the region for every possible parameter value. Computational difficulties come from the number of inequalities to consider - as highlighted by the chapter so far - and from the necessity of obtaining their sample analogues for every possible parameter value - because the bounding probabilities are often very complicated multi-dimension integrals.

In this respect, when conduction inference on Θ^o, one gets computational gains from two sources, under Assumption 1. Firstly, the number of inequalities to consider is notably diminished: from $2^{2^{n(n-1)}}-2$ to $2n\times 2^{n-1}$, with respect to $\Theta^\star$. Secondly, checking the violation in the data of those inequalities is easy. More precisely, it can be observed that, $\forall g_{\cdot j} \in \{0,1\}^{n-1}$ and $\forall j \in \mathcal{N}$,

$$\mathbb{P}(G_{\cdot j} \in \{0,1\}^{n-1} \setminus \{g_{\cdot j}\} | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(\{0,1\}^{n-1} \setminus \{g_{\cdot j}\})$$

is equivalent to

$$\mathbb{P}(G_{\cdot j} = g_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}) \geq \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j}) = \{g_{\cdot j}\}} d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon) \tag{2.9}$$

and

$$\mathbb{P}(G_{\cdot j} \in \{g_{\cdot j}\} | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(\{g_{\cdot j}\})$$

is equivalent to

$$\mathbb{P}(G_{\cdot j} = g_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}) \leq \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } g_{\cdot j} \in \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j})} d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon) \tag{2.10}$$

where $\tilde{F}_{\cdot j}(\cdot; \theta_\varepsilon)$ is $\varepsilon_{\cdot j}$'s cdf, the first multi-dimensional integral is the probability that $g_{\cdot j}$ is the unique PSNE of the *section j game*, and the second multi-dimensional integral is the probability that $g_{\cdot j}$ is a PSNE of the *section j game*. Hence, obtaining the sample analogues of those inequalities involves the computation of the multi-dimensional integrals in (2.9) and (2.10) which can be done via the simple frequency simulator proposed by McFadden (1989) and Pakes and Pollard (1989). In principle, one would need to draw several values of preference shocks and verify whether each of all possible 2^{n-1} realisations of $G_{\cdot j}$ is a PSNE of the *section j game* for every drawn value and $\forall j \in \mathcal{N}$, generating an extremely demanding routine even for moderate n. However, by exploiting some properties of the set of PSNE of the *section j game* for any $j \in \mathcal{N}$, the whole process can be significantly sped up. Specifically, when during the inference procedure a candidate parameter value is such that $v(\cdot; \delta)$ is monotone increasing, Tarski's fixed point theorem guarantees existence of a greatest and lowest fixed points. These two fixed points can be quickly obtained by implementing the algorithm in Jia (2008). It follows that one only has to check whether the realisations of $G_{\cdot j}$ lying between the greatest and lowest fixed points are PSNE of the *section j game*[58]. Viceversa, when during the inference procedure a candidate a parameter value is such that $v(\cdot;\delta)$ is monotone decreasing, the structure of the *section j game* is similar to the structure of an entry game with substitution effects, as explained in Section 2.3.3. Berry (1992) shows that in an entry game with substitution effects all the equilibria are characterised by the same number of firms entering the market. Reinterpreting this result for the *section j game*, it can be proved that all the PSNE of the *section j game* are characterised by the same number, $n^*_{\cdot j}$, of players linking to player j[59]. $n^*_{\cdot j}$ can be quickly obtained by implementing the constructive algorithm used in Section 2.3.3 to show existence of a PSNE of the *section j game* when $v(\cdot;\delta)$ is monotone decreasing[60]. Thus, one only has to check whether the realisations of $G_{\cdot j}$ characterised by $n^*_{\cdot j}$ players linking to player j are PSNE of the *section j game*, for a total of $\frac{(n-1)!}{n^*_{\cdot j}!(n-1-n^*_{\cdot j})!} < 2^{n-1}$ realisations.

Overall, Monte Carlo experiments reveal that conducting inference on Θ^o is computationally manageable with relatively limited computational resources up to $n = 20$. Though, it is important to highlight that the proposed methodology is not applicable to very large networks, as the number of inequalities to consider, even if remarkably diminished with respect to $\Theta^\star$, still depends on n[61].

Comparison with Tamer (2003) and CT Tamer (2003) and CT illustrate an entry game with complete information and focus on the outer set of parameter values such that the empirical probability of each realisation of the vector of firms' actions is between the probability of such a realisation being the unique equilibrium of the entry game and the probability of such a realisation being a possible equilibrium of the entry game, conditional on players' observed characteristics. Thus, as a direct application of such a strategy, one might characterise an outer set, Θ^o_{CT}, collecting

[58]Miyauchi (2016) restricts to positive externalities and uses the same intuition to simplify the computational burden of inference.

[59]See the end of the proof of Lemma 3 in Appendix B.

[60]See the proof of Lemma 3 in Appendix B.

[61]Moreover, the bounds in (2.8) vanish as $n \to \infty$. Hence, when n is very large, instead of the methodology described here, it may be more appropriate to follow procedures developed for situations in which one or few large networks are observed (Leung, 2015; 2016; de Paula, Richards-Shubik and Tamer, 2016; Menzel, 2016).

the parameter values such that the empirical probability of each realisation of the network is between the probability of such a realisation being the unique equilibrium of the network formation game and the probability of such a realisation being a possible equilibrium of the network formation game, conditional on $\boldsymbol{X}$.

More formally, consider Artstein's inequalities for $\boldsymbol{G}$ indexed by the compact sets $K := \{\boldsymbol{g}\} \in \mathcal{K}_{\mathcal{G}}$ and $K := \mathcal{G} \setminus \{\boldsymbol{g}\} \in \mathcal{K}_{\mathcal{G}}$ $\forall \boldsymbol{g} \in \mathcal{G}$. Let

$$\begin{aligned}\Theta^o_{CT} := \Big\{ \theta \in \Theta | \mathbb{P}(\boldsymbol{G} \in \mathcal{G} \setminus \{\boldsymbol{g}\} | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(\mathcal{G} \setminus \{\boldsymbol{g}\}) \\ \mathbb{P}(\boldsymbol{G} \in \{\boldsymbol{g}\} | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(\{\boldsymbol{g}\}) \; \forall \boldsymbol{g} \in \mathcal{G}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \Big\}\end{aligned} \tag{2.11}$$

As for Θ^o, the inequalities above can be rewritten by using the probability that $\boldsymbol{g}$ is the unique PSNE of the network formation game and the probability that $\boldsymbol{g}$ is a PSNE of the network formation game. Moreover, by (2.6), $\Theta^o_{CT} \supseteq \Theta^\star$.

However, computational gains generated by Θ^o_{CT} may be insufficient because conducting inference on Θ^o_{CT} requires checking the violation in the data of $2 \times 2^{n(n-1)}$ inequalities $\forall \theta \in \Theta$ and $\forall \boldsymbol{x} \in \mathcal{X}$. Instead, Θ^o brings greater computational advantages by considering $2n \times 2^{n-1}$ inequalities $\forall \theta \in \Theta$ and $\forall \boldsymbol{x} \in \mathcal{X}$. For example, with 15 players as in the data used for the empirical illustration in Chapter 3, Θ^o_{CT} is defined by 3.291×10^{63} inequalities, while Θ^o involves $491{,}520$ inequalities, $\forall \theta \in \Theta$ and $\forall \boldsymbol{x} \in \mathcal{X}$.

Additionally, in terms of informativeness of bounds, Θ^o_{CT} delivers wider bounds than Θ^o when independence among the *section* 1 *game*, *section* 2 *game*,..., *section n game* is imposed. Indeed, if $\theta \in \Theta^o$, then all the inequalities determining Θ^o_{CT} can be reconstructed by taking products of specific inequalities characterising Θ^o. Viceversa, by summing inequalities representing Θ^o_{CT}, one obtains bounds that are larger than those in (2.8). Specifically,

Proposition 4. (Comparison with Tamer (2003) and CT) Under Assumptions 1 and 2, $\Theta^o \subseteq \Theta^o_{CT}$. ◇

2.3.8 Extensions

All the results illustrated so far can be extended to situations in which the formation of the link ij requires the consent of both players i and j, while deletion can be done unilaterally, as for PHBIs and trading connections. More details on such a bilateral game follow.

Strategies Players reveal the desired outgoing and incoming links, and only reciprocally announced ties are formed. For any $i \in \mathcal{N}$, a pure strategy vector of player i is $s^i \in \{0,1\}^{2(N-1)}$ collecting s^i_{ij} and s^i_{ji} $\forall j \neq i \in \mathcal{N}$, where s^i_{ij} is a scalar equal to 1 if player i is willing to form the link ij and 0 otherwise. A pure strategy profile of the game is $s \in \{0,1\}^{2N(N-1)}$ listing s^i $\forall i \in \mathcal{N}$. Mutual consent is needed to form links, i.e., $G_{ij} = s^i_{ij} s^j_{ij}$ $\forall i \in \mathcal{N}, \forall j \in \mathcal{N}$ with $i \neq j$.

Preferences and information Assumptions on preferences and information are as in Section 2.2. However, now each player $i \in \mathcal{N}$ can decide also on incoming links according to the payoff

$$U_i(\boldsymbol{G},\boldsymbol{X},\varepsilon;\theta_u) := \sum_{j=1}^{N} G_{ji} \times \left[b(X_i,X_j;\gamma) + \varepsilon^i_{ji}\right] + \sum_{j=1}^{N} G_{ij} \times \left[z(X_i,X_j;\beta) + v(\sum_{k\neq i}^{N} G_{kj};\delta) + \varepsilon^i_{ij}\right] \tag{2.12}$$

where ε^i_{ji} and ε^i_{ij} are scalars, unobserved by the researcher, listing the residual variables affecting the payoff that player i receives, respectively, from the formation of the link ji, and of the link ij. The first term of the sum represents the net benefits that player i gets from her incoming connections, where $b(\cdot;\gamma)$ is any function of X_i and X_j known by the researcher up to a vector of parameters γ. The second term of the sum is as in (2.1). Payoffs are non-transferable.

Equilibrium Agents play PSPNE[62], and the resulting network is a pure strategy pairwise Nash stable (hereafter PSPNS) network. A network $\boldsymbol{G}$ is a PSPNS network when it is robust to unilateral multi-link deletion and bilateral one-link formation. More formally, let $\boldsymbol{G}$'s dependence on s be indicated by $\boldsymbol{G}(s)$. Additionally, let s^{-i} be the vector s without s^i.

[62] Jackson and Wolinski (1996); Calvó-Armengol (2004); Bloch and Jackson (2006); Goyal and Joshi (2006); Calvó-Armengol and Ilkiliç (2009).

Definition 3. (PSPNS network) s is a PSPNE of the bilateral network formation game if

$$U_i(\boldsymbol{G}(s), \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\boldsymbol{G}(s_\star^i, s^{-i}), \boldsymbol{X}, \varepsilon; \theta_u)$$

$\forall s_\star^i \neq s^i \in \{0,1\}^{2(N-1)}$ and $\forall i \in \mathcal{N}$, and there does not exist a pair of players $(i, j) \in \mathcal{N}$ such that, when $G_{ij}(s) = 0$,

$$U_i(\boldsymbol{G}(s) + ij, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\boldsymbol{G}(s), \boldsymbol{X}, \varepsilon; \theta_u)$$

and

$$U_j(\boldsymbol{G}(s) + ij, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_j(\boldsymbol{G}(s), \boldsymbol{X}, \varepsilon; \theta_u)$$

with strict inequality for at least one of the two players, where $\boldsymbol{G}(s) + ij$ denote the matrix $\boldsymbol{G}(s)$ when the link ij is added.
$\boldsymbol{G}$ is a PSPNS network if there exists a PSPNE s of the bilateral network formation game such that $\boldsymbol{G} = \boldsymbol{G}(s)$. ⋆

Remark 5. (Observations on the equilibrium concept) Alternative equilibrium concepts adopted in bilateral games are pairwise stability[63] and Nash stability[64]. Pairwise stable networks are robust to unilateral one-link deletion and bilateral one-link formation. Pairwise stability is an equilibrium notion independent of any network formation procedure and has nice computational properties. However, it only considers very simple deviations and, hence, it may be too tolerant in classifying a network as stable, especially when there are few players. On the other hand, Nash stable networks are constructed by letting players announce desired outgoing and incoming links, according to PSNE, and, then, forming mutually beneficial links. Using PSNE in a bilateral game induces coordination problems because link creation requires the consent of the two involved parties. This causes the game displaying a multiplicity of Nash stable networks, always including the empty network, as playing zero is weakly optimal even when forming a link would be profitable to both players. In order to solve this issue, PSPNE allows players to coordinate their deci-

[63] Jackson and Wolinski (1996).
[64] Myerson (1991).

sions, and, by not leaving aside any reciprocally beneficial link, it refines the set of stable networks. In particular, the set of PSPNS networks is the intersection of the set of Nash stable networks and the set of pairwise stable networks. Additionally, within this model, the set of PSPNS networks and the set of pairwise stable networks coincide, by the additively separability of $U_i(\cdot;\theta_u)$ over player i's incoming and outgoing links (Gilles and Sarangi, 2005).

Also, Definition 3 assumes that players' payoffs are non-transferable. Adapting the results of the chapter to the case of transferable payoffs, with transfers made between the players involved in a link or coming from outside players (Bloch and Jackson, 2005; 2006), is not a trivial extension and, therefore, is left to future analysis. $\triangle$

Lastly, by exploiting the additive separability of $U_i(\cdot;\theta_u)$ over player i's outgoing and incoming links, Lemma 4 maintains that the inequalities in Definition 3 simplify to a system of $N(N-1)$ equations whose solution is a PSPNS network.

Lemma 4. (Characterisation of a PSPNS network) $\boldsymbol{G}$ is a PSPNS network if and only if

$$G_{ij} = \mathbb{1}\{z(X_i,X_j;\beta)+v(\sum_{k\neq i}^{N} G_{kj};\delta)+\varepsilon_{ij}^{i}\geq 0\}\mathbb{1}\{b(X_i,X_j;\gamma)+\varepsilon_{ij}^{j}\geq 0\} \quad \forall i\in\mathcal{N}, j\in\mathcal{N}, i\neq j \tag{2.13}$$

$\diamond$

Identification Equilibrium existence for every value of payoff-relevant variables and parameters can be shown following the steps illustrated in Section 2.3.3, after having adapted the constructive proof in Berry (1992) to the bilateral setting considered here. More details are in Appendix A. Moreover, the game admits multiple equilibria for some values of payoff-relevant variables and parameters. Hence, without restrictions on equilibrium selection, partial identification arguments analogous to those discussed in sections 2.3.5 and 2.3.7 can be derived, just by replacing the equilibrium concept and imposing $\varepsilon_{ij} := (\varepsilon_{ij}^{i}, \varepsilon_{ij}^{j})$[65].

[65]Differently from the unilateral case discussed up to Section 2.3.7, it should be noticed that, in the bilateral situation considered here, the tightness of bounds depends also on the proportion of

2.4 Inference

After having transformed Artstein's inequalities in (2.8) into moment inequalities, inference on the outer set Θ^o can be performed choosing among several available techniques: if $\mathcal{X}$ is finite e.g., Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Romano and Shaikh (2008; 2010), Rosen (2008), Stoye (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano, Shaikh, and Wolf (2014), Pakes, et al. (2015), Bugni, et al. (2016), Chernozhukov, Chetverikov and Kato (2016), Kaido, Molinari and Stoye (2016), Bugni, Canay and Shi (2017); if $\mathcal{X}$ is not finite e.g., Andrews and Shi (2013, 2017), Chernozhukov, Lee, and Rosen (2013), Lee, Song, and Whang (2013, 2014), Amstrong (2014), Chetverikov (2017).

This section briefly discusses how to construct a $(1-\alpha)\%$ confidence region following the generalized moment selection procedure developed by Andrews and Soares (2010) and Andrews and Shi (2013).

Firstly, it is convenient to express Θ^o as

$$\begin{aligned}\Theta^o = \Big\{\theta \in \Theta | H^l_{g_{\cdot j},\boldsymbol{x},n}(\theta) \leq \mathbb{P}(G_{\cdot j} = g_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}, N = n) \leq H^u_{g_{\cdot j},\boldsymbol{x},n}(\theta) \\ \forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.}, \forall n \in \mathbb{N} \setminus \{1,2\}\Big\}\end{aligned} \tag{2.14}$$

where

$$H^l_{g_{\cdot j},\boldsymbol{x},n}(\theta) := \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } \mathcal{S}_{\cdot j,\theta_u}(\boldsymbol{x},n,e_{\cdot j}) = \{g_{\cdot j}\}} d\tilde{F}_{\cdot j}(e_{\cdot j};\theta_\varepsilon) \tag{2.15}$$

and

$$H^u_{g_{\cdot j},\boldsymbol{x},n}(\theta) := \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } g_{\cdot j} \in \mathcal{S}_{\cdot j,\theta_u}(\boldsymbol{x},n,e_{\cdot j})} d\tilde{F}_{\cdot j}(e_{\cdot j};\theta_\varepsilon) \tag{2.16}$$

It follows that

$$\begin{aligned}\Theta^o = \Big\{\theta \in \Theta | E\Big[\mathbb{1}(G_{\cdot j} = g_{\cdot j}) - H^l_{g_{\cdot j},\boldsymbol{X},N}(\theta) | \boldsymbol{X} = \boldsymbol{x}, N = n\Big] \geq 0, \\ E\Big[H^u_{g_{\cdot j},\boldsymbol{X},N}(\theta) - \mathbb{1}(G_{\cdot j} = g_{\cdot j}) | \boldsymbol{X} = \boldsymbol{x}, N = n\Big] \geq 0 \\ \forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.}, \forall n \in \mathbb{N} \setminus \{1,2\}\Big\}\end{aligned} \tag{2.17}$$

directed links observed in the sample.

or, equivalently by stacking all the integrants in a vector $m(\boldsymbol{G},\boldsymbol{X},N;\boldsymbol{\theta})$ of dimension $(2n2^{n-1}) \times 1$,

$$\Theta^o = \Big\{\theta \in \Theta | E\Big[m(\boldsymbol{G},\boldsymbol{X},N;\theta)|\boldsymbol{X} = \boldsymbol{x}, N = n\Big] \geq 0 \;\forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.}, \forall n \in \mathbb{N} \setminus \{1,2\}\Big\} \tag{2.18}$$

Secondly, the conditional moment inequalities in (2.18) should be transformed into unconditional moment inequalities by considering the set

$$\Theta^o(\mathcal{P}) = \Big\{\theta \in \Theta | E\Big[m(\boldsymbol{G},\boldsymbol{X},N;\theta)p(\boldsymbol{X},N)\Big] \geq 0 \;\forall p \in \mathcal{P}\Big\} \tag{2.19}$$

where $p \in \mathcal{P}$ is a function $(\boldsymbol{x},n) \in \mathcal{X} \times \mathbb{N} \setminus \{1,2\} \mapsto p(\boldsymbol{x},n) \in \mathbb{R}^{2n \times 2^{n-1}}$, and $\mathcal{P}$ is chosen such that $\Theta^o = \Theta^o(\mathcal{P})$.

For example, when $\mathcal{X}$ is finite,

$$\mathcal{P} := \{p \text{ s.t. } p(\boldsymbol{x},n) = \mathbb{1}\{\boldsymbol{X} = \boldsymbol{x}, N = n\} \times \mathbf{1}_{2n2^{n-1}} \;\forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.}, \forall n \in \mathbb{N} \setminus \{1,2\}\} \tag{2.20}$$

where $\mathbf{1}_{2n2^{n-1}}$ denotes the vector of ones with dimension $(2n2^{n-1}) \times 1$. When $\mathcal{X}$ is not finite, details on the construction of $\mathcal{P}$ are given by Andrews and Shi (2013), section 9.

The researcher is now ready to construct an appropriate test statistic $S_M(\theta)$ $\forall \theta \in \Theta$ just by replacing the expectation in (2.19) with its sample analogue and imposing a penalty for each inequality violated in the data. A $(1-\alpha)\%$ confidence region for each $\theta \in \Theta^o$ is hence

$$CS_M := \Big\{\theta \in \Theta — S_M(\theta) \leq \hat{c}_{M,1-\alpha}(\theta)\Big\} \tag{2.21}$$

where $\hat{c}_{M,1-\alpha}(\theta)$ is an estimate of the $1-\alpha$ quantile of the asymptotic probability distribution of $S_M(\theta)$. More details on how to compute $S_M(\theta)$ and $\hat{c}_{M,1-\alpha}(\theta)$ are in Andrews and Soares (2010), Andrews and Shi (2013), and Appendix C.

2.5 Monte Carlo simulations

This section reports the results of some Monte Carlo experiments on the outer set Θ^o run assuming that $\mathcal{X}$ is finite and following the inference method developed by Andrews and Soares (2010), whose main steps are illustrated in Appendix C. The focus is on the unilateral case with the following model specification

$$U_i(\boldsymbol{G},\boldsymbol{X},\varepsilon;\theta_u) := \sum_{j=1}^{N} G_{ij} \times \left[\beta|X_i - X_j| + \delta \sum_{k\neq i} G_{kj} + \varepsilon_{ij}\right] \tag{2.22}$$

where $X_i \sim U([0,1])$, and $\{\varepsilon_{ij}\}_{\forall i,j\in\mathcal{N},i\neq j}$ are i.i.d. across ij with ε_{ij} distributed as a standard normal. Let $\theta := (\beta,\delta)$.

Firstly, the behaviour of $\frac{1}{M}S_M(\theta_0)$ is investigated for different values of N, M, and θ_0. in Figure 2.2. Panel (a) is obtained by setting $N = 3$, $M = 200,500,800$, and $\theta_0 = (0.4,-0.3)$. Panel (b) is obtained by setting $N = 7$, $M = 200,500,800$, and $\theta_0 = (-1.5,1.2)$. Panel (c) is obtained by setting $N = 20$, $M = 200,500,800$, and $\theta_0 = (-5,-6)$. For all the panels, the number of simulations to compute the multidimensional integrals discussed in Section 2.4 is imposed equal to $\frac{M}{2}$. As expected, the empirical probability distribution function of $\frac{1}{M}S_M(\theta_0)$ shrinks around zero as M increases. Regarding the computational performance using 12 cores: when $N = 3$ and $M = 200,500,700$, the average time per iteration is, respectively, $0.005,0.020,0.056$ sec.; when $N = 7$ and $M = 200,500,800$, the average time per iteration is, respectively, $0.015,0.102,0.266$ sec.; when $N = 20$ and $M = 200,500,800$, the average time per iteration is, respectively, $0.613,4.413,35.033$ sec.

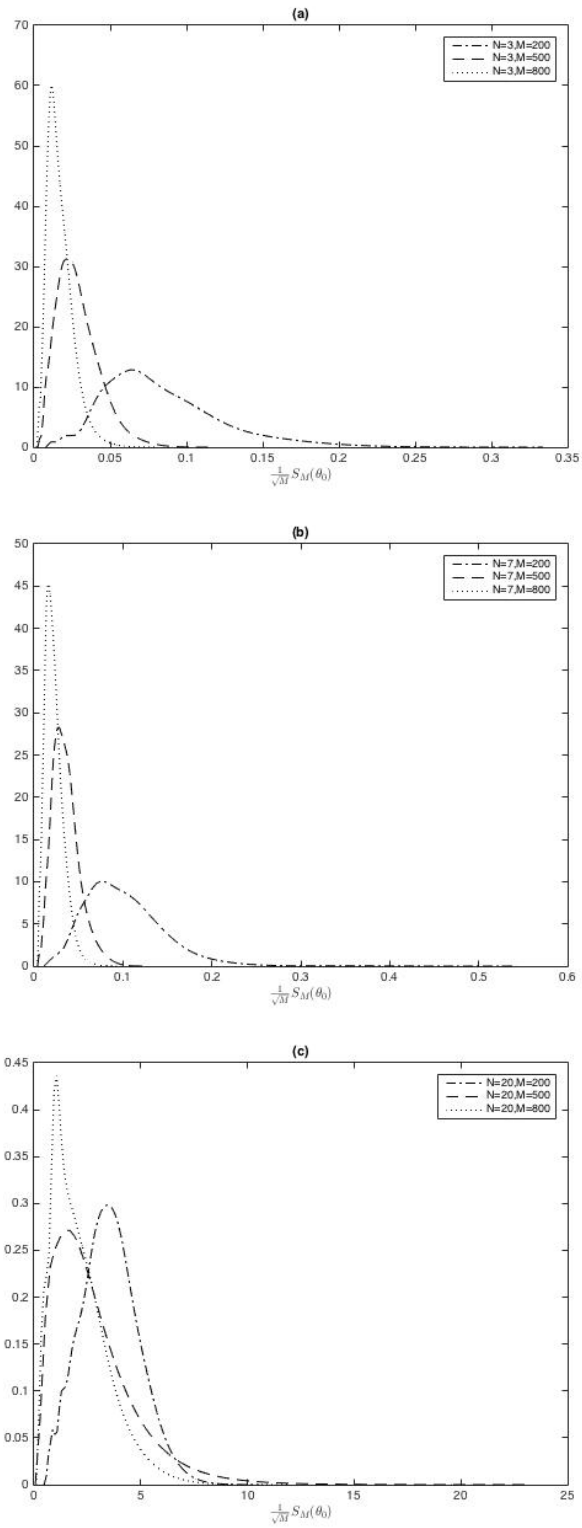


Figure 2.2: The figure reports the estimated probability distribution function of $\frac{1}{M}S_M(\theta_0)$ for different values of N, M, and θ_0. Panel (a) is obtained by setting $N = 3$, $M = 200, 500, 800$, and $\theta_0 = (0.4, -0.3)$. Panel (b) is obtained by setting $N = 7$, $M = 200, 500, 800$, and $\theta_0 = (-1.5, 1.2)$. Panel (c) is obtained by setting $N = 20$, $M = 200, 500, 800$, and $\theta_0 = (-5, -6)$. For all the panels, the number of simulations to compute the multidimensional integrals discussed in Section 2.4 is imposed equal to $\frac{M}{2}$.

Secondly, the coverage probability of θ_0 by the 95% confidence region, constructed as discussed in Section 2.4, is examined. Specifically, Table 2.1 reports the fraction of Monte Carlo experiments such that θ_0 belongs to the 95% confidence region over 500 replications, for different values of N, M, and θ_0. As expected, such a fraction is equal to or greater than 0.95. The number of simulations to compute the multidimensional integrals and the number of bootstrapped samples to obtain the critical values are set respectively equal to 50 and 100.

Table 2.1: Fraction of Monte Carlo experiments such that θ_0 belongs to the 95% confidence region over 500 replications.

	$M = 100$	M=400
$N = 3, \theta_0 = (0.4, -0.3)$	1	1
$N = 7, \theta_0 = (-1.5, 1.2)$	0.96	0.98
$N = 20, \theta_0 = (-5, -6)$	0.95	0.95

2.6 Conclusions

The chapter studies identification of the parameters governing agents' preferences in a static game of network formation, where links represent asymmetric relations between players. Agents have complete information and play PSNE if link formation can be unilaterally established, or PSPNE in the bilateral case. Payoffs are non-transferable. Link decisions are interdependent, as the payoff that player i receives from linking to player j is affected by the number of additional players doing the same. After having shown existence of an equilibrium and assuming that several equilibrium networks are observed, partial identification arguments are provided without restrictions on equilibrium selection in the presence of multiple equilibria. The identification methodology attenuates the usual computational difficulties arising at the inference stage - due to the large number of possible sets of equilibria - by giving up on sharpness and restricting the attention to some local games of the network formation game. Overall, Monte Carlo exercises show that constructing a confidence region for the suggested identified region of parameters is computationally manageable using relatively limited computational resources, with up to 20 players. As an empirical illustration of the methodology, the chapter investigates firms' incentives for having executives sitting on the board of competitors, using data on Italian joint stock companies. It is found that firm i prefers its executives sitting on the board of rival j when executives of other competitors are hosted too, possibly because it enables i to engage with them in "cheap talking" about past or future choices, besides having the opportunity to learn about j's decision making process.

There are some avenues of future research. Specifically, there could be other interdependencies among link decisions to consider. For example, player i's payoff from linking to player j may also depend on the i's connections, or on the identity, links and characteristics of the additional agents connecting to j. It may be worth enriching players' payoffs in this direction and investigating how the identification results proposed here can be extended to such more complicated settings. Another idea could be to examine how the identification analysis changes if one removes the monotonicity of $v(\cdot;\delta)$ - e.g., one may wonder how to adjust bounds when $v(\cdot;\delta)$ has a "U" shape -, or the additive separability over outgoing and incoming links characterising payoffs.

Chapter 3

An empirical application to board interlocks[66]

3.1 Introduction

The chapter shows that that the methodology developed in Chapter 3 can deliver economically meaningful estimates. Specifically, the procedure is used to investigate firms' incentives for having executives sitting on the board of competitors (also called primary horizontal board interlocks, hereafter PHBIs).

Most organisations are governed by a board of directors composed of executives and non-executives. The former lead the decision making process, the latter are involved in the supervision and advising of executives. PHBIs are a common arrangement of firms' governance structure in several European countries. Deeply analysed by corporate governance experts, they also draw the attention of economists because they may help firms to exchange information, and, in turn, reduce strategic uncertainty, transmit tacit knowledge, increase transparency, or encourage coordination. In such a scenario, interdependence among companies' decisions for forming PHBIs becomes crucial because it allows them to expand and radiate the flow of information. Indeed, firm i could find extremely attractive to have one of its executives sitting on the board of rival j when executives of other competitors are hosted too, as it would enable i to engage with them in "cheap talk-

[66]The views expressed in this chapter are those of the author and do not necessarily reflect those of the Bank of Italy.

ing" about past of future choices, besides having the opportunity to learn about j's decision making process.

At the same time, such interdependence causes endogeneity, and, thus, prevents the possibility of using standard econometric techniques to conduct inference on firms' preferences behind PHBIs. Applying the methodology illustrated in Chapter 2 represents an alternative. In particular, a 95% confidence region for the suggested outer set is constructed using Italian data. In line with the intuition above, results reveal that firms prefer to have their executives sitting on the board of a rival when executives of other competitors are appointed too. For the aim of simplification, firms' characteristics are discretised and the inference method developed by Andrews and Soares (2010) is followed, as illustrated in Appendix C[67].

Literature review The exercise falls within the study of inter-organisational ties, e.g., board interlocks, cross-ownerships, joint ventures, or supply and distribution channels. More specifically, the corporate governance literature on the topic is divided into two groups. According to the inter-organizational linkage perspective, embraced by this chapter, companies are entities that possess interests. In pursuit of them, they form relations with other firms. For example, they share board members as an attempt to transfer information, and, consequently, decrease investment uncertainty, anticipate disturbances, promote coordination, or convey expertise (Thompson and McEwen, 1958; Dooley, 1969; Allen, 1974; Pfeffer and Salancik, 1978; Aldrich, 1979), especially when happening with competitors (Carrington, 1981; Leslie, 2004; Gabrielsen, Hjelmeng, and Sørgard, 2011; Waller, 2011)[68] and through the exchange of executives (Mintz and Schwartz, 1981; Mizruchi and Bunting, 1981; Stokman, Wasseur and Elsas, 1985; Stokman, Van Der Knoop and Wasseur, 1988; Mizruchi and Stearns, 1994). Conversely, according to the class alliance view, directors are actors with career and reputation goals. To achieve

[67] One could also proceed without discretising firms' characteristics and apply the inference procedure developed by Andrews and Shi (2013; 2017). However, when N is not small, these methodologies become computationally burdensome as the dimension of the conditioning variable is equal to $3N$.

[68] Also some policy reports, e.g., OECD (2008; 2010), debate on information exchanges between competitors through board interlocks.

them, they establish connections with other peers, for example by sitting on multiple boards (Zeitlin, 1976; Zajac, 1988)[69].

Most of the empirical findings on board interlocks focus on the correlation between a firm's profitability or size, and the intensity of board interlocks (Dooley, 1969; Pfeffer, 1972; Allen, 1974; Bunting, 1976; Pennings, 1980; Carrington, 1981; Burt, 1983; Meeusen and Cuyvers, 1985; Mizruchi and Stearns, 1988; Kaplan and Reishus, 1990; Booth and Deli, 1996). Some empirical works analyse the impact of board interlocks (exogenously taken or instrumented) on firms' internal decisions, e.g., executives' compensations (Hallock, 1997; Core, Holthausen, and Larcker, 1999; Patnam, 2013; Gayle, Golan, and Miller, 2015), patenting and R&D spending (Helmers, Patnam and Rau, 2015), and hiring choices (Lalanne, 2016). The chapter aims to shade further light on the role of board interlocks as information transmitters, by estimating, for the first time, a model in which firms' decisions on board interlocks are interdependent as a result of organisations' incentives for taking advantage of information streams.

The rest of the chapter is organised as follows: Section 3.2 summarises the italian context; Section 3.3 illustrates the specification of the game for PHBIs; Section 3.4 describes the data; Section 3.5 reports the results; Section 3.6 provides conclusions and directions for future research.

3.2 The Italian context

The firms considered by the empirical illustration are Italian joint stock companies (*Societá per Azioni*). Joint stock companies, i.e, business entities where shareholders' liability is limited to the nominal value of held shares, represent the largest organisations in Italy. They are not necessarily listed on the stock exchange and are governed by a board of directors, which is a collegial body appointed by the shareholders who are free to choose its size. The board of directors can delegate its executive duties to one or more of its members. If the mandate is conferred, then executives have to report to the board with a frequency determined by the company's

[69] Detailed reviews on the two approaches are e.g., in Palmer (1983), Ornstein (1984), and Mizruchi (1996).

statute, and, in any case, at least every six months. Moreover, delegators can ask executives to provide the board with any information related to the management of the company and have advice duties over executives' conduct on the basis of the information received[70].

The sharing of board members between joint stock companies is a distinguish feature of Italian capitalism since the end of the nineteenth century (Luzzatto Fegiz, 1928; Bianco and Pagnoni, 1997; Barbi, 2000; Rinaldi and Vasta, 2005; Bertoni and Randone, 2006; Ciocca, 2007; Santella, et al., 2009), with relatively stable dynamics over time (Vasta and Baccini, 1997; Rinaldi and Vasta, 2005; 2012; Santella, Drago and Polo, 2009; Bellenzier and Grassi, 2014). Additionally, the Italian law system does not impose any clear and general prohibition on such a practice, or on the number of different appointments that a director can hold. The only exception is the Law 214 of the year 2011, which forbids organisations to share of board members with companies or groups operating in the banking, insurance and financial services sectors.

3.3 Applying the network formation game to board interlocks

There are $N \in [3, \infty)$ firms operating in the same industry and simultaneously deciding which links to form according to the rules of the game described in Section 2.3.8. The link ij exists when firm i has one of its executives sitting on the board of firm j, and it represents an asymmetric exchange of information between i and j. Indeed, i has the right to know about j's decision making process, because, by the Italian corporate law, j should truthfully report to i of any past or future action, during board meetings held on a regular basis. Moreover, when executives of other competitors are appointed on the board of j too, i may engage with them in "cheap talking", i.e., costless, informal, and unverifiable discussions, about prior or planned choices[71]. Viceversa, i has advice duties towards j. Thus, j receives

[70] Articles 2381, 2392, and 2393 of the Italian Civil Code.
[71] Farrell and Rabin (1996).

suggestions and recommendations from i, potentially including "cheap talk" confidences. Such asymmetry is reflected in the structure of the payoff function (2.12) and is highlighted also e.g., by Warner and Unwalla (1967), Mizruchi and Bunting (1981), Palmer (1983), and Richardson (1987).

Remark 6. (Observations on the payoff function)

Suppose the links ij, kj, and ik are formed. It turns out that firm k's "cheap talk" announcements during firm j's board meetings are verifiable by firm i, through the attendance of k's board meetings. Still, (2.12) assumes that i takes into account the opportunity for seeing k at j's assembly. The intuition behind is that i sustains costs in order to check k's "cheap talk" announcements made during j's board meetings, e.g., reserving time for k's board meetings. Hence, any occasion for unofficial conversations with k outside k's assembly is valuable to i.

Suppose now the links ij, kj, ih, kh are formed. (2.12) pretends that firms i and h take into account twice the possibility of seeing each other. The intuition behind is that both opportunities are equally relevant to the eyes of i and h because they may give rise to "cheap talk" discussions different for content or quality.

Also, as mentioned in Remark 1, (2.12) postulates that, when firm i decides about the link ij, it does not care about identity, characteristics and existing links of firm k with an executive sitting on company j's board. Despite all these effects might be attractive and reasonable[72], the technical complexity of the problem urges to leave them for future research. Indeed, they would all break the proof for the existence of a PSPNS network and nullify the computational gains brought by the outer set of parameter values Θ^o.

Moreover, the Italian law does not impose any restriction on board size. This, combined with the fact that the industry size in the data used is at most 15, seems to suggest that there is no need to add cost functions or constraints depending on the number of links in firms' payoffs.

Lastly, it seems reasonable to assume that firms' payoffs are non-transferable, as the net benefits that companies receive from information flows spreading through

[72] See footnote 43 for examples.

links are hardly comparable[73]. △

As an example, Figure 3.1 pictures a network arising in an industry composed by four firms. $\forall i \in \{1, ..., 4\}$, the board composition of firm i is indicated by two sets of letters. Each letter denotes an individual. The first set of letters is the set of executives. The second set of letters is the set of non-executives.

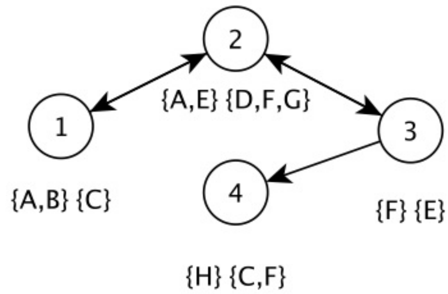


Figure 3.1: Example of a directed network arising in an industry composed by four firms. For every $i \in \{1, ..., 4\}$, the board composition of firm i is indicated by two sets of letters. Each letter denotes an individual. The first set of letters is the set of executives. The second set of letters is the set of non-executives.

Remark 7. (Observations on the construction of networks) Three considerations follow. Firstly, directors' identities are ignored. In fact, according to the Italian legal framework, firms are the ultimate arbiters of link decisions, as a director needs her original board's approval to join the board of a competitor[74]. Nevertheless, there may be contexts in which a link decision between two competitors is motivated by the outstanding and exogenous capacities of an individual, rather than for transferring information. These situations are excluded by the present work to preserve the tractability of the analysis[75].

Secondly, and related to the first point, executives are identified with their companies. For example, consider director F in Figure 3.1. Despite she sits on the boards of firms 2 and 4, she does not create any link between them because she acts

[73]In reality, there are money transfers from firms to directors. However, as explained in Remark 7, the present analysis ignores directors' identities to preserve tractability.

[74]Article 2390 of the Italian Civil Code.

[75]I.e., it is assumed that any link is motivated by the will of transferring information between firms, independently of the characteristics of the shared directors.

on both boards as a representative of firm 3 where she has executive duties. On top of that, every circumstances in which companies share directors lacking executive roles at all the companies in a given industry are omitted, as, most of the times, these individuals sit on multiple boards because of their technical skills rather than in order to transfer information across firms. In support of this argument, several works in the corporate governance literature sustain that only ties involving executive powers can represent long term economic and institutional relations between companies (Mizruchi and Bunting, 1981; Stokman, Wasseur and Elsas, 1985; Stokman, Van Der Knoop and Wasseur, 1988). Moreover, consider director A in Figure 3.1 with executive duties in firms 1 and 2. The analysis here pretends that, during firm 2's board meetings, A participates to "cheap talk" discussions with director F speaking in the name of firm 1[76].

Lastly, cases in which firms share board members with companies in other industries are not taken into account. In fact, before the year 2011 (after it is forbidden by law), these events are mostly driven by connections with financial institutions, possibly arising because experts in the financial industry are useful to firms operating in other industries, or because lending banks want to be represented on and control the boards of debtors. As these incentives could follow patterns different from the information exchange arguments illustrated earlier, their study may require a payoff structure more sophisticated than (2.12) and is postponed to future research. Additionally, the majority of the policy debate today is focused on the incentives behind the existence of ties within industries, rather than across industries, given the potential impact on competition.

$\triangle$

3.4 Data

The sources of data are the Registro Imprese and the Cerved databases, whose access has been provided by the Bank of Italy. The Registro Imprese is a database in which all Italian companies are required to enrol through the Chamber of Commerce

[76]Notice also that, during firm 2's board meetings, director A cannot participate to "cheap talk" discussions with director F speaking in the name of company 2.

in their territorial province and is the primary source of certification of their constituent data. It offers detailed and updated information on individual firms (e.g., legal status, year of registration, composition of governance bodies, geographical location, principle line of activity) and on important changes related to their existence (e.g., termination, liquidation, bankruptcy, mergers and acquisitions). The Cerved database contains information useful for measuring the credit risk of Italian limited companies[77], and, among other data, provides balance sheet details.

The considered sample collects all the Italian joint stock companies with a governance organised under the Articles 2380/2409-septies of the Italian Civil Code (*Sociatá per Azioni con sistema tradizionale*) and whose data for the year 2010 were available[78], i.e., 2599 firms operating in 386 industries.

The board composition, together with the role of each director (executive or non-executive), are extracted from the Registro Imprese database. Industries are constructed considering firms' principal lines of activity provided by the ATECO 2002 code from the Registro Imprese database. The ATECO 2002 code is similar to the SIC code in the UK and U.S. It is an alpha-numeric code with varying degrees of detail - the letters indicate the macroeconomic sector while the numbers represent subsectors. It is developed in five levels: sections (letter), subsections (two letters, optional), divisions (2 digits), groups (3 digits), classes (4 digits) and categories (5

[77] Joint stock companies are a category of limited companies.

[78] The Italian law system offers several ways to organise the governance of a joint stock company. The *sistema tradizionale* is the default rule, adopted by the majority of joint stock companies. Identifying from the Registro Imprese database joint stock companies with a governance organised differently from the *sistema tradizionale* is a delicate task, that requires careful investigation of the power relations among governance bodies and is beyond the scope of the present work.

digits)[79,80].

In line with some empirical studies on board interlocks, e.g., Dooley (1969), Pfeffer (1972), Allen (1974), Bunting (1976), Pennings (1980), and Carrington (1981), firms' dimensions and profitabilities, together with industry sizes, are set as exogenous variables influencing link decisions.

As per Dooley (1969), Allen (1974) and Mizruchi and Stearns (1988; 1994), a firm's dimension is measured using total assets (hereafter TA)[81], extracted from the Cerved database. As per Baysinger and Butler (1985) and Fligstein and Brantley (1992), a firm's profitability is measured using return on equity (hereafter ROE)[82], extracted from the Cerved database. Lastly, in order to apply the inference method proposed by AS and discussed in Section 2.4, TA and ROE are discretised into ten separate bins, according to their 10, 20, ..., 90th quantiles. Consequently, TA and ROE take values in $\{1,2,...,10\}$. Additional data cleaning steps are in Appendix D.1.

Some descriptive statistics for industry size, TA and ROE are in Table 3.1. Some network summary statistics (averaged over industries) are in Table 3.2. Over-

[79]For example:

- A: Agriculture, hunting and fishing
- 01: Agriculture, hunting and related service activities
- 01.1: Crops
- 01.11: Growing of cereals and other arable crops
- 01.11.1: Growing of cereals (rice included)
- 01.11.2: Growing of oil seeds
- ...

[80]In 2008 the ATECO 2002 code was replaced by the ATECO 2007 code, whose structure preserves the same general characteristics. However, this chapter uses the ATECO 2002 code as its data quality is remarkably higher for the year 2010.

[81]Pfeffer (1972) measures a firm's size using total sales. Booth and Deli (1996) propose the natural log of the sum of the market value of the firm equity plus the book value of preferred stock.

[82]Alternative measures of a firm's profitability include: price-cost margins (Collins and Preston, 1969; Carrington, 1981); market value, price-earnings ratio and debt-equity ratio (Fligstein and Brantley, 1992); return on sales (Mizruchi and Stearns, 1988; Fligstein and Brantley, 1992); return on assets (Richardson, 1987; Mizruchi and Stearns, 1988; Fligstein and Brantley, 1992); return on shareholders' investment and net interest on assets (Bernstein, 1978; Pennings, 1980; Richardson, 1987); dividend cuts (Kaplan and Reishus, 1990); return on invested capital (Bunting, 1976); average Tobin's q (Hermalin and Weisbach, 1991).

all, networks look disconnected with several isolated nodes. Furthermore, the reported network summary statistics seem relatively stable over the years 2005-2010 (Tables 3.3), which legitimates modelling the formation of PHBIs as a static game with complete information. More comments are in Appendix D.2.

Table 3.1: Descriptive statistics for industry size, TA, ROE

	Mean	St.dev	Min	Max	$[0.25;0.50;0.75]$	Skewness	Kurtosis	Total number of firms	Total number of industries
N	6.733	3.483	3	15	$[4;6;9]$	0.812	2.645	2599	386
TA ($\times 10^6$ €)	117.281	$1,567.453$	0.067	$73,916.239$	$[6.998;15.653;39.984]$	41.471	$1,903.552$	✓	✓
ROE (%)	1.267	24.589	-128.410	69.820	$[-2.382;2.360;11.402]$	-1.600	9.071	✓	✓

Table 3.2: Some network summary statistics. Definitions are in Appendix D.2.

	Mean	St.dev	Min	Max	$[0.25;0.50;0.75]$ quantiles	Skewness	Kurtosis
Density	0.005	0.026	0	0.333	$[0;0;0]$	8.462	88.322
Average degree	0.023	0.096	0	1	$[0;0;0]$	5.905	45.181
% Isolated nodes	97.666	8.758	33.333	100	$[100;100;100]$	-4.299	22.587
Number of links	0.163	0.617	0	6	$[0;0;0]$	4.859	32.750

Table 3.3: Mean values for some network summary statistics across the years 2005-2010

Year	Density	Average degree	% Isolated nodes	Number of links
2005	0.006	0.025	97.604	0.167
2006	0.007	0.023	97.641	0.135
2007	0.009	0.028	96.974	0.159
2008	0.006	0.028	97.027	0.208
2009	0.005	0.027	97.206	0.208
2010	0.005	0.023	97.666	0.163

3.5 Results

Inference is conducted on the following model specification first

$$G_{ij} = \mathbb{1}\{\beta_0 + \beta_1(TA_j - TA_i) + \beta_2(ROE_j - ROE_i) + \delta \sum_{k \neq i}^{N} G_{kj} + \varepsilon_{ij}^{i} \geq 0\} \times \\ \mathbb{1}\{\gamma_0 + \gamma_1(TA_j - TA_i) + \gamma_2(ROE_j - ROE_i) + \varepsilon_{ij}^{j} \geq 0\} \tag{3.1}$$

$\forall i, j \in \mathcal{N}, i \neq j$. According to (3.1), firms i and j decide about the link ij considering their differences in size and profitability which affect payoffs in a linear way. Moreover, i takes into account the number of additional companies with an executive sitting on the board of j, as a measure of "cheap talk" opportunities for the reasons discussed earlier. To simplify inference, $\{\varepsilon_{ij}\}_{i,j \in \mathcal{N}, i \neq j}$ are assumed i.i.d. across ij, with $\varepsilon_{ij} := (\varepsilon_{ij}^{i}, \varepsilon_{ij}^{j})$ distributed as a standard bivariate normal.

Table 3.4 reports the hypercube that contains the 95% confidence region for each parameter value in Θ^{o}. The sign of various effects, as measured by projections of this hypercube, is analysed first. The projection for the parameter δ is $[2.129,\ 20.909]$. The positive sign reveals that, all else equal, firm i's payoff from appointing as executive a board member of firm j increases with the number of additional competitors doing the same. Such a result confirms the intuition that i finds attractive to have one of its executives sitting on the board of j when executives of other competitors are hosted too, as it enables i to engage with them in "cheap talk" communications, besides having the opportunity to learn about j's decision making process. The projections for the parameters β_1 and β_2 are, respectively, $[0.022,\ 8.381]$ and $[0.012,\ 7.486]$, and indicate that, all else equal, firm i prefers its executives sitting on the board of firm j when j is larger and more profitable than i. Indeed, it may be that firms prefer their executives sitting on the board of larger and more profitable competitors because these represent major sources of tacit knowledge or strategic uncertainty, and, therefore their decision making process is more worth to be observed. Conversely, the projections for the parameters γ_1 and γ_2 are, respectively, $[-4.327,\ -0.004]$ and $[-9.655,\ -0.016]$, and indicate that, all else equal, firm j prefers hosting on its board executives of firm i when j is smaller and less profitable than i. Indeed, it may be that smaller and less profitable firms are not considered capable of offering precious advice, or that their "cheap talk" confidentialities are not valuable enough.

Table 3.4: Projections of the 95% confidence region for each $\theta \in \Theta^o$ according to specification (3.1).

β_0	$[-15.399,\ -0.783]$
β_1	$[0.022,\ 8.381]$
β_2	$[0.012,\ 7.486]$
δ	$[2.129,\ 20.909]$
γ_0	$[-0.469,\ 37.490]$
γ_1	$[-4.327,\ -0.004]$
γ_2	$[-9.655,\ -0.016]$

One idea to discuss the magnitude of results is considering the ratio between the change induced by a given unit increase in one variable relative to the change induced by a one unit increase in another. In this sense, it can be seen that, to keep the payoff that firm i receives from the link ij constant when a link pointing to firm j is added, one would need to reduce $(ROE_j - ROE_i)$ of approximatively 2 bins at least, or reduce $(TA_j - TA_i)$ of approximatively 3 bins at least.

Differences in size and profitability may affect firms' payoffs non-linearly. In order to study such potential non-linearities, the following model specification is additionally considered

$$\begin{aligned}
G_{ij} = &\mathbb{1}\{\beta_0 + \beta_1 \mathbb{1}\{-9 \leq TA_j - TA_i \leq -5\} + \beta_2 \mathbb{1}\{-4 \leq TA_j - TA_i \leq 0\} + \beta_3 \mathbb{1}\{1 \leq TA_j - TA_i \leq 5\} + \\
&+ \beta_4 \mathbb{1}\{-9 \leq ROE_j - ROE_i \leq -5\} + \beta_5 \mathbb{1}\{-4 \leq ROE_j - ROE_i \leq 0\} + \beta_6 \mathbb{1}\{1 \leq ROE_j - ROE_i \leq 5\} + \\
&+ \delta \sum_{k \neq i} G_{kj} + \varepsilon_{ij}^i \geq 0\} \times \\
&\mathbb{1}\{\gamma_0 + \gamma_1 \mathbb{1}\{-9 \leq TA_j - TA_i \leq -5\} + \gamma_2 \mathbb{1}\{-4 \leq TA_j - TA_i \leq 0\} + \gamma_3 \mathbb{1}\{1 \leq TA_j - TA_i \leq 5\} + \\
&+ \gamma_4 \mathbb{1}\{-9 \leq ROE_j - ROE_i \leq -5\} + \gamma_5 \mathbb{1}\{-4 \leq ROE_j - ROE_i \leq 0\} + \gamma_6 \mathbb{1}\{1 \leq ROE_j - ROE_i \leq 5\} + \\
&+ \varepsilon_{ij}^j \geq 0\}
\end{aligned} \tag{3.2}$$

$\forall i, j \in \mathcal{N}, i \neq j$. As earlier, $\{\varepsilon_{ij}\}_{i,j \in \mathcal{N}, i \neq j}$ are assumed i.i.d. across ij, with ε_{ij} distributed as a standard bivariate normal.

Table 3.5 reports the hypercube that contains the 95% confidence region for each $\theta \in \Theta^o$. Consider first the sign of various effects as measured by projections of this hypercube. The projection for the parameter δ is $[31.523,\ 35.902]$ and has a positive sign, as for specification (3.1). The interpretation of the indicator function parameters is slightly more complicated and done as follows. Let the base group be $TA_j - TA_i$ and $ROE_j - ROE_i$ both between 6 and 9. Table 3.6 reports the confidence intervals for sum of pairs of parameters via projections relative to other combinations of realisations of $TA_j - TA_i$ and $ROE_j - ROE_i$. Overall, the base group is always favoured by firm i, i.e., i prefers its executives sitting on the board of firm j when j is significantly larger and more profitable than i. An exception is represented by the projection for $\beta_2 + \beta_6$ that includes both positive and negative values. This means that the corresponding indicator functions may have a positive or a negative effect on payoffs. Conversely, the base group is never favoured by firm j, i.e., j prefers hosting on its board executives of firm i when j is not significantly larger and more profitable than i. An exception is when $TA_j - TA_i$ is between -9 and -5 and $ROE_j - ROE_i$ is between -4 and 0, which is less favoured by j than the base group, possibly because j sees itself excessively vulnerable and exposed in front of i. Moreover, the projections for $\gamma_1 + \gamma_4$ and $\gamma_2 + \gamma_5$ include both positive and negative values.

Table 3.5: Projections of the 95% confidence region for each $\theta \in \Theta^o$ according to specification (3.2).

β_0	$[-7.120,\ -3.431]$
β_1	$[-2.006 \times 10^3,\ -1.998 \times 10^3]$
β_2	$[6.977,\ 12.562]$
β_3	$[0.202,\ 2.629]$
β_4	$[-23.251,\ -15.473]$
β_5	$[-25.678,\ -22.059]$
β_6	$[-14.954,\ -9.652]$
δ	$[31.523,\ 35.902]$
γ_0	$[0.845,\ 2.679]$
γ_1	$[-12.762,\ -7.030]$
γ_2	$[-7.958,\ -4.965]$
γ_3	$[-0.723,\ 1.785]$
γ_4	$[7.290,\ 11.360]$
γ_5	$[4.584,\ 7.291]$
γ_6	$[13.156,\ 16.360]$

Table 3.6: Projections of sums for interpreting signs according to specification (3.2).

$\beta_1 + \beta_4$	$[-2.029 \times 10^3,\ -2.013 \times 10^3]$
$\beta_1 + \beta_5$	$[-2.029 \times 10^3,\ -2.022 \times 10^3]$
$\beta_1 + \beta_6$	$[-2.017 \times 10^3,\ -2.010 \times 10^3]$
$\beta_2 + \beta_4$	$[-14.348,\ -5.527]$
$\beta_2 + \beta_5$	$[-17.541,\ -10.957]$
$\beta_2 + \beta_6$	$[-6.401,\ 1.372]$
$\beta_3 + \beta_4$	$[-22.973,\ -14.692]$
$\beta_3 + \beta_5$	$[-25.097,\ -20.983]$
$\beta_3 + \beta_6$	$[-14.112,\ -7.810]$
$\gamma_1 + \gamma_4$	$[-3.776,\ 1.016]$
$\gamma_1 + \gamma_5$	$[-6.954,\ -2.445]$
$\gamma_1 + \gamma_6$	$[2.224,\ 8.544]$
$\gamma_2 + \gamma_4$	$[1.266,\ 5.295]$
$\gamma_2 + \gamma_5$	$[-1.449,\ 0.111]$
$\gamma_2 + \gamma_6$	$[7.057,\ 10.608]$
$\gamma_3 + \gamma_4$	$[7.096,\ 13.145]$
$\gamma_3 + \gamma_5$	$[4.403,\ 7.670]$
$\gamma_3 + \gamma_6$	$[13.162,\ 17.150]$

To comment the magnitude of results, the strategy used earlier can not be applied to specification (3.2) due to non-linearities. An alternative is studying how bounds on density, average degree, percentage of isolated nodes and number of links vary as a consequence of changes in TA or ROE. Various experiments are possible. As an example, Table 3.7 reports the outcome of the following procedure: for each industry and value of parameters in the 95% confidence region, the discretised amount of total assets of the smallest firms is equalised to the discretised amount of total assets of the biggest firms, hence reducing size heterogeneity within industries; several realisations of preference shocks are drawn; for each drawn realisation, PSPNS networks are found; the density, the average degree, the percentage of isolated nodes and the total number of links in each PSPNS network are computed, and their minimum and maximum values across PSPNS networks are recorded; bounds are then averaged across drawn realisations and industries; finally, the smallest lower bound and largest upper bounds across values of parameters are reported in the second and third columns of Table 3.7. The same experiment is repeated keeping the amount of total assets within each industry at the observed values and results are reported in the fourth and sixth columns of Table 3.7. Lastly, observed empirical values are in the fifth column of Table 3.7. As a consequence of the simulated shift, the upper bounds on density, average degree and number of links decrease. The lower bound on the percentage of isolated nodes decreases. Hence, by reducing heterogeneity in firms' size, networks can have more isolated nodes and can become more disconnected.

Table 3.7: Bounds on some network summary statistics according to specification (3.2) when the following experiment is run: within each industry, the discretised amount of total assets of the smallest firms is equalised to the discretised amount of total assets of the biggest firms.

	New lower bound	New upper bound	Old lower bound	Empirical value	Old upper bound
Density	0	0.800	0	0.005	0.886
Average Degree	0	4.640	0	0.022	5.153
% Isolated nodes	0.063	100	0.139	97.666	100
Number of links	0	41.096	0	0.163	45.808

3.6 Conclusions

The chapter provides an illustration of the methodology developed in Chapter 2 to empirically investigate which preferences are behind firms' decisions to appoint competitors' directors as executives, using data on Italian joint stock companies. It is found that a firm i prefers its executives sitting on the board of a rival j when executives of other competitors are hosted too, possibly because it enables i to engage with them in "cheap talk" communications, besides having the opportunity to learn about j's decision making process.

There are some avenues of future research. Firstly, companies can be connected also through cross-ownerships. It could be worth investigating whether and how these ties relate to PHBIs. Moreover, the chapter postulates that executives act as representatives of their companies' will. In reality, principal-agent issues are pervasive and one possibility would be to consider a more sophisticated structural model including them, together with board members' identities. Furthermore, the study takes a firm's entrance in an industry as an exogenous event. One option to remove such assumption might be to build an unique model of entrance and PHBIs formation. Lastly, from a policy perspective, what is the impact of a potential information exchange happening through PHBIs on market structures remains an open question: from one hand, it may help to develop or sustain collusive behaviours; from the other it could improve competition by increasing transparency. Providing an answer requires a richer model and more data, a direction which might be valuable to explore.

General conclusions

This dissertation addresses topics in the econometrics of network formation models. Chapter 1 provides a review of the literature. Statistical models focus on the specification of the probability distribution of the network. Examples include models in which nodes are born sequentially and meet existing vertices according to random meetings and network-based meetings. Within this group of models, special attention is reserved to the milestone work by Jackson and Rogers (2007): after having discussed and replicated the main results of the paper, an extension of the original model is examined and fitted to a dataset of Google Plus users.

Even if statistical models can reproduce relatively well the main characteristics of real networks, they usually lack of microfundation, essential for counterfactual analysis. The chapter hence moves to considering the econometrics of economic models of network formation, where agents form links in order to maximise a payoff function. Within this framework, Chapter 2 studies identification of the parameters governing agents' preferences in a static game of network formation, where links represent asymmetric relations between players. Agents have complete information and play PSNE when link formation is unilaterally decided, or PSPNE in the bilateral case. Payoffs are non-transferable. Link decisions are interdependent, as the payoff that player i receives from linking to player j is assumed to be monotonically affected by the number of additional players doing the same. After having shown existence of an equilibrium, partial identification arguments are provided without restrictions on equilibrium selection. The usual computational difficulties are attenuated by restricting the attention to some local games of the network formation game and giving up on sharpness. Overall, Monte Carlo exercises show that conducting inference on the suggested outer set is computationally manageable using relatively limited computational resources, with up to 20 players. The chapter offers some avenues of future research. Specifically, there could be other interdependencies among link decisions to consider. For example, player i's payoff from linking to player j may also be affected on the number of connections already formed by i, or on the links and characteristics of the additional agents connecting to j. It may be worth enriching players' payoffs in this direction and investigating whether the identification results proposed here can be extended to such more complicated settings. Another option could be to examine how the identification analysis changes if one removes the additive separability over outgoing and incoming links characterising payoffs. For example, one may wonder how to adjust bounds when $v(\cdot;\delta)$ has a "U" shape.

Chapter 3 applies the methodology developed in Chapter 2 to empirically investigate which preferences are behind firms' decisions to appoint competitors' directors as executives. Using data on Italian companies, it is found that a firm i prefers its executives sitting on the board of a rival j when executives of other competitors are hosted too, possibly because it enables i to engage with them in "cheap talk" communications, besides having the opportunity to learn about j's decision making process. Extensions are possible. Firstly, companies may be connected also through ownership participations. It may be worth investigating whether and how these ties relate to PHBIs. Secondly, the chapter postulates that firms are identified with their executives. In reality, principal-agent issues might be pervasive and one possibility would be to consider a more sophisticated structural model including them, together with board members' identities. Thirdly, the study takes a firm's entrance in an industry as an exogenous event. One option to remove such assumption might be to build an unique model of entrance and PHBIs formation. Lastly, from a policy perspective, the impact of the information exchange happening through PHBIs on market structures remains an open question: from one hand, it may help to develop or sustain collusive behaviours; from the other it could improve competition by increasing transparency. Providing an answer requires a richer model and more data, a direction which might be valuable to explore.

Appendix A

Existence of a PSPNS network

Existence of a PSPNS network for every value of payoff-relevant variables and parameters can be shown following the strategy adopted for the unilateral case in section 2.3.3. Specifically, for any $j \in \mathcal{N}$, in the *bilateral section j game* players other than player j simultaneously reveal whether they want to form a link pointing to j, j replies and only mutually announced links are created. A pure strategy vector of player j is $s^j_{\cdot j} \in \{0,1\}^{n-1}$ collecting s^j_{ij} $\forall i \neq j \in \mathcal{N}$. For any $i \neq j \in \mathcal{N}$, a pure strategy of player i is $s^i_{ij} \in \{0,1\}$. A pure strategy profile of the game is $s_{\cdot j} \in \{0,1\}^{2(n-1)}$ listing s^i_{ij} $\forall i \neq j \in \mathcal{N}$ and $s^j_{\cdot j}$. Each player $i \neq j \in \mathcal{N}$ gets as payoff

$$U^j_i(G_{\cdot j}, \boldsymbol{X}, \varepsilon^i_{\cdot j}; \theta_u) := G_{ij} \times \left[z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon^i_{ij} \right]$$

Player j gets as payoff

$$U^j_j(G_{\cdot j}; \boldsymbol{X}, \varepsilon^j_{\cdot j}; \theta_u) := \sum_{i=1}^{n} G_{ij} \times \left[b(X_i, X_j; \gamma) + \varepsilon^j_{ij} \right]$$

Agents play PSPNE and the resulting *section j* is a PSPNS *section j*. Definitions are now given. Let the dependence of $G_{\cdot j}$ on $s_{\cdot j}$ be denoted by $G_{\cdot j}(s_{\cdot j})$.

Definition 4. (PSPNS *section j*) $s_{\cdot j}$ is a PSPNE of the *section j game* if

$$s^i_{ij} = \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}(s_{\cdot j}); \delta) + \varepsilon^i_{ij} \geq 0\}$$

and

$$s_{ij}^{j} = \mathbb{1}\{b(X_i, X_j; \gamma) + \varepsilon_{ij}^{j} \geq 0\}$$

$\forall i \neq j \in \mathcal{N}$. $G_{\cdot j}$ is a PSPNS *section j* if there exists a PSPNE $s_{\cdot j}$ of the *section j game* such that $G_{\cdot j} = G_{\cdot j}(s_{\cdot j})$, i.e.,

$$G_{ij} = \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon_{ij}^{i} \geq 0\}\mathbb{1}\{b(X_i, X_j; \gamma) + \varepsilon_{ij}^{j} \geq 0\} \quad \forall i \neq j \in \mathcal{N}$$

$\star$

As for the unilateral case,

Lemma 5. (Decomposing the bilateral network formation game) ***G*** is a PSPNS network if and only if $G_{\cdot j}$ is a PSPNS *section j* $\forall j \in \mathcal{N}$. $\diamond$

Moreover, using Tarski's fixed point theorem when $v(\cdot; \delta)$ is monotone increasing and a bilateral game reinterpretation of the constructive proof in Berry (1992) when $v(\cdot; \delta)$ is monotone decreasing, it can be shown that

Lemma 6. (Existence of a PSPNS *section j*) There exists a PSPNS *section j* $\forall j \in \mathcal{N}$. $\diamond$

Hence, by Lemmas 5 and 6,

Proposition 5. (Existence of a PSPNS network) There exists a PSPNS network. $\diamond$

Proofs of Lemmas 5, 6 and Proposition 5 are in Appendix B.

Appendix B

Proofs

Proof of Lemma 1 (Lemma 4). Before starting the proof, notice that

$$G_{ij} = \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{N} G_{kj}; \delta) + \varepsilon_{ij} \geq 0\} \; \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \, i \neq j$$

is equivalent to

$$U_i(G_{ij}, \boldsymbol{G}_{-\{ij\}}, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\tilde{G}_{ij}, \boldsymbol{G}_{-\{ij\}}, \boldsymbol{X}, \varepsilon; \theta_u) \text{ for } \tilde{G}_{ij} \neq G_{ij} \in \{0,1\}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \, i \neq j \tag{B.1}$$

where $\boldsymbol{G}_{-\{ij\}}$ is the matrix $\boldsymbol{G}$ with ijth element deleted.

It is firstly proved that if $\boldsymbol{G}$ is a PSNE of the network formation game, then (B.1) is satisfied. For any $i \in \mathcal{N}, j \in \mathcal{N}$ with $i \neq j$, let $G_{i\cdot -\{ij\}}$ be the vector $G_{i\cdot}$ with ijth element removed. By setting $\tilde{G}_{i\cdot} = (\tilde{G}_{ij}, G_{i\cdot -\{ij\}})$ with $\tilde{G}_{ij} \neq G_{ij}$ in $U_i(G_{i\cdot}, \boldsymbol{G}_{-\{i\cdot\}}, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\tilde{G}_{i\cdot}, \boldsymbol{G}_{-\{i\cdot\}}, \boldsymbol{X}, \varepsilon; \theta_u)$ of Definition 1, it follows that $U_i(G_{ij}, \boldsymbol{G}_{-\{ij\}} \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\tilde{G}_{ij}, \boldsymbol{G}_{-\{ij\}} \boldsymbol{X}, \varepsilon; \theta_u)$ and this is verified $\forall i \in \mathcal{N}, \forall j \in \mathcal{N}$ with $i \neq j$.

Conversely, it is proved that if (B.1) holds, then $\boldsymbol{G}$ is a PSNE of the network formation game. For any $i \in \mathcal{N}$, if $U_i(G_{ij}, \boldsymbol{G}_{-\{ij\}}, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\tilde{G}_{ij}, \boldsymbol{G}_{-\{ij\}}, \boldsymbol{X}, \varepsilon; \theta_u)$, then, by the additive separability of $U_i(\cdot; \theta_u)$ over player i's outgoing links, $U_i(G_{i\cdot}, \boldsymbol{G}_{-\{i\cdot\}}, \boldsymbol{X}, \varepsilon; \theta_u) \geq U_i(\tilde{G}_{i\cdot}, \boldsymbol{G}_{-\{i\cdot\}}, \boldsymbol{X}, \varepsilon; \theta_u) \; \forall \tilde{G}_{i\cdot} \neq G_{i\cdot} \in \{0,1\}^{N-1}$ and this is verified $\forall i \in \mathcal{N}$.

Lemma 4 can be shown analogously. $\square$

Proof of Lemma 2 (Lemma 5). It is firstly proved that if $\boldsymbol{G}$ is a PSNE of the network formation game, then $G_{\cdot j}$ is a PSNE of the *section j game* $\forall j \in \mathcal{N}$. By Lemma 1, if $\boldsymbol{G}$ is a PSNE of the network formation game, then $G_{ij} = \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon_{ij} \geq 0\}$ $\forall i \in \mathcal{N}, \forall j \in \mathcal{N}$ with $i \neq j$. This set of conditions also includes those defining $G_{\cdot j}$ as a PSNE of the *section j game* $\forall j \in \mathcal{N}$. Therefore, $G_{\cdot j}$ is a PSNE of the *section j game* $\forall j \in \mathcal{N}$.

Conversely, it is proved that if $G_{\cdot j}$ is a PSNE of the *section j game* $\forall j \in \mathcal{N}$, then $\boldsymbol{G}$ is a PSNE of the network formation game. $\forall j \in \mathcal{N}$, if $G_{\cdot j}$ is a PSNE of the *section j game*, then, by Definition 2, $G_{ij} = \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon_{ij} \geq 0\}$ $\forall i \neq j \in \mathcal{N}$. Hence, the conditions of Lemma 1 are satisfied and $\boldsymbol{G}$ is a PSNE of the network formation game.

Lemma 5 can be shown analogously. □

Theorem 1. (Tarski's fixed point theorem) Let $F(x)$ be a monotone increasing function from a non-empty complete lattice X into X. Then,

(i) the set of fixed points of $F(x)$ in X is non-empty, where $\sup_x(\{x \in X, x \leq F(x)\})$ and $\inf_x(\{x \in X, x \geq F(x)\})$ denote, respectively, the greatest and the least fixed points;

(ii) the set of fixed points of $F(x)$ in X is a non-empty complete lattice.

◇

Proof of Lemma 3 (Lemma 6). Consider any $j \in \mathcal{N}$. It is firstly discussed the case in which $v(\sum_{k \neq i}^{n} G_{kj}; \delta)$ is monotone increasing. Let

$$h_{ij}(G_{\cdot j}) := \mathbb{1}\{z(X_i, X_j; \beta) + v(\sum_{k \neq i}^{n} G_{kj}; \delta) + \varepsilon_{ij} \geq 0\} \; \forall i \neq j \in \mathcal{N}$$

and

$$h(G_{\cdot j}) := \begin{pmatrix} h_{1j}(G_{\cdot j}) \\ h_{2j}(G_{\cdot j}) \\ \vdots \\ h_{j-1\,\mathrm{j}}(G_{\cdot j}) \\ h_{j+1\,\mathrm{j}}(G_{\cdot j}) \\ \vdots \\ h_{nj}(G_{\cdot j}) \end{pmatrix}$$

Hence, $h : \{0,1\}^{n-1} \to \{0,1\}^{n-1}$. It is possible to show that the function h satisfies the sufficient conditions of Theorem 1 when $v(\sum_{k\neq i}^{n} G_{kj};\delta)$ is monotone increasing, which, in turn, guarantees existence of a PSNE of the *section j game* when $v(\sum_{k\neq i}^{n} G_{kj};\delta)$ is monotone increasing. Indeed, let the comparison between vectors be coordinate-wise, i.e., for any $G_{\cdot j} \in \{0,1\}^{n-1}, G'_{\cdot j} \in \{0,1\}^{n-1}$

$$G_{\cdot j} \geq G'_{\cdot j} \Leftrightarrow G_{ij} \geq G'_{ij} \ \forall i \neq j \in \mathcal{N}$$

Thus, $G_{\cdot j} = G'_{\cdot j}$ if and only if $G_{\cdot j} \geq G'_{\cdot j}$ and $G_{\cdot j} \leq G'_{\cdot j}$. Moreover, $G_{\cdot j}$ and $G'_{\cdot j}$ are unordered if and only if neither $G_{\cdot j} \geq G'_{\cdot j}$ nor $G_{\cdot j} \leq G'_{\cdot j}$. Therefore, $\{0,1\}^{n-1}$ is a lattice, i.e, a set with a partial order. As $\{0,1\}^{n-1}$ is a finite lattice, it is complete. Furthermore, if $v(\sum_{k\neq i}^{n} G_{kj};\delta)$ is monotone increasing, then h is a monotone increasing function. In fact, consider two vectors $G_{\cdot j} \geq G'_{\cdot j}$. Since $G_{ij} \geq G'_{ij} \ \forall i \neq j \in \mathcal{N}$, then $z(X_i,X_j;\beta) + v(\sum_{k\neq i}^{n} G_{kj};\delta) + \varepsilon_{ij} \geq z(X_i,X_j;\beta) + v(\sum_{k\neq i}^{n} G'_{kj};\delta) + \varepsilon_{ij} \ \forall i \neq j \in \mathcal{N}$. Hence, $h(G_{\cdot j}) \geq h(G'_{\cdot j})$ and the sufficient conditions of the theorem are met.

Now, the case in which $v(\sum_{k\neq i}^{n} G_{kj};\delta)$ is monotone decreasing is considered. As explained in section 2.3.3, it can be noticed that the structure of the *section j game* when $v(\cdot;\delta)$ is monotone decreasing is similar to the structure of an entry game with substitution effects. Existence of a PSNE in an entry game with substitution effects is proved by Berry (1992) by means of a constructive proof which can be reinterpreted for the *section j game* as follows. Let $Y_{ij} := z(X_i,X_j;\beta) + \varepsilon_{ij}$. The elements $(Y_{ij})_{\forall i\neq j\in\mathcal{N}}$ are ordered from largest to smallest. Let $k \in \{1,\ldots,n-1\}$ denote the position of Y_{ij} in the ordered list and let π be a function such that $\pi(ij) = k$,

$\forall i \neq j \in \mathcal{N}$. By replacing the subscript ij with k, the ordered sequence is

$$Y_1 \geq Y_2 \geq \ ... \ \geq Y_k \geq \ ... \ \geq Y_{n-1}$$

Let $Y_0 := \max\{Y_1, -v(-1;\delta)\}$. $n^*_{\cdot j}$ is defined as the largest element of the set of integers $\{0, 1, \ ... \ , k, \ ... \ , n-1\}$ satisfying $Y_{n^*_{\cdot j}} + v(n^*_{\cdot j} - 1;\delta) \geq 0$, i.e.

$$n^*_{\cdot j} := \max\{k \in \{0, ..., n-1\} | Y_k + v(k-1;\delta) \geq 0\}$$

Consider $G_{\cdot j}$ with $G_{ij} = 1$ if $\pi(ij) \leq n^*_{\cdot j}$ and $G_{ij} = 0$ otherwise. One can see that $G_{\cdot j}$ is a PSNE of the *section j game*. In fact choosing $n^*_{\cdot j}$ according to the previous criterion means that

$$\begin{aligned}
Y_0 + v(-1;\delta) \geq 0 \quad & \text{(a)} \\
Y_1 + v(0;\delta) \geq 0 \quad & \text{(b)} \\
& ... \\
Y_{n^*_{\cdot j}} + v(n^*_{\cdot j} - 1;\delta) \geq 0 \quad & \text{(c)} \\
Y_{n^*_{\cdot j}+1} + v(n^*_{\cdot j};\delta) < 0 \quad & \text{(d)} \\
Y_{n^*_{\cdot j}+2} + v(n^*_{\cdot j} + 1;\delta) < 0 \quad & \text{(e)} \\
& ... \\
Y_{n-1} + v(n-2;\delta) < 0 \quad & \text{(f)}
\end{aligned}$$

For $G_{\cdot j}$ being a PSNE of the *section j game*, the following inequalities should be satisfied

$$Y_1 + v(n^*_{\cdot j} - 1; \delta) \geq 0 \quad \text{(g)}$$

$$Y_2 + v(n^*_{\cdot j} - 1; \delta) \geq 0 \quad \text{(h)}$$

$$...$$

$$Y_{n^*_{\cdot j}} + v(n^*_{\cdot j} - 1; \delta) \geq 0 \quad \text{(i)}$$

$$Y_{n^*_{\cdot j}+1} + v(n^*_{\cdot j}; \delta) < 0 \quad \text{(l)}$$

$$Y_{n^*_{\cdot j}+2} + v(n^*_{\cdot j}; \delta) < 0 \quad \text{(m)}$$

$$...$$

$$Y_{n-1} + v(n^*_{\cdot j}; \delta) < 0 \quad \text{(n)}$$

By observing that inequalities (g), (h), ... ,(i) are implied by inequality (c) and all the other inequalities follow from inequality (d), it can be concluded that $G_{\cdot j}$ is a PSNE of the *section j game*.

Lemma 6 can be proved analogously after having replaced PSNE with PSPNE and imposed

$$Y_{ij} := \begin{cases} z(X_i, X_j; \beta) + \varepsilon^i_{ij} \text{ if } b(X_i, X_j; \gamma) + \varepsilon^j_{ij} \geq 0 \\ -\infty \text{ otherwise} \end{cases}$$

Moreover, Berry (1992) shows that in an entry game with substitution effects all the equilibria are characterised by the same number of firms entering the market. Reinterpreting this result for the *section j game*, it can be proved that all the PSNE of the *section j game* are characterised by the same number, $n^*_{\cdot j}$, of players linking to player j.

In fact, suppose there is some equilibrium with $k^* > n^*_{\cdot j}$ edges. None of the players whom inequalities (l), (m),..., (n) above are referred to is willing to form a link pointing to player j, as inequalities (l), (m),..., (n) imply that

$$Y_{n^*_{\cdot j}+1} + v(k^* - 1; \delta) < 0 \quad \text{(l')}$$

$$Y_{n^*_{\cdot j}+2} + v(k^* - 1; \delta) < 0 \quad \text{(m')}$$

$$\ldots$$

$$Y_{n-1} + v(k^* - 1; \delta) < 0 \quad \text{(n')}$$

It follows that there cannot be some equilibrium with $k^* > n^*_{\cdot j}$ edges.

Viceversa, suppose there is some equilibrium with $k^* < n^*_{\cdot j}$ edges. All of the players whom inequalities (g), (h),..., (i) above are referred to are willing to form a link pointing to player j, as inequalities (g), (h),..., (i) imply that

$$Y_1 + v(k^*; \delta) \geq 0 \quad \text{(g')}$$

$$Y_2 + v(k^*; \delta) \geq 0 \quad \text{(h')}$$

$$\ldots$$

$$Y_{n^*_{\cdot j}} + v(k^*; \delta) \geq 0 \quad \text{(i')}$$

It follows that there cannot be some equilibrium with $k^* < n^*_{\cdot j}$ edges.

□

Proof of Proposition 1 (Proposition 5). By Lemma 3, there exists a PSNE of the *section j game* $\forall j \in \mathcal{N}$. By Lemma 2, if $G_{\cdot j}$ is a PSNE of the *section j game* $\forall j \in \mathcal{N}$, then $\boldsymbol{G}$ is a PSNE of the network formation game. Thus, the network formation game has a PSNE.

Proposition 5 can be shown analogously.

□

Construction of $\mathcal{A}_{\mathcal{G}}$ Consider the sets $K_{\cdot 1} \in \mathcal{K}_{\{0,1\}^{n-1}}, \ldots, K_{\cdot n} \in \mathcal{K}_{\{0,1\}^{n-1}}$. Construct the set $\mathcal{B}_{K_{\cdot 1},\ldots,K_{\cdot n}} := \bigtimes_{j=1}^{n} K_{\cdot j}$. Hence, $\mathcal{B}_{K_{\cdot 1},\ldots,K_{\cdot n}}$ is a collection of $L := \Pi_{j=1}^{n} |K_{\cdot j}|$ sets and it can be written as $\{B_l\}_{l=1}^{L}$. Any set $B_l \in \mathcal{B}_{K_{\cdot 1},\ldots,K_{\cdot n}}$ is composed by n vectors of dimension $(n-1) \times 1$. Hence, $B_l := \{b_{l,1}, \ldots, b_{l,n}\}$ with $b_{l,h} := (b^1_{l,h} \ \ldots \ b^{n-1}_{l,h})'$

$\forall h \in \{1,...,n\}$, $\forall l \in \{1,...,L\}$. Create the $n \times n$ matrix

$$C_l := \begin{pmatrix} 0 & b^1_{l,2} & b^1_{l,3} & \dots & b^1_{l,n} \\ b^1_{l,1} & 0 & b^2_{l,3} & \dots & b^2_{l,n} \\ b^2_{l,1} & b^2_{l,2} & 0 & \dots & b^3_{l,n} \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ b^{n-2}_{l,1} & b^{n-2}_{l,2} & b^{n-2}_{l,3} & \dots & b^{n-1}_{l,n} \\ b^{n-1}_{l,1} & b^{n-1}_{l,2} & b^{n-1}_{l,3} & \dots & 0 \end{pmatrix}$$

$\forall l \in \{1,...,L\}$. Let $A := \{C_1,...,C_L\}$. Repeat the procedure for all possible ordered n-tuples with repetition of $\mathcal{K}_{\{0,1\}^{n-1}}$ and denominate the family of sets A's as $\mathcal{A}_{\mathcal{G}}$. Notice that $|\mathcal{A}_{\mathcal{G}}| = (2^{2^{n-1}} - 1)^n < |\mathcal{K}_{\mathcal{G}}| = 2^{2^{n(n-1)}} - 1$.

For example, suppose $n = 3$. Hence,

$$\{0,1\}^2 := \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \}$$

with $|\{0,1\}^2| = 4$,

$$\mathcal{K}_{\{0,1\}^2} := \Big\{ \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \}, \{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \}, \{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \},$$

$$\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \}, \{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \},$$

$$\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \}, \{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \},$$

$$\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \} \Big\}$$

with $|\mathcal{K}_{\{0,1\}^2}| = 15$ and

$$\mathcal{K}_{\mathcal{G}} := \left\{ \{ \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \}, \ldots \right\}$$

with $|\mathcal{K}_{\mathcal{G}}| = 2^{64} - 1$.

Let $K_{\cdot 1} := \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}$, $K_{\cdot 2} := \{0,1\}^2$, $K_{\cdot 3} := \{0,1\}^2$. Hence,

$$
\begin{aligned}
&\mathcal{B}_{K_{\cdot 1},K_{\cdot 2},K_{\cdot 3}} := \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \} \bigtimes \{0,1\}^2 \bigtimes \{0,1\}^2 \\
&= \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \} \bigtimes \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \} \bigtimes \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \} \\
&= \Big\{ \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \\
&\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \\
&\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \\
&\{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \{ \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \\
&\{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \}, \\
&\{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix} \}, \\
&\{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix} \}, \\
&\{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \}, \{ \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix} \} \Big\}
\end{aligned}
$$

with cardinality $L = 32$. Therefore,

$$
\begin{aligned}
A := \{ &\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \\
&\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \}
\end{aligned}
$$

with $|A| = L$.

Proof of Proposition 2. The proof is articulated as follows: in step 1, any set $K \in$

$\mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1, ..., |\mathcal{G}| - 2\}$ satisfying the sufficient conditions of Proposition 2 is considered; in steps 2 and 3, Artstein's inequalities for the sets $\{K \cup D\}$ and $\{C \setminus D\}$ are rewritten in equivalent, but useful, ways; by combining steps 2 and 3 with Artstein's inequality for the set $\mathcal{G}$ (trivial Artstein's inequality) in step 4, it follows that

$$\mathbb{P}(G \in K | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap K \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \tag{B.2}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s.

Step 1 Consider a set $K \in \mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1, ..., |\mathcal{G}| - 2\}$, where the set $C := \mathcal{G} \setminus K$ is such that $\exists$ a non-empty set $D \subset C$ with $\{\tilde{D} \cup \tilde{C}\} \notin \mathcal{A}_{\mathcal{G}}$ $\forall \tilde{D} \subseteq D$ and $\forall \tilde{C} \subseteq \{C \setminus D\}$[83]. Assume that Artstein's inequality for the set $\{K \cup D\}$

$$\mathbb{P}(\boldsymbol{G} \in \{K \cup D\} | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap \{K \cup D\} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \tag{B.3}$$

and for the set $\{K \cup \{C \setminus D\}\}$

$$\mathbb{P}(\boldsymbol{G} \in \{K \cup \{C \setminus D\}\} | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap \{K \cup \{C \setminus D\}\} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \tag{B.4}$$

are satisfied, $\forall \boldsymbol{x} \in \mathcal{X}$ a.s.

Step 2 (B.3) is equivalent to

$$1 - \mathbb{P}(\boldsymbol{G} \in \{K \cup D\} | \boldsymbol{X} = \boldsymbol{x}) \geq 1 - \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap \{K \cup D\} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \tag{B.5}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s., which is equivalent to

$$\mathbb{P}(\boldsymbol{G} \in \{C \setminus D\} | \boldsymbol{X} = \boldsymbol{x}) \geq \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \text{ hits } \{C \setminus D\} \text{ only } | \boldsymbol{X} = \boldsymbol{x})] \tag{B.6}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s.

[83]Notice that $|K| \leq |\mathcal{G}| - 2$ because $D = \emptyset$ for $|K| \in \{|\mathcal{G}| - 1, |\mathcal{G}|\}$.

Step 3 (B.4) is equivalent to

$$1-\mathbb{P}(\boldsymbol{G}\in\{K\cup\{C\setminus D\}\}|\boldsymbol{X}=\boldsymbol{x})\geq 1-\mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)\cap\{K\cup\{C\setminus D\}\}\neq\emptyset|\boldsymbol{X}=\boldsymbol{x}) \tag{B.7}$$

$\forall\boldsymbol{x}\in\mathcal{X}$ a.s., which is equivalent to

$$\mathbb{P}(\boldsymbol{G}\in D|\boldsymbol{X}=\boldsymbol{x})\geq\mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)\text{ hits }D\text{ only }|\boldsymbol{X}=\boldsymbol{x})] \tag{B.8}$$

$\forall\boldsymbol{x}\in\mathcal{X}$ a.s.

Step 4 Moreover, consider Artstein's inequality for the set $\mathcal{G}$

$$\begin{aligned}1=&\mathbb{P}(G\in K|\boldsymbol{X}=\boldsymbol{x})+\mathbb{P}(G\in\{C\setminus D\}\boldsymbol{X}=\boldsymbol{x})+\mathbb{P}(G\in D|\boldsymbol{X}=\boldsymbol{x})=\\&\mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)\cap K\neq\emptyset|\boldsymbol{X}=\boldsymbol{x})+\mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)\text{ hits }\{C\setminus D\}\text{ only }|\boldsymbol{X}=\boldsymbol{x})+\\&\mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)\text{ hits }D\text{ only }|\boldsymbol{X}=\boldsymbol{x})+\underbrace{\mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)\text{ hits both }\{C\setminus D\},D|\boldsymbol{X}=\boldsymbol{x})}_{=0\text{ by construction}}\end{aligned} \tag{B.9}$$

$\forall\boldsymbol{x}\in\mathcal{X}$ a.s.

Hence, if (B.6), (B.8), and (B.9) hold $\forall\boldsymbol{x}\in\mathcal{X}$ a.s., then (B.2) should be satisfied $\forall\boldsymbol{x}\in\mathcal{X}$ a.s.

□

Proof of Corollary 1. The proof is articulated as follows: step 1 shows the sufficiency of the condition provided by Corollary 1 by considering the construction of $\mathcal{A}_{\mathcal{G}}$; step 2 shows its necessity by contradiction.

Step 1 It is shown that if, given a set $K\in\mathcal{K}_{\mathcal{G}}$ with $|K|\in\{1,...,|\mathcal{G}|-2\}$, the set $C:=\mathcal{G}\setminus K$ is such that $\exists$ a non-empty $D\subset C$ with all the pairs of matrices $\boldsymbol{g}_D\in D,\boldsymbol{g}_{\{C\setminus D\}}\in\{C\setminus D\}$ differing for at least two rows, then $\{\tilde{D}\cup\tilde{C}\}\notin\mathcal{A}_{\mathcal{G}}$ $\forall\tilde{D}\subseteq D$ and $\forall\tilde{C}\subseteq\{C\setminus D\}$.

This comes from the fact that, by construction, any set $K\in\mathcal{A}_{\mathcal{G}}$ cannot be partitioned into two non-empty subsets K_1,K_2 with all pairs of matrices $\boldsymbol{g}_{K_1}\in K_1,\boldsymbol{g}_{K_2}\in K_2$ differing for two rows at least.

Step 2 It is shown that if, given a set $K \in \mathcal{K}_\mathcal{G}$ with $|K| \in \{1, ..., |\mathcal{G}|-1\}$, the set $C := \mathcal{G} \setminus K$ is such that $\exists$ a non-empty $D \subset C$ with $\{\tilde{D} \cup \tilde{C}\} \notin \mathcal{A}_\mathcal{G}$ $\forall \tilde{D} \subseteq D$ and $\forall \tilde{C} \subseteq \{C \setminus D\}$, then all the pairs of matrices $\boldsymbol{g}_D \in D, \boldsymbol{g}_{\{C \setminus D\}} \in \{C \setminus D\}$ differ for at least two rows.

By contradiction: suppose that there is a pair of matrices $\boldsymbol{g}_D \in D, \boldsymbol{g}_{\{C \setminus D\}} \in \{C \setminus D\}$ differing for one row only. Then, the set $\{\boldsymbol{g}_D, \boldsymbol{g}_{\{C \setminus D\}}\} \in \mathcal{A}_\mathcal{G}$ which contradicts the assumptions.

$\square$

Proof of Corollary 2. The proof is articulated as follows: step 1 represents the matrices in the set $\mathcal{G}$ as a graph which is shown to be distance-regular; step 2 introduces a result about the vertex connectivity of such a graph; step 3 shows that, by step 2, if a set $K \in \mathcal{K}_\mathcal{G}$ has cardinality strictly less than $n(2^{n-1})$, then the necessary and sufficient condition of Corollary 1 is violated; step 4 uses step 3 to derive an upper bound on the number of sets satisfying the necessary and sufficient condition of Corollary 1; step 5 provides a lower bound on the number of sets satisfying the necessary and sufficient condition of Corollary 1.

Step 1 Consider the graph $\mathcal{W}$ of size $|\mathcal{G}|$ where the nodes represent the matrices in the set $\mathcal{G}$ and there is a link (undirected) between two nodes if the corresponding matrices differ for one row only. It can be noticed that the graph $\mathcal{W}$ is distance-regular, meaning that: (i) it is connected, i.e., there is a path[84] between every pair of nodes; (ii) each node has the same degree; (iii) for every two nodes $\boldsymbol{g}_1 \in \mathcal{G}, \boldsymbol{g}_2 \in \mathcal{G}$ the number of vertices at distance[85] d_1 from $\boldsymbol{g}_1$ and at distance d_2 from $\boldsymbol{g}_2$ depends only upon d_1, d_2, and the distance between $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$.

While (i) holds by construction, (ii) comes from the fact that, for any matrix $\boldsymbol{g} \in \mathcal{G}$, $2^{n-1}-1$ is the number of possible variations of a given row of $\boldsymbol{g}$ and n is the number of rows of $\boldsymbol{g}$. Hence, each node has degree $n(2^{n-1}-1)$. In the remaining of step 1, (iii) is shown.

[84] A path of length k is a sequence of k links which connect a sequence of k nodes.

[85] The distance between two nodes is the length of the shortest path connecting them.

Step 1.1 Let $\bar{d}$ be the diameter[86] of the graph $\mathcal{W}$. As stated in Brouwer, Cohen, and Neumaier (1989), (iii) is equivalent to say that there are constants a_i, b_i, c_i such that, $\forall i \in \{0, ..., \bar{d}\}$ and for all nodes $\boldsymbol{g}_1 \in \mathcal{G}, \boldsymbol{g}_2 \in \mathcal{G}$ at distance i, there are c_i neighbours[87] of $\boldsymbol{g}_2$ at distance $i-1$ from $\boldsymbol{g}_1$, b_i neighbours of $\boldsymbol{g}_2$ at distance $i+1$ from $\boldsymbol{g}_1$, and a_i neighbours of $\boldsymbol{g}_2$ at distance i from $\boldsymbol{g}_1$. The next steps show that such constants a_i, b_i, c_i exist.

Step 1.2 It can be observed that the graph $\mathcal{W}$ has diameter n because any two matrices $\boldsymbol{g}_1 \in \mathcal{G}, \boldsymbol{g}_2 \in \mathcal{G}$ can differ for at most n rows. Moreover if two nodes $\boldsymbol{g}_1 \in \mathcal{G}, \boldsymbol{g}_2 \in \mathcal{G}$ are at distance $i \in \{0, ..., n\}$, then they differ for i rows.

Step 1.3 Consider any two nodes $\boldsymbol{g}_1 \in \mathcal{G}, \boldsymbol{g}_2 \in \mathcal{G}$ at distance i. Let the node $\boldsymbol{g}_3 \in \mathcal{G}$ be a neighbour of $\boldsymbol{g}_2$. If $\boldsymbol{g}_3$ is at distance $i-1$ to $\boldsymbol{g}_1$, then the matrix $\boldsymbol{g}_3$ should coincide with the matrix $\boldsymbol{g}_2$ except for one row, among the i rows at which $\boldsymbol{g}_1$ differs from $\boldsymbol{g}_2$, where it is, instead, equal to $\boldsymbol{g}_1$. Hence, the number of admissible matrices $\boldsymbol{g}_3$ is equivalent to the number of rows in which the matrix $\boldsymbol{g}_1$ is different from the matrix $\boldsymbol{g}_2$, i.e., i. Let $c_i := i$.

If $\boldsymbol{g}_3$ is at distance i to $\boldsymbol{g}_1$, then the matrix $\boldsymbol{g}_3$ should coincide with the matrix $\boldsymbol{g}_2$ except for one row, among the i rows at which $\boldsymbol{g}_1$ differs from $\boldsymbol{g}_2$, where it also differs from $\boldsymbol{g}_1$. Hence, the number of admissible matrices $\boldsymbol{g}_3$ is equivalent to the number of rows in which the matrix $\boldsymbol{g}_1$ is different from the matrix $\boldsymbol{g}_2$, i.e., i, times the number of values that a given row can take different from the value of the same rows in $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, i.e., $2^{n-1}-2$. Let $a_i := i(2^{n-1}-2)$.

If $\boldsymbol{g}_3$ is at distance $i+1$ to $\boldsymbol{g}_1$, then the matrix $\boldsymbol{g}_3$ should coincide with the matrix $\boldsymbol{g}_2$ except for one row that is not among the i rows at which $\boldsymbol{g}_1$ differs from $\boldsymbol{g}_2$. Hence, the number of admissible matrices $\boldsymbol{g}_3$ is equivalent to the number of rows in which the matrix $\boldsymbol{g}_1$ is equivalent to the matrix $\boldsymbol{g}_2$, i.e., $n-i$, times the number of values that a given row can take different from the value of the same rows in $\boldsymbol{g}_2$, i.e., $2^{n-1}-1$. Let $b_i := (n-i)(2^{n-1}-1)$.

Therefore, there are constants a_i, b_i, c_i such that, $\forall i \in \{0, ..., \bar{d}\}$ and for all nodes $\boldsymbol{g}_1 \in \mathcal{G}, \boldsymbol{g}_2 \in \mathcal{G}$ at distance i, there are c_i neighbours of $\boldsymbol{g}_2$ at distance $i-1$ from $\boldsymbol{g}_1$,

[86] The diameter is the maximum distance between all pairs of nodes.

[87] A vertex's neighbour is a node at distance 1 from that vertex.

b_i neighbours of $\boldsymbol{g}_2$ at distance $i+1$ from $\boldsymbol{g}_1$, and a_i neighbours of $\boldsymbol{g}_2$ at distance i from $\boldsymbol{g}_1$.

Step 2 The first part of Theorem in Brouwer and Koolen (2009) states that in a distance-regular graph the vertex connectivity, i.e., the minimum number of nodes whose deletion disconnects the graph, equals the vertex degree. This suggests that one can split the graph $\mathcal{W}$ in at least 2 separate components by deleting at least $n(2^{n-1}-1)$ nodes.

Step 3 Consider a set $K \in \mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1,...,|\mathcal{G}|-2\}$. If $1 \leq |K| < n(2^{n-1}-1)$, or, equivalently, if the set $C := \mathcal{G} \setminus K$ is such that $|C| \in \{|\mathcal{G}|-n(2^{n-1}-1),...,|\mathcal{G}|-1\}$, then, by step 2, there exists no non-empty set $D \subset C$ with all the pairs of matrices $\boldsymbol{g}_D \in D, \boldsymbol{g}_{\{C \setminus D\}} \in \{C \setminus D\}$ differing for at least two rows. Indeed, C corresponds to deleting less than $n(2^{n-1}-1)$ vertices from the graph $\mathcal{W}$ which delivers still a connected graph by step 2. Hence, it should be $n(2^{n-1}-1) \leq |K| \leq |\mathcal{G}|-2$, or, equivalently, $2 \leq |C| \leq |\mathcal{G}|-n(2^{n-1}-1)$.

Step 4 The number of sets $K \in \mathcal{K}_{\mathcal{G}}$ with cardinality $|K| \in \{n(2^{n-1}-1),...,|\mathcal{G}|-2\}$ is $\sum_{k=n(2^{n-1}-1)}^{|\mathcal{G}|-2} \binom{2^{n(n-1)}}{k}$.

Hence, combining steps 3 and 4, it can be concluded that the number of sets $K \in \mathcal{K}_{\mathcal{G}}$ with $|K| \in \{1,...,|\mathcal{G}|-2\}$ such that, given the set $C := \mathcal{G} \setminus K$, $\exists$ a non-empty set $D \subset C$ with all pairs of matrices $\boldsymbol{g}_D \in D, \boldsymbol{g}_{\{C \setminus D\}} \in \{C \setminus D\}$ differing for at least two rows is $a \leq \sum_{k=n(2^{n-1}-1)}^{|\mathcal{G}|-2} \binom{2^{n(n-1)}}{k}$.

Step 5 The second part of Theorem in Brouwer and Koolen (2009) states that in a distance-regular graph the only disconnecting sets of vertices with size equal to the vertex degree are the nodes' neighbourhoods[88]. Hence, in the graph $\mathcal{W}$, there are $|\mathcal{G}|$ possible ways to delete $n(2^{n-1}-1)$ vertices, one for each vertex[89]. Therefore, it can be concluded that $a \geq |\mathcal{G}|$.

$\square$

Proof of Proposition 3. The proof is articulated as follows: step 1 shows that, under Assumption 1, $\Theta^{\star\star} \supseteq \Theta^{\star}$ by considering Artsteins' inequalities $\forall K \in \mathcal{A}_{\mathcal{G}}$; step

[88]The neighbourhood of a vertex is the set of its neighbours.

[89]Notice that, since $n > 2$, there is no pair of nodes in $\mathcal{W}$ with the same neighbourhood.

2 shows that, under Assumptions 1 and 2, $\Theta^{\star\star} = \Theta^{\star}$ and it is divided into two sub-steps; step 2.1 shows that, under Assumptions 1 and 2, if Arstein's inequalities involving the outcomes of the local games are satisfied, then Arstein's inequalities $\forall K \in \mathcal{A}_{\mathcal{G}}$ are satisfied; step 2.2 shows that, under Assumptions 1 and 2, if Arstein's inequalities involving the outcomes of the local games are satisfied, then Arstein's inequalities $\forall K \in \mathcal{K}_{\mathcal{G}} \setminus \mathcal{A}_{\mathcal{G}}$ are satisfied.

Step 1 It is shown that, under Assumption 1, $\Theta^{\star\star} \supseteq \Theta^{\star}$. Specifically, it is proved that, under Assumption 1, if a $\theta \in \Theta$ is such that

$$\mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K) \; \forall K \in \mathcal{K}_{\mathcal{G}}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.10}$$

then

$$\mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X},\varepsilon_{\cdot j})|\boldsymbol{X}=\boldsymbol{x}}(K_{\cdot j}) \; \forall K_{\cdot j} \in \mathcal{K}_{\{0,1\}^{n-1}}, \forall j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.11}$$

For any $j \in \mathcal{N}$ and $K_{\cdot j} \in \mathcal{K}_{\{0,1\}^{n-1}}$, take $K \in \mathcal{A}_{\mathcal{G}}$ corresponding to

$$\underbrace{\{0,1\}^{n-1} \times ... \times \{0,1\}^{n-1}}_{j-1 \text{ times}} \times K_{\cdot j} \times \underbrace{\{0,1\}^{n-1} \times ... \times \{0,1\}^{n-1}}_{n-j \text{ times}}$$

Consider a $\theta \in \Theta$ such that (B.10) holds. Hence,

$$\mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K)$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s., which is equivalent, by Lemma 2, to

$$\begin{aligned}
&\mathbb{P}(G_{\cdot 1} \in \{0,1\}^{n-1}, ..., G_{\cdot j-1} \in \{0,1\}^{n-1}, G_{\cdot j} \in K_{\cdot j}, G_{\cdot j+1} \in \{0,1\}^{n-1}, ..., G_{\cdot n} \in \{0,1\}^{n-1} | \boldsymbol{X} = \boldsymbol{x}) \\
&\quad \leq \mathbb{P}(\mathcal{S}_{\theta_u,\cdot 1}(\boldsymbol{X}, \varepsilon_{\cdot 1}) \cap \{0,1\}^{n-1} \neq \emptyset, ..., \mathcal{S}_{\theta_u,\cdot j-1}(\boldsymbol{X}, \varepsilon_{\cdot j-1}) \cap \{0,1\}^{n-1} \neq \emptyset, \mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset, \\
&\qquad \mathcal{S}_{\theta_u,\cdot j+1}(\boldsymbol{X}, \varepsilon_{\cdot j+1}) \cap \{0,1\}^{n-1} \neq \emptyset, ..., \mathcal{S}_{\theta_u,\cdot n}(\boldsymbol{X}, \varepsilon_{\cdot n}) \cap \{0,1\}^{n-1} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x})
\end{aligned}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s., which is equivalent to

$$\mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x})$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. By repeating the same arguments $\forall K_{\cdot j} \in \mathcal{K}_{\{0,1\}^{n-1}}$ and $\forall j \in \mathcal{N}$, all the inequalities in (B.11) are obtained.

Step 2 It is shown that, under Assumptions 1 and 2, $\Theta^{\star\star} = \Theta^{\star}$. As discussed in step 1, under Assumption 1, if $\theta \in \Theta^{\star}$, then $\theta \in \Theta^{\star\star}$. Hence, in what follows it is proved that, under Assumptions 1 and Assumptions 2, if $\theta \in \Theta^{\star\star}$, then $\theta \in \Theta^{\star}$. This is equivalent to show that, if a $\theta \in \Theta$ is such that (B.11) holds, then (B.10) is satisfied.

Step 2.1 It is shown that if a $\theta \in \Theta$ is such that (B.11) holds, then

$$\mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K) \ \forall K \in \mathcal{A}_{\mathcal{G}}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.12}$$

Consider any $K \in \mathcal{A}_{\mathcal{G}}$, which corresponds, by definition of $\mathcal{A}_{\mathcal{G}}$, to some $K_{\cdot 1} \times \ldots \times K_{\cdot n}$ with $K_{\cdot j} \in \mathcal{K}_{\{0,1\}^{n-1}}$ $\forall j \in \mathcal{N}$. Take a $\theta \in \Theta$ such that (B.11) holds. This implies that

$$\begin{cases} \mathbb{P}(G_{\cdot 1} \in K_{\cdot 1} | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u, \cdot 1}(\boldsymbol{X}, \varepsilon_{\cdot 1}) \cap K_{\cdot 1} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \\ \mathbb{P}(G_{\cdot 1} \in K_{\cdot 2} | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u, \cdot 2}(\boldsymbol{X}, \varepsilon_{\cdot 2}) \cap K_{\cdot 2} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \\ \vdots \\ \mathbb{P}(G_{\cdot n} \in K_{\cdot n} | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u, \cdot n}(\boldsymbol{X}, \varepsilon_{\cdot n}) \cap K_{\cdot n} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \end{cases}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. By taking the product, it follows that

$$\prod_{j=1}^{n} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}) \leq \prod_{j=1}^{n} \mathbb{P}(\mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \tag{B.13}$$

$\forall \boldsymbol{x} \in \mathcal{X}$. Moreover, $\forall \theta \in \Theta$, under Assumptions 1 and 2, it holds that

$$\prod_{j=1}^{n} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}; \theta) = \mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}; \theta) \tag{B.14}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. Indeed, the model predicts

$$\prod_{j=1}^{n} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}; \theta) = \prod_{j=1}^{n} \int_{e_{\cdot j} \in \mathbb{R}^{n-1}} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j})) d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon)$$
$$= \int_{e_{\cdot 1} \in \mathbb{R}^{n-1}} \int_{e_{\cdot 2} \in \mathbb{R}^{n-1}} \dots \int_{e_{\cdot n} \in \mathbb{R}^{n-1}} \prod_{j=1}^{n} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j})) d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon)$$
$$= \int_{e_{\cdot 1} \in \mathbb{R}^{n-1}} \int_{e_{\cdot 2} \in \mathbb{R}^{n-1}} \dots \int_{e_{\cdot n} \in \mathbb{R}^{n-1}} \prod_{j=1}^{n} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j})) \prod_{j=1}^{n} d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon)$$
$$\underbrace{=}_{\substack{\text{Ass. 2 (i) and} \\ \text{Fubini's Theorem}}} \int_{e := (e_{\cdot 1}, \dots, e_{\cdot n}) \in \mathbb{R}^{n(n-1)}} \prod_{j=1}^{n} \mathbb{P}(G_{\cdot j} \in K_{\cdot j} | \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j})) dF(e; \theta_\varepsilon)$$
$$\underbrace{=}_{\text{Ass 2 (ii)}} \int_{e := (e_{\cdot 1}, \dots, e_{\cdot n}) \in \mathbb{R}^{n(n-1)}} \mathbb{P}(G_{\cdot 1} \in K_{\cdot 1}, \dots, G_{\cdot n} \in K_{\cdot n} | \mathcal{S}_{\theta_u, \cdot 1}(\boldsymbol{x}, e_{\cdot 1}), \dots, \mathcal{S}_{\theta_u, \cdot n}(\boldsymbol{x}, e_{\cdot n})) dF(e; \theta_\varepsilon)$$
$$\underbrace{=}_{\text{Lemma 2}} \int_{e := (e_{\cdot 1}, \dots, e_{\cdot n}) \in \mathbb{R}^{n(n-1)}} \mathbb{P}(\boldsymbol{G} \in \bigtimes_{j=1}^{n} K_{\cdot j} | \mathcal{S}_{\theta_u}(\boldsymbol{x}, e)) dF(e; \theta_\varepsilon)$$
$$= \mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}; \theta)$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. Furthermore, $\forall \theta \in \Theta$, under Assumptions 1 and 2, it holds that

$$\prod_{j=1}^{n} \mathbb{P}(\mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap K \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \tag{B.15}$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. Indeed, the model predicts

$$\prod_{j=1}^{n} \mathbb{P}(\mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}) \underbrace{=}_{\text{Ass. 1 (ii)}} \prod_{j=1}^{n} \mathbb{P}(\mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset)$$
$$\underbrace{=}_{\text{Ass. 2 (i)}} \mathbb{P}(\mathcal{S}_{\theta_u, \cdot 1}(\boldsymbol{x}, \varepsilon_{\cdot 1}) \cap K_{\cdot 1} \neq \emptyset, \dots, \mathcal{S}_{\theta_u, \cdot n}(\boldsymbol{x}, \varepsilon_{\cdot n}) \cap K_{\cdot n} \neq \emptyset) \underbrace{=}_{\text{Lemma 2}} \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{x}, \varepsilon) \cap K \neq \emptyset)$$
$$\underbrace{=}_{\text{Ass. 1 (ii)}} \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap K \neq \emptyset | \boldsymbol{X} = \boldsymbol{x})$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. Therefore, inserting (B.14) and (B.15) in (B.13) after having replaced the probabilities predicted by the model with their empirical counterparts, under Assumption 1 and 2, it holds that

$$\mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \leq \mathbb{P}(\mathcal{S}_{\theta_u}(\boldsymbol{X}, \varepsilon) \cap K \neq \emptyset | \boldsymbol{X} = \boldsymbol{x})$$

$\forall \boldsymbol{x} \in \mathcal{X}$ a.s. By repeating the same arguments $\forall K \in \mathcal{A}_{\mathcal{G}}$, all the inequalities in (B.12) are obtained.

Step 2.2 It is shown that if a $\theta \in \Theta$ is such that (B.11) holds, then

$$\mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}) \leq T_{\mathcal{S}_{\theta_u}(\boldsymbol{X},\varepsilon)|\boldsymbol{X}=\boldsymbol{x}}(K) \ \forall K \in \mathcal{K}_{\mathcal{G}} \setminus \mathcal{A}_{\mathcal{G}}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.16}$$

where $\mathcal{K}_{\mathcal{G}} \setminus \mathcal{A}_{\mathcal{G}}$ is the collection of sets not included in $\mathcal{A}_{\mathcal{G}}$ obtained by taking the union of elements of $\mathcal{A}_{\mathcal{G}}$.

□

Proof of Proposition 4. The proof is articulated as follows: step 1 shows that, under Assumptions 1 and 2, if $\theta \in \Theta^o$, then $\theta \in \Theta^o_{CT}$; step 2 shows that, under Assumptions 1 and 2, the converse does not necessarily hold; it follows that, under Assumptions 1 and 2, $\Theta^o \subseteq \Theta^o_{CT}$.

Step 1 It is shown that, under Assumptions 1 and 2, if $\theta \in \Theta^o$, then $\theta \in \Theta^o_{CT}$. This comes from step 2.1 in the proof of Proposition 3 with $K_{\cdot j} := \{g_{\cdot j}\}$ and $K_{\cdot j} := \{0,1\}^{n-1} \setminus \{g_{\cdot j}\}$ $\forall g_{\cdot j} \in \{0,1\}^{n-1}$ and $\forall j \in \mathcal{N}$.

Step 2 It is shown that, under Assumptions 1 and 2, if $\theta \in \Theta^o_{CT}$, then it does not necessarily follow that $\theta \in \Theta^o$. Consider any $\theta \in \Theta^o_{CT}$, i.e.,

$$\int_{e \in \mathbb{R}^{n(n-1)} \text{ s.t. } \mathcal{S}_{\theta_u}(\boldsymbol{x},e)=\{\boldsymbol{g}\}} dF(e;\theta_\varepsilon) \leq \mathbb{P}(\boldsymbol{G} = \boldsymbol{g} | \boldsymbol{X} = \boldsymbol{x}) \leq \int_{e \in \mathbb{R}^{n(n-1)} \text{ s.t. } \boldsymbol{g} \in \mathcal{S}_{\theta_u}(\boldsymbol{x},e)} dF(e;\theta_\varepsilon) \ \forall \boldsymbol{g} \in \mathcal{G}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.17}$$

Let $\boldsymbol{G}_{-\{\cdot j\}}$ be the matrix $\boldsymbol{G}$ with jth column deleted. By integrating out $\boldsymbol{G}_{-\{\cdot j\}}$, (B.17) implies that

$$\sum_{\boldsymbol{g}_{-\{\cdot j\}}} \int_{e \in \mathbb{R}^{n(n-1)} \text{ s.t. } \mathcal{S}_{\theta_u}(\boldsymbol{x},e)=\{\boldsymbol{g}\}} dF(e;\theta_\varepsilon) \leq \mathbb{P}(G_{\cdot j} = \boldsymbol{g}_{\cdot j} | \boldsymbol{X} = \boldsymbol{x}) \leq \sum_{\boldsymbol{g}_{-\{\cdot j\}}} \int_{e \in \mathbb{R}^{n(n-1)} \text{ s.t. } \boldsymbol{g} \in \mathcal{S}_{\theta_u}(\boldsymbol{x},e)} dF(e;\theta_\varepsilon)$$
$$\forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.18}$$

which is equivalent to

$$\mathbb{P}(\exists \boldsymbol{g}_{-\{\cdot j\}} \text{ s.t. } (g_{\cdot j}, \boldsymbol{g}_{-\{\cdot j\}}) \text{ is the unique equilibrium}; \boldsymbol{\theta}) \leq \mathbb{P}(G_{\cdot j} \in \{\boldsymbol{g}_{\cdot j}\} | \boldsymbol{X} = \boldsymbol{x})$$
$$\leq \sum_{\boldsymbol{g}_{-\{\cdot j\}}} \mathbb{P}((g_{\cdot j}, \boldsymbol{g}_{-\{\cdot j\}}) \text{ is an equilibrium}; \boldsymbol{\theta}) \ \forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \text{ a.s.} \tag{B.19}$$

It can be noticed that

$$\mathbb{P}(\exists \boldsymbol{g}_{-\{\cdot j\}} \text{ s.t. } (g_{\cdot j}, \boldsymbol{g}_{-\{\cdot j\}}) \text{ is the unique equilibrium}; \boldsymbol{\theta}) \leq \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j}) = \{g_{\cdot j}\}} d\tilde{F}_{\cdot j}(e_{\cdot j}; \boldsymbol{\theta}_{\varepsilon}) \tag{B.20}$$

and

$$\sum_{\boldsymbol{g}_{-\{\cdot j\}}} \mathbb{P}((g_{\cdot j}, \boldsymbol{g}_{-\{\cdot j\}}) \text{ is an equilibrium}; \boldsymbol{\theta}) \geq \mathbb{P}(\exists \boldsymbol{g}_{-\{\cdot j\}} \text{ s.t. } (g_{\cdot j}, \boldsymbol{g}_{-\{\cdot j\}}) \text{ is an equilibrium}; \boldsymbol{\theta})$$
$$\underbrace{=}_{\text{Lemma 2}} \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } g_{\cdot j} \in \mathcal{S}_{\theta_u, \cdot j}(\boldsymbol{x}, e_{\cdot j})} d\tilde{F}_{\cdot j}(e_{\cdot j}; \boldsymbol{\theta}_{\varepsilon}) \tag{B.21}$$

$\forall g_{\cdot j} \in \{0,1\}^{n-1}$, $\forall j \in \mathcal{N}$, and $\forall \boldsymbol{x} \in \mathcal{X}$, i.e., Θ^o_{CT} implies wider bounds for $\mathbb{P}(G_{\cdot j} = \boldsymbol{g}_{\cdot j} | \boldsymbol{X} = \boldsymbol{x})$ than those imposed by Θ^o $\forall g_{\cdot j} \in \{0,1\}^{n-1}$, $\forall j \in \mathcal{N}$, and $\forall \boldsymbol{x} \in \mathcal{X}$ a.s. Moreover, Assumption 2 does not help to refine expressions (B.20) and (B.21). Hence, $\theta \in \Theta^o_{CT}$ does not signify that $\theta \in \Theta^o$. □

Appendix C

Inference when $\mathcal{X}$ is finite

This section discusses how to construct a 95% confidence region for the outer set Θ^o when $\mathcal{X}$ is finite and observations are i.i.d., following the method of Andrews and Soares (2010). In order to simplify the exposition and without loss of generality, in the remaining of the section it is assumed that N is a degenerate random variable with support $\{n\}$, for $n \in \mathbb{N} \setminus \{1,2\}$.

From (2.19) combined with (2.20),

$$\Theta^o = \left\{ \theta \in \Theta | \tilde{H}^l_{g_{\cdot j},\boldsymbol{x}}(\theta) \leq \mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x}) \leq \tilde{H}^u_{g_{\cdot j},\boldsymbol{x}}(\theta) \ \forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \right\} \tag{C.1}$$

where

$$\tilde{H}^l_{g_{\cdot j},\boldsymbol{x}}(\theta) := \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } \mathcal{S}_{\cdot j,\theta_u}(\boldsymbol{x}, e_{\cdot j}) = \{g_{\cdot j}\}} d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon) \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \tag{C.2}$$

and

$$\tilde{H}^u_{g_{\cdot j},\boldsymbol{x}}(\theta) := \int_{e_{\cdot j} \in \mathbb{R}^{n-1} \text{ s.t. } g_{\cdot j} \in \mathcal{S}_{\cdot j,\theta_u}(\boldsymbol{x}, e_{\cdot j})} d\tilde{F}_{\cdot j}(e_{\cdot j}; \theta_\varepsilon) \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \tag{C.3}$$

A preliminary step needed to conduct inference on Θ^o is estimation of $\mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x})$ $\forall g_{\cdot j} \in \{0,1\}^{n-1}$, $\forall j \in \mathcal{N}$, and $\forall \boldsymbol{x} \in \mathcal{X}$. Moreover, the inference algorithm requires computation of $\tilde{H}^l_{g_{\cdot j},\boldsymbol{x}}(\theta)$ and $\tilde{H}^u_{g_{\cdot j},\boldsymbol{x}}(\theta)$ $\forall g_{\cdot j} \in \{0,1\}^{n-1}$, $\forall j \in \mathcal{N}$, $\forall \boldsymbol{x} \in \mathcal{X}$, and $\forall \theta \in \Theta$.

Estimating $\mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x})$, for example via a frequency estimator, is complicated by the fact that, under Assumption 1 (i), players' identities vary across networks. Indeed, when $\mathcal{X}$ is finite, within each network, there may be observationally identical players which could be labelled arbitrarily by the researcher, hence producing different estimates of $\mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x})$[90]. A similar problem arises when computing $\tilde{H}^l_{g_{\cdot j}, \boldsymbol{x}}(\theta)$ and $\tilde{H}^u_{g_{\cdot j}, \boldsymbol{x}}(\theta)$. To solve this issue one possibility is adopting the strategy proposed by Sheng (2016) that relies on joint exchangeability of networks[91], as to ensure that labelling loses any relevance. The procedure is described below in 4 steps.

Step 1 This step imposes sufficient conditions for joint exchangeability of networks as in Sheng (2016).

Assumption 3. (Exchangeability)

(i) The finite sequence of random variables $(\varepsilon_{ij})_{\forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j}$ is jointly exchangeable.

(ii) The finite sequence of random variables $(X_i)_{\forall i \in \mathcal{N}}$ is exchangeable[92].

(iii) The equilibrium selection mechanisms adopted by players in the network formation game is independent of players' labels, i.e., for every permutation φ of the labels in $\mathcal{N}$, $\forall K \in \mathcal{K}_{\mathcal{G}}$, and $\forall \theta \in \Theta$,

$$\mathbb{P}(\boldsymbol{G} \in K | \mathcal{S}_{\theta_u}(\boldsymbol{x}, e)) = \mathbb{P}(\boldsymbol{G} \in K^{\varphi} | \mathcal{S}_{\theta_u}(\boldsymbol{x}^{\varphi}, e^{\varphi}))$$

$\forall \boldsymbol{x} \in \mathcal{X}$ and $\forall e \in \mathbb{R}^{n(n-1)}$, where K^{φ} and e^{φ} are obtained by applying φ respectively to K and e.

•

Remark 8. (Discussion on Assumption 3) Assumption 3 restricts the correlation among players' characteristics and the set of admissible equilibrium selection mechanisms for the network formation game.

More specifically, Assumption 3 (i) is satisfied if, $\forall i \in \mathcal{N}$ and $\forall j \in \mathcal{N}$ with $i \neq j$, $\varepsilon_{ij} = \alpha_i + \beta_j + \xi_{ij}$, where $\Big((\alpha_i)_{\forall i \in \mathcal{N}}, (\beta_j)_{\forall j \in \mathcal{N}}, (\xi_{ij})_{\forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j}\Big)$ are i.i.d.

[90]When players' observed characteristics are continuous this problem does not arise, as observing identical players is a zero probability event.

[91]Joint exchangeability is defined e.g., by Kallengberg (2005).

[92]Exchangeability is defined e.g., by Schervish (1995).

(random effects across i and j). Instead, Assumption 3 (i) is violated if the random variables forming the sequence $(\varepsilon_{ij})_{\forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j}$ are identically distributed across ij but their joint probability distribution is not invariant to labelling. Similar considerations can be made for Assumption 3 (ii).

Assumption 3 (iii) excludes that, in case the network formation game admits multiple equilibria, players coordinate on a specific outcome in the equilibrium set considering agents' labels. For example, Assumption 3 (ii) is satisfied if the equilibrium selection rule of the network formation game assigns a uniform probability distribution over the outcomes in the equilibrium set. Assumption 3 (ii) is also met if players select the outcome providing the highest total payoff from the equilibrium set. Instead, Assumption 3 (ii) is violated if players choose the outcome generating the highest payoff for agent 1 from the equilibrium set .

Assumptions 1 and 3 imply the testable prediction that networks are jointly exchangeable[93]. Indeed, for every permutation φ of the labels in $\mathcal{N}$, $\forall K \in \mathcal{K}_{\mathcal{G}}$, $\forall \boldsymbol{x} \in \mathcal{X}$, and $\forall \theta \in \Theta$, under Assumptions 1 and 3, the model states that

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{G} \in K, \boldsymbol{X} = \boldsymbol{x}; \theta) &= \mathbb{P}(\boldsymbol{G} \in K | \boldsymbol{X} = \boldsymbol{x}; \theta)\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \int_{e \in \mathbb{R}^{n(n-1)}} \mathbb{P}(\boldsymbol{G} \in K | \mathcal{S}_{\theta_u}(\boldsymbol{x}, e)) dF(e; \theta_\varepsilon)\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \\
&= \int_{e^\varphi \in \mathbb{R}^{n(n-1)}} \underbrace{\mathbb{P}(\boldsymbol{G} \in K^\varphi | \mathcal{S}_{\theta_u}(\boldsymbol{x}^{\boldsymbol{\varphi}}, e^\varphi))}_{\text{Ass. 3 (iii)}} \underbrace{dF(e^\varphi; \theta_\varepsilon)}_{\text{Ass. 3 (i)}} \underbrace{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}^\varphi)}_{\text{Ass. 3 (ii)}} \\
&= \mathbb{P}(\boldsymbol{G} \in K^\varphi | \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}}; \theta)\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}^\varphi) = \mathbb{P}(\boldsymbol{G} \in K^\varphi, \boldsymbol{X} = \boldsymbol{x}^\varphi; \theta)
\end{aligned}
$$

Lastly, under Assumptions 1 and 3, also the *section* 1, ..., *section n* are jointly exchangeable, i.e., $\forall j \in \mathcal{N}$, for every permutation φ of the labels in $\mathcal{N}$, $\forall K_{\cdot j} \in \mathcal{K}_{\{0,1\}^{n-1}}$, $\forall \boldsymbol{x} \in \mathcal{X}$, and $\forall \theta \in \Theta$, under Assumptions 1 and 3, the model predicts

$$
\mathbb{P}(G_{\cdot j} \in K_{\cdot j}, \boldsymbol{X} = \boldsymbol{x}; \theta) = \mathbb{P}(G_{\cdot \varphi(j)} \in K^\varphi_{\cdot \varphi(j)}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}}; \theta) \tag{C.4}
$$

where $K^\varphi_{\cdot \varphi(j)}$ is obtained by applying φ to $K_{\cdot j}$[94].

[93]This, in turn, suggests that, under Assumptions 1 and 3, networks are dense by Aldous-Hoover Representation Theorem (Orbanz and Roy, 2015).

[94]Notice that, under Assumption 3 (i), $\forall j \in \mathcal{N}$, the finite sequence of random variables $(\varepsilon_{ij})_{\forall i \in \mathcal{N}, i \neq j}$ is jointly exchangeable, because any finite subsequence of the finite sequence of random variables $(\varepsilon_{ij})_{\forall i \in \mathcal{N}, \forall j \in \mathcal{N}, i \neq j}$ is jointly exchangeable (Proposition 1.12 in Schervish, 1995)

$\triangle$

Step 2 This step shows that, under Assumptions 1 and 3, it is sufficient to focus on the *section j game* for a $j \in \mathcal{N}$, rather than $\forall j \in \mathcal{N}$. Specifically, it is proved that, under Assumptions 1 and 3, $\Theta^o = \Theta^o_{\cdot j}\ \forall j \in \mathcal{N}$, where

$$\Theta^o_{\cdot j} := \left\{\theta \in \Theta | \tilde{H}^l_{g_{\cdot j},\boldsymbol{x}}(\theta) \le \mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x}) \le \tilde{H}^u_{g_{\cdot j},\boldsymbol{x}}(\theta)\ \forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall \boldsymbol{x} \in \mathcal{X}\right\} \tag{C.5}$$

Proof. It can be seen that if $\theta \in \Theta^o$, then $\theta \in \Theta^o_{\cdot j}\ \forall j \in \mathcal{N}$. Hence, in what follows it is proved that, $\forall j \in \mathcal{N}$, if $\theta \in \Theta^o_{\cdot j}$, then $\theta \in \Theta^o$. This is the same as showing that $\forall \theta \in \Theta$ and $\forall j \in \mathcal{N}$, if

$$\tilde{H}^l_{g_{\cdot j},\boldsymbol{x}}(\theta) \le \mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x}) \le \tilde{H}^u_{g_{\cdot j},\boldsymbol{x}}(\theta)\ \forall g_{\cdot j} \in \{0,1\}^{n-1}, \forall \boldsymbol{x} \in \mathcal{X} \tag{C.6}$$

then

$$\tilde{H}^l_{g_{\cdot h},\boldsymbol{x}}(\theta) \le \mathbb{P}(G_{\cdot h} = g_{\cdot h}, \boldsymbol{X} = \boldsymbol{x}) \le \tilde{H}^u_{g_{\cdot h},\boldsymbol{x}}(\theta)\ \forall g_{\cdot h} \in \{0,1\}^{n-1}, \forall h \neq j \in \mathcal{N}, \forall \boldsymbol{x} \in \mathcal{X} \tag{C.7}$$

Firstly, $\forall \theta \in \Theta$, $\forall j \in \mathcal{N}$, $\forall g_{\cdot j} \in \{0,1\}^{n-1}$, $\forall \boldsymbol{x} \in \mathcal{X}$, and for every permutation φ of the labels in $\mathcal{N}$ such that $\varphi(j) \neq j$, under Assumptions 1 and 3, it holds that

$$\mathbb{P}(\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset, \boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}(\mathcal{S}_{\theta_u,\cdot \varphi(j)}(\boldsymbol{X}, \varepsilon_{\cdot \varphi(j)}) \cap K^{\varphi}_{\cdot \varphi(j)} \neq \emptyset, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}}) \tag{C.8}$$

when $K_{\cdot j} := \{g_{\cdot j}\}$ and $K_{\cdot j} := \{0,1\}^{n-1} \setminus \{g_{\cdot j}\}$. Indeed,

$$\begin{aligned}
&\mathbb{P}(\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset, \boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}(\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{X}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x})\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) \\
&\underbrace{=}_{\text{Ass. 1 (ii)}} \mathbb{P}(\mathcal{S}_{\theta_u,\cdot j}(\boldsymbol{x}, \varepsilon_{\cdot j}) \cap K_{\cdot j} \neq \emptyset)\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \underbrace{\mathbb{P}(\mathcal{S}_{\theta_u,\cdot \varphi(j)}(\boldsymbol{x}^{\boldsymbol{\varphi}}, \varepsilon_{\cdot \varphi(j)}) \cap K^{\varphi}_{\cdot \varphi(j)} \neq \emptyset)}_{\text{Ass. 1 (i), Ass. 3 (i)}}\underbrace{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}})}_{\text{Ass. 3 (ii)}} \\
&\underbrace{=}_{\text{Ass. 1 (ii)}} \mathbb{P}(\mathcal{S}_{\theta_u,\cdot \varphi(j)}(\boldsymbol{X}, \varepsilon_{\cdot \varphi(j)}) \cap K^{\varphi}_{\cdot \varphi(j)} \neq \emptyset | \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}})\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}}) = \mathbb{P}(\mathcal{S}_{\theta_u,\cdot \varphi(j)}(\boldsymbol{X}, \varepsilon_{\cdot \varphi(j)}) \cap K^{\varphi}_{\cdot \varphi(j)} \neq \emptyset, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}})
\end{aligned}$$

Therefore, combining (C.4) with (C.8) $\forall \theta \in \Theta$, $\forall j \in \mathcal{N}$, $\forall K_{\cdot j} \in \mathcal{K}_{\{0,1\}^{n-1}}$, and $\forall$ permutation of labels φ such that $\varphi(j) \neq j$, and replacing the probabilities predicted by the model with their empirical counterparts, it holds that (C.6) is equivalent to (C.7), under Assumptions 1 and 3. $\square$

Step 3 Moreover, under Assumptions 1 and 3, the inequalities in (C.5) indexed by realisations of $G_{\cdot j}, \boldsymbol{X}$ that are equivalent up to a permutation of the labels in $\mathcal{N}$ other than label j are identical.

Proof. It is now shown that, some inequalities in (C.6) are redundant. This comes from the fact that, by (C.4) and (C.8), for every permutation φ of the labels in $\mathcal{N}$ such that $\varphi(j) = j$,

$$\tilde{H}^l_{g_{\cdot j}, \boldsymbol{x}}(\theta) \leq \mathbb{P}(G_{\cdot j} = g_{\cdot j}, \boldsymbol{X} = \boldsymbol{x}) \leq \tilde{H}^u_{g_{\cdot j}, \boldsymbol{x}}(\theta)$$

is equivalent to

$$\tilde{H}^l_{g^{\varphi}_{\cdot j}, \boldsymbol{x}^{\boldsymbol{\varphi}}}(\theta) \leq \mathbb{P}(G_{\cdot j} = g^{\varphi}_{\cdot j}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}}) \leq \tilde{H}^u_{g^{\varphi}_{\cdot j}, \boldsymbol{x}^{\boldsymbol{\varphi}}}(\theta)$$

$\forall \theta \in \Theta$, $\forall g_{\cdot j} \in \{0,1\}^{n-1}$, and $\forall \boldsymbol{x} \in \mathcal{X}$. $\square$

Step 4 By steps 2 and 3, under Assumptions 1 and 3, conducting inference on Θ^o is equivalent to conducting inference on

$$\Theta^o_{\cdot 3} = \left\{ \theta \in \Theta | \tilde{H}^l_{g_{\cdot 3}, \boldsymbol{x}}(\theta) \leq \mathbb{P}(G_{\cdot 3} = g_{\cdot 3}, \boldsymbol{X} = \boldsymbol{x}) \leq \tilde{H}^u_{g_{\cdot 3}, \boldsymbol{x}}(\theta) \ \forall (g_{\cdot 3}, \boldsymbol{x}) \in \mathcal{W} \right\} \tag{C.9}$$

where the subscript j is fixed to 3 without loss of generality, and $\mathcal{W} \subseteq \{0,1\}^{n-1} \times \mathcal{X}$ denotes the collection of realisations of $(G_{\cdot 3}, \boldsymbol{X})$ left over after having deleted those generating redundant inequalities when applying all the permutations φ of the labels in $\mathcal{N}$ with $\varphi(3) = 3$ as explained in step 3. Appendix C.1 illustrates an algorithm to construct $\mathcal{W}$. It should be noticed that the set $\mathcal{W}$ is not unique because one is free to keep any of the realisations of $(G_{\cdot 3}, \boldsymbol{X})$ producing identical inequalities.

Let $C_{g_{\cdot 3}, \boldsymbol{x}} \subset \{0,1\}^{n-1} \times \mathcal{X}$ be the collection of realisations of $(G_{\cdot 3}, \boldsymbol{X})$ giving rise to inequalities identical to the inequalities indexed by $(g_{\cdot 3}, \boldsymbol{x})$ when applying all the permutations φ of the labels in $\mathcal{N}$ with $\varphi(3) = 3$[95]. Sheng (2016) observes that, under Assumptions 1 and 3, (C.9) can be rewritten as

$$\Theta^o_{\cdot 3} = \Big\{ \theta \in \Theta | \tilde{H}^l_{C_{g\cdot 3,\boldsymbol{x}}}(\theta) \leq \mathbb{P}((G_{\cdot 3}, \boldsymbol{X}) \in C_{g\cdot 3,\boldsymbol{x}}) \leq \tilde{H}^u_{C_{g\cdot 3,\boldsymbol{x}}}(\theta) \; \forall (g_{\cdot 3}, \boldsymbol{x}) \in \mathcal{W} \Big\} \tag{C.10}$$

where $\tilde{H}^l_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)$ is the probability that every PSNE of the *section* 3 *game* combined with $\boldsymbol{X}$ falls in $C_{g\cdot 3,\boldsymbol{x}}$ and $\tilde{H}^u_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)$ is the probability that at least one PSNE of the *section* 3 *game* combined with $\boldsymbol{X}$ falls in $C_{g\cdot 3,\boldsymbol{x}}$, given $\theta \in \Theta$. A proof of the equivalence between (C.9) and (C.10) is given in Appendix C.2.

It can be noticed that (C.10) is a convenient way of rewriting $\Theta^o_{\cdot 3}$ as estimates of $\mathbb{P}((G_{\cdot 3}, \boldsymbol{X}) \in C_{g\cdot 3,\boldsymbol{x}})$ do not depend on how players are labelled by the researcher. Similarly, the computation of $\tilde{H}^l_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)$ and $\tilde{H}^u_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)$ is not affected by assigned labels.

Let $\hat{\mathbb{P}}_{C_{g\cdot 3,\boldsymbol{x}}}$ denote an unbiased estimator of $\mathbb{P}((G_{\cdot 3}, \boldsymbol{X}) \in C_{g\cdot 3,\boldsymbol{x}})$. Appendix C.3 describes a procedure to estimate $\mathbb{P}((G_{\cdot 3}, \boldsymbol{X}) \in C_{g\cdot 3,\boldsymbol{x}})$. Moreover, Appendix C.4 provides a procedure to compute $\tilde{H}^l_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)$ and $\tilde{H}^u_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)$.

By unbiasedness of $\hat{\mathbb{P}}_{C_{g\cdot 3,\boldsymbol{x}}}$,

$$\Theta^o_{\cdot 3} = \Big\{ \theta \in \Theta | E(\hat{\mathbb{P}}_{C_{g\cdot 3,\boldsymbol{x}}} - \tilde{H}^l_{C_{g\cdot 3,\boldsymbol{x}}}(\theta)) \geq 0, E(\tilde{H}^u_{C_{g\cdot 3,\boldsymbol{x}}}(\theta) - \hat{\mathbb{P}}_{C_{g\cdot 3,\boldsymbol{x}}}) \geq 0 \; \forall (g_{\cdot 3}, \boldsymbol{x}) \in \mathcal{W} \Big\} \tag{C.11}$$

Reintroducing the subscript m and collecting the lhs of all inequalities in $E[b_m(\theta)]$,

$$\Theta^o_{\cdot 3} = \Big\{ \theta \in \Theta | E[b_m(\theta)] \geq 0 \Big\} \tag{C.12}$$

Assumption 1, combined with an i.i.d. sampling scheme and other regularity conditions listed by Andrews and Soares (2010), imply that a valid $(1-\alpha)\%$ confidence region for each $\theta \in \Theta^o_{\cdot 3}$ can be constructed as follows. Let $\bar{b}_M(\theta) :=$

[95] In network theory, all realisations of $(G_{\cdot 3}, \boldsymbol{X})$ in $C_{g\cdot 3,\boldsymbol{x}}$ are called *isomorphic* and $C_{g\cdot 3,\boldsymbol{x}}$ is an *equivalence class* for $(G_{\cdot 3}, \boldsymbol{X})$.

$\frac{1}{M}\sum_{m=1}^{M} b_m(\theta)$ and $\bar{b}_{k,M}(\theta)$ denote its kth element. Let

$$S_M(\theta) := \sum_k \left(\min\left\{\frac{\sqrt{M}\bar{b}_{k,M}(\theta)}{\tilde{\sigma}_{k,M}(\theta)}, 0\right\}\right)^2$$

where $\tilde{\sigma}_{k,M}(\theta)$ is a consistent estimator of the asymptotic standard deviation of $\sqrt{M}\bar{b}_{k,M}(\theta)$. A $(1-\alpha)\%$ confidence region for each $\theta \in \Theta^o_{\cdot 3}$ is

$$CS_M := \left\{\theta \in \Theta — S_M(\theta) \leq \hat{c}_{M,1-\alpha}(\theta)\right\} \tag{C.13}$$

where $\hat{c}_{M,1-\alpha}(\theta)$ is an estimate of the $1-\alpha$ quantile of the asymptotic probability distribution of $S_M(\theta)$, obtainable following the bootstrap procedure with hard threshold in Andrews and Soares (2010). More details on the construction of $S_M(\theta)$ and $\hat{c}_{M,1-\alpha}(\theta)$ are in Appendix C.5.

C.1 Construction of $\mathcal{W}$

This section illustrates a way to construct the set $\mathcal{W}$.

1. Rewrite each realisation $(g_{\cdot 3}, \boldsymbol{x})$ of $(G_{\cdot 3}, \boldsymbol{X}) \in \{0,1\}^{n-1} \times \mathcal{X}$ by listing in a row vector

   (i) x_3;

   (ii) g_{i3} $\forall i \neq 3 \in \mathcal{N}$ such that $g_{i3} = 1$, disposing them with respect to x_i in ascending order; if there are $i,k \in \mathcal{N}$ with $i \neq k \neq 3$ such that $g_{i3} = g_{k3} = 1$ and $x_i = x_k$, any order is allowed;

   (iii) g_{i3} $\forall i \neq 3 \in \mathcal{N}$ such that $g_{i3} = 0$, disposing them with respect to x_i in ascending order; if there are $i,k \in \mathcal{N}$ with $i \neq k \neq 3$ such that $g_{i3} = g_{k3} = 0$ and $x_i = x_k$, any order is allowed;

   (iv) x_i $\forall i \neq 3 \in \mathcal{N}$ according to the disposition of players adopted in the previous steps.

2. For each row that is repeated once or more, delete all duplications from the second.

3. Collect the saved rows and rearrange each of them in its original order. The resulting set is $\mathcal{W}$.

As an example, assume $n = 3$ and $\mathcal{X} := \{0,1\}$. The set $\{0,1\}^2 \times \mathcal{X}$ is represented in Table C.1. The realisations of $(G_{\cdot 3}, \boldsymbol{X})$ giving rise to equivalent inequalities have a symbol of the same colour. Table C.2 reports in blue the rows of Table C.1 reordered according to step 1. above. It can be noticed that the algorithm described in step 1. detects all the realisations of $(G_{\cdot 3}, \boldsymbol{X})$ associated to the same colour in Table C.1. Lastly, Table C.3 lists the elements of the set $\mathcal{W}$.

Table C.1: Representation of $\{0,1\}^2 \times \mathcal{X}$.

G_{13}	G_{23}	X_1	X_2	X_3
1	1	1	1	1
1	1	1	0	1
1	1	1	1	0
1	1	1	0	0
1	1	0	1	1
1	1	0	0	1
1	1	0	1	0
1	1	0	0	0
1	0	1	1	1
1	0	1	0	1
1	0	1	1	0
0	0	1	0	0
1	0	0	1	1
1	0	0	0	1
1	0	0	1	0
1	0	0	0	0
0	1	1	1	1
0	1	1	0	1
0	1	1	1	0
0	1	1	0	0
0	1	0	1	1
0	1	0	0	1
1	0	0	0	1
0	1	0	0	0
0	0	1	1	1
0	0	1	0	1
0	0	1	1	0
0	0	1	0	0
0	0	0	1	1
0	0	0	0	1
0	0	0	1	0
0	0	0	0	0

Table C.2: Reordering the rows of Table C.1.

G_{13}	G_{23}	X_1	X_2	X_3					
1	1	1	1	1	1	1	1	1	1
1	1	1	0	1	1	1	1	1	0
1	1	1	1	0	0	1	1	1	1
1	1	1	0	0	0	1	1	1	0
1	1	0	1	1	1	1	1	1	0
1	1	0	0	1	1	1	1	0	0
1	1	0	1	0	0	1	1	1	0
1	1	0	0	0	0	1	1	0	0
1	0	1	1	1	1	1	0	1	1
1	0	1	0	1	1	1	0	1	0
1	0	1	1	0	0	1	0	1	1
1	0	1	0	0	0	1	0	1	0
1	0	0	1	1	1	1	0	0	1
1	0	0	0	1	1	1	0	0	0
1	0	0	1	0	0	1	0	0	1
1	0	0	0	0	0	1	0	0	0
0	1	1	1	1	1	1	0	1	1
0	1	1	0	1	1	1	0	0	1
0	1	1	1	0	0	1	0	1	1
0	1	1	0	0	0	1	0	0	1
0	1	0	1	1	1	1	0	1	0
0	1	0	0	1	1	1	0	0	0
0	1	0	1	0	0	1	0	1	0
0	1	0	0	0	0	1	0	0	0
0	0	1	1	1	1	0	0	1	1
0	0	1	0	1	1	0	0	1	0
0	0	1	1	0	0	0	0	1	1
0	0	1	0	0	0	0	0	1	0
0	0	0	1	1	1	0	0	1	0
0	0	0	0	1	1	0	0	0	0
0	0	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0

Table C.3: Representation of $\mathcal{W}$.

G_{13}	G_{23}	X_1	X_2	X_3
1	1	1	1	1
1	1	1	1	0
0	1	1	1	1
0	1	1	1	0
1	1	1	0	0
0	1	1	0	0
1	1	0	1	1
1	1	0	1	0
0	1	0	1	1
0	1	0	1	0
1	1	0	0	1
1	1	0	0	0
0	1	0	0	1
0	1	0	0	0
1	0	0	1	1
1	0	0	1	0
0	0	0	1	1
0	0	0	1	0
1	0	0	0	0
0	0	0	0	0

C.2 Equivalence between (C.9) and (C.10)

As seen in (C.4) and after having replaced the probabilities predicted by the model with their empirical counterparts, under Assumptions 1 and 3,

$$\mathbb{P}(G_{\cdot i} = g_{\cdot i}, \boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}(G_{\cdot\varphi(i)} = g^{\varphi}_{\cdot\varphi(i)}, \boldsymbol{X} = \boldsymbol{x}^{\boldsymbol{\varphi}}) \tag{C.14}$$

$\forall i \in \mathcal{N}$, for every permutation φ of the labels in $\mathcal{N}$, and $\forall (g_{\cdot i}, \boldsymbol{x}) \in \{0,1\}^{n-1} \times \mathcal{X}$.

By (C.14) applied for every permutation φ of the labels in $\mathcal{N}$ such that $\varphi(3) = 3$,

$$\mathbb{P}((G_{\cdot 3}, \boldsymbol{X}) \in C_{g_{\cdot 3},\boldsymbol{x}}) = |C_{g_{\cdot 3},\boldsymbol{x}}| \times \mathbb{P}(G_{\cdot 3} = g_{\cdot 3}, \boldsymbol{X} = \boldsymbol{x}) \tag{C.15}$$

In a similar way, $\tilde{H}^{l}_{C_{g_{\cdot 3},\boldsymbol{x}}}(\theta)$ and $\tilde{H}^{u}_{C_{g_{\cdot 3},\boldsymbol{x}}}(\theta)$ can be shown being, respectively, equivalent to $|C_{g_{\cdot 3},\boldsymbol{x}}| \times \tilde{H}^{l}_{g_{\cdot 3},\boldsymbol{x}}$ and $|C_{g_{\cdot 3},\boldsymbol{x}}| \times \tilde{H}^{u}_{g_{\cdot 3},\boldsymbol{x}}$.

Hence,

$$\left\{\theta \in \Theta | \tilde{H}^{l}_{C_{g_{\cdot 3},\boldsymbol{x}}}(\theta) \leq \mathbb{P}((G_{\cdot 3}, \boldsymbol{X}) \in C_{g_{\cdot 3},\boldsymbol{x}}) \leq \tilde{H}^{u}_{C_{g_{\cdot 3},\boldsymbol{x}}}(\theta)\ \forall (g_{\cdot 3}, \boldsymbol{x}) \in \mathcal{W}\right\}$$
$$\underbrace{=}_{\text{(C.15)}} \left\{\theta \in \Theta || C_{g_{\cdot 3},\boldsymbol{x}}| \times \tilde{H}^{l}_{g_{\cdot 3},\boldsymbol{x}}(\theta) \leq |C_{g_{\cdot 3},\boldsymbol{x}}| \times \mathbb{P}(G_{\cdot 3} = g_{\cdot 3}, \boldsymbol{X} = \boldsymbol{x}) \leq |C_{g_{\cdot 3},\boldsymbol{x}}| \times \tilde{H}^{u}_{g_{\cdot 3},\boldsymbol{x}}(\theta) \forall (g_{\cdot 3}, \boldsymbol{x}) \in \mathcal{W}\right\} =$$

C.3 Computation of $\hat{\mathbb{P}}_{C_{g_{\cdot 3},\boldsymbol{x}}}$

Consider any $i \in \mathcal{N}$ and $(\tilde{g}_{\cdot i}, \tilde{\boldsymbol{x}}) \in \{0,1\}^{n-1} \times \mathcal{X}$ such that $\exists$ a permutation φ of the labels in $\mathcal{N}$ with $\varphi(i) = 3$ generating $(\tilde{g}^{\varphi}_{\cdot\varphi(i)}, \tilde{\boldsymbol{x}}^{\varphi}) = (g_{\cdot 3}, \boldsymbol{x})$. By (C.14),

$$\mathbb{P}(G_{\cdot i} = \tilde{g}_{\cdot i}, \boldsymbol{X} = \tilde{\boldsymbol{x}}) = \mathbb{P}(G_{\cdot 3} = g_{\cdot 3}, \boldsymbol{X} = \boldsymbol{x}) \tag{C.16}$$

Consider $C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}} \subseteq \{0,1\}^{n-1} \times \mathcal{X}$. By (C.14) applied for every permutation φ of the labels in $\mathcal{N}$ with $\varphi(i) = i$,

$$\mathbb{P}((G_{\cdot i}, \boldsymbol{X}) \in C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}) = |C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}| \times \mathbb{P}(G_{\cdot i} = \tilde{g}_{\cdot i}, \boldsymbol{X} = \tilde{\boldsymbol{x}}) \tag{C.17}$$

Hence,

$$\mathbb{P}((G_{\cdot i},\boldsymbol{X}) \in C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}) \underbrace{=}_{(C.17)} |C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}| \times \mathbb{P}(G_{\cdot i} = \tilde{g}_{\cdot i},\boldsymbol{X} = \tilde{\boldsymbol{x}}) \underbrace{=}_{(C.16)} |C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}| \times \mathbb{P}(G_{\cdot 3} = g_{\cdot 3},\boldsymbol{X} = \boldsymbol{x})$$
$$\underbrace{=}_{|C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}|=|C_{g_{\cdot 3},\boldsymbol{x}}|} |C_{g_{\cdot 3},\boldsymbol{x}}| \times \mathbb{P}(G_{\cdot 3} = g_{\cdot 3},\boldsymbol{X} = \boldsymbol{x}) \underbrace{=}_{(C.15)} \mathbb{P}((G_{\cdot 3},\boldsymbol{X}) \in C_{g_{\cdot 3},\boldsymbol{x}}) \tag{C.18}$$

Let

$$\hat{\mathbb{P}}_{C_{g_{\cdot 3},\boldsymbol{x}}} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}((G_{\cdot i},\boldsymbol{X}) \in C_{\tilde{g}_{\cdot i},\tilde{\boldsymbol{x}}}) \tag{C.19}$$

From (C.18), $\hat{\mathbb{P}}_{C_{g_{\cdot 3},\boldsymbol{x}}}$ is an unbiased estimator for $|C_{g_{\cdot 3},\boldsymbol{x}}| \times \mathbb{P}(G_{\cdot 3} = g_{\cdot 3},\boldsymbol{X} = \boldsymbol{x}) = \mathbb{P}((G_{\cdot 3},\boldsymbol{X}) \in C_{g_{\cdot 3},\boldsymbol{x}})$ and does not depend on assigned labels.

An algorithm to compute $\hat{\mathbb{P}}_{C_{g_{\cdot 3},\boldsymbol{x}}}$ is the following:

1. Rewrite $(g_{\cdot 3},\boldsymbol{x})$ by listing

   (i) x_3;

   (ii) g_{h3} $\forall h \neq 3 \in \mathcal{N}$ such that $g_{h3} = 1$, disposing them with respect to x_h in ascending order; if there are $h,k \in \mathcal{N}$ with $h \neq k \neq 3$ such that $g_{h3} = g_{k3} = 1$ and $x_h = x_k$, any order is allowed;

   (iii) g_{h3} $\forall h \neq 3 \in \mathcal{N}$ such that $g_{h3} = 0$, disposing them with respect to x_h in ascending order; if there are $h,k \in \mathcal{N}$ with $h \neq k \neq 3$ such that $g_{h3} = g_{k3} = 0$ and $x_h = x_k$, any order is allowed;

   (iv) x_h $\forall h \neq 3 \in \mathcal{N}$ according to the disposition adopted in the previous steps.

2. Call A_3 the obtained row of values.

3. $\forall i \in \mathcal{N}$ in the dataset, list

   (i) x_i;

   (ii) g_{hi} $\forall h \neq i \in \mathcal{N}$ such that $g_{hi} = 1$, disposing them with respect to x_h in ascending order; if there are $h,k \in \mathcal{N}$ with $h \neq k \neq i$ such that $g_{hi} = g_{ki} = 1$ and $x_h = x_k$, any order is allowed;

(iii) g_{hi} $\forall h \neq i \in \mathcal{N}$ such that $g_{hi} = 0$, disposing them with respect to x_h in ascending order; if there are $h, k \in \mathcal{N}$ with $h \neq k \neq i$ such that $g_{hi} = g_{ki} = 0$ and $x_h = x_k$, any order is allowed;

(iv) x_h $\forall h \neq i \in \mathcal{N}$ according to the disposition adopted in the previous steps.

4. Call A_i the obtained row of values $\forall i \in \mathcal{N}$.

Hence,

$$\hat{\mathbb{P}}_{C_{g.3,\boldsymbol{x}}} := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{A_i = A_3\}$$

C.4 Computation of $\tilde{H}^l_{C_{g.3,\boldsymbol{x}}}(\theta)$ and $\tilde{H}^u_{C_{g.3,\boldsymbol{x}}}(\theta)$

The computation of $\tilde{H}^l_{C_{g.3,\boldsymbol{x}}}(\theta)$ and $\tilde{H}^u_{C_{g.3,\boldsymbol{x}}}(\theta)$ can be done via the simple frequency simulator proposed by McFadden (1989) and Pakes and Pollard (1989). Specifically, $\forall i \in \mathcal{N}$, R_M[96] realisations of $\varepsilon_{\cdot i}$ are randomly drawn from its distribution. Let $\varepsilon_{\cdot i,r}$ denote the random vector for the rth draw $\forall i \in \mathcal{N}$. Hence,

$$\hat{H}^l_{C_{g.3,\boldsymbol{x}}}(\theta) := \frac{1}{R_M \times n} \sum_{r=1}^{R_M} \sum_{i=1}^{n} \mathbb{1}(\text{all outcomes of the } \textit{section i game} \text{ fall in } C_{\tilde{g}.i,\tilde{\boldsymbol{x}}}) \tag{C.20}$$

and

$$\hat{H}^l_{C_{g.3,\boldsymbol{x}}}(\theta) := \frac{1}{R_M \times n} \sum_{r=1}^{R_T} \sum_{i=1}^{n} \mathbb{1}(\text{at least one outcome of the } \textit{section i game} \text{ falls in } C_{\tilde{g}.i,\tilde{\boldsymbol{x}}}) \tag{C.21}$$

In the empirical application, $R_M = 100$. In order to establish the value of the indicators function the algorithm illustrated in Appendix C.3 can be employed, combined with the the simplifications described in section 2.3.7 in order to reduce the number of potential equilibria to evaluate at each iteration.

C.5 Steps to construct a confidence region

This section illustrates how to obtain the test statistic $S_M(\theta)$ and the critical value $\hat{c}_{M,1-\alpha}(\theta)$ when constructing a $(1-\alpha)\%$ confidence region for each $\theta \in \Theta^o_{\cdot 3}$. After

[96]The subscript M reminds that R should increase to infinity with sample size to avoid not vanishing simulations errors (CT).

having designed a grid of candidate parameter values[97], for each θ in the grid:

(i) Compute $b_m(\theta)$ $\forall m \in \{1,...,M\}$.

(ii) Compute $\bar{b}_M(\theta) := \frac{1}{M}\sum_{m=1}^{M} b_m(\theta)$. Let $\bar{b}_{k,M}(\theta)$ denote its kth element and $\tilde{\sigma}_{k,M}(\theta)$ a consistent estimator of the asymptotic standard deviation of $\sqrt{M}\bar{b}_{k,M}(\theta)$. Specifically,

$$\tilde{\sigma}_{k,M}(\theta) := \sqrt{\frac{1}{M}\sum_{m=1}^{M}(b_{k,m}(\theta) - \bar{b}_{k,M}(\theta))^2}$$

(iii) Compute the test statistic $S_M(\theta) := \sum_k \left(\min\left\{\frac{\sqrt{M}\bar{b}_{k,M}(\theta)}{\tilde{\sigma}_{k,M}(\theta)}, 0\right\}\right)^2$.

(iv) For each k, compute $\xi_{k,M}(\theta) := \frac{1}{\sqrt{\log(M)}}\sqrt{M}\frac{\bar{b}_{k,t}(\theta)}{\tilde{\sigma}_{k,M}(\theta)}$.

(v) For each k, choose the hard threshold $\zeta_{k,M}(\theta) := \begin{cases} 0 & \text{if } \xi_{k,M}(\theta) \leq 1 \\ \infty & \text{otherwise} \end{cases}$

(vi) Draw with replacement R bootstrap samples i.i.d. over r. In the empirical application, $R = 120$.

(vii) For $r = 1,...,R$

  (a) Repeat steps (i) and (ii) and obtain $\bar{b}^{\star}_{k,M,r}(\theta)$ and $\tilde{\sigma}^{\star}_{k,M,r}(\theta)$ for each k.

  (b) Compute $L_{M,r}(\theta) := \sum_k \left(\min\left\{\frac{\sqrt{M}(\bar{b}^{\star}_{k,M,r}(\theta) - \bar{b}_{k,M}(\theta))}{\tilde{\sigma}^{\star}_{k,M,r}(\theta)} + \zeta_{k,M}(\theta), 0\right\}\right)^2$.

(viii) Take the GMS critical value, $\hat{c}_{M,1-\alpha}(\theta)$, as the $(1-\alpha)$ sample quantile of $\{L_{M,r}(\theta)\}_{r=1}^{R}$.

(ix) Reject if $S_M(\theta) > \hat{c}_{M,1-\alpha}(\theta)$.

Hence, the $(1-\alpha)$% confidence region for each $\theta \in \Theta^o_{\cdot 3}$ is

$$CS_M = \{\theta \in \Theta \text{ such that } S_M(\theta) \leq \hat{c}_{M,1-\alpha}(\theta)\}$$

[97]See Appendix C.6.

C.6 Construction of the initial grid of parameters

One difficulty with conducting inference on sets is scanning over a multi-dimensional parameter space. In practice, what the researcher can do is exploring the parameter space around the global minimum of $S_M(\theta)$ in some rational way. For the empirical application, the slice sampling method used by Kline and Tamer (2016) is employed to construct the initial grid of parameter values. The procedure is as follows:

(i) List many starting values for θ, one of which has all entries equal to zero, others are constructed using the results of simple probits.

(ii) From each starting value, minimise $S_M(\theta)$ running a global optimisation algorithm in Matlab; specifically, a pattern search algorithm (*psearch*) with different polling strategies and a genetic algorithm (*ga*) were used.

(iii) Let s be the global minimum of $S_M(\theta)$.

(iv) Save one vector of parameters solving $S_M(\theta) = s$ and call it θ_s.

(v) Run the pre-implemented slice sampling routine in Matlab (*slicesample*) setting $\mathbb{1}\{S_M(\theta) = s\}$ as the un-normalized density and θ_s as the starting value; save the results of each iteration in the course of the algorithm.

(vi) Look at the parameter values encountered in the course of the algorithm and draw a random sample of 500 points. This sample is the initial grid of parameters.

To guarantee a better exploration of all relevant regions of the parameter space, steps (iv), (v) and (vi) were repeated for each vector of parameters found in step (ii) and solving $S_M(\theta) = s$, and the grids obtained from step (vi) were merged. Moreover, robustness checks on the number of random draws from the un-normalised density were conducted.

Alternative procedures are the simulated annealing method proposed by CT and the differential evolution algorithm described by BMM.

Appendix D

Empirical application

D.1 Data construction and cleaning

In order to extract and merge the information from the Registro Imprese database, each firm is uniquely identified by combining its (i) Chamber of Commerce's territorial province, (ii) R.E.A[98] code, and (iii) tax code. The R.E.A. code is a number assigned to each company when enrolling at the Registro Imprese database. The tax code is a numeric code of 16 digits.

Each board member is uniquely identified by her tax code, which is an alphanumeric code of 16 characters, similar to the Social Security Number in the United States or the National Insurance Number in the United Kingdom.

In order to merge the information from the Registro Imprese database with that from the Cerved database, firms' tax codes are used.

Moreover, industries composed of 1 or 2 firms (because $N \geq 3$) and the industry Holdings (because atypical as composed by firms not involved in the production or exchange of goods or services but only owning owns other companies' stock)[99] are dropped.

D.2 Some network summary statistics

For the purpose of measuring the degree of cohesion, the density of a network $\boldsymbol{G}$ is the fraction between the total number of links in the network and the total number

[98] R.E.A. stands for *Repertorio Economico Amministrativo.*

[99] ATECO 2002 code: 74.15.0.

of possible links

$$\frac{\sum_{i \in \mathcal{N}, j \in \mathcal{N}, i \neq j} G_{ij}}{N(N-1)} \in [0,1]$$

A higher density denotes tighter relations between firms. It can be observed from Table 3.2 that the density varies between 0 and 0.333 and has an average value across industries of 0.005.

The average degree of a node tells how many links a node has on average

$$\frac{1}{N} \sum_{i \in \mathcal{N}, j \in \mathcal{N}} G_{ij} \in [0, N-1]$$

It can be observed from Table 3.2 that it varies between 0 and 1 and has an average value (approximated to the nearest integer) across industries of 0. Such a low average value is in line with the low average density commented above.

The percentage of isolated nodes is computed as

$$100 \times \frac{1}{N} \sum_{i \in \mathcal{N}} \mathbb{1}\{G_{ij} = G_{ji} = 0 \; \forall j \neq i \in \mathcal{N}\} \in [0\%, 100\%]$$

It can be observed from Table 3.2 that it varies between 33.333% and 100% with an average value across industries of 97.665%.

Lastly, the total number of links is obtained as

$$\sum_{i \in \mathcal{N}, j \in \mathcal{N}, i \neq j} G_{ij} \in [0, N(N-1)]$$

It can be observed from Table 3.2 that it varies between 0 and 6 with an average value (approximated to the nearest integer) across industries of 0.

References

Aldrich, H.E. (1979): *Organizations and Environments*, Englewood Cliffs, NJ: Prentice-Hall.

Allen, M.P. (1974): “The Structure of Interorganizational Elite Cooptation: Interlocking Corporate Directorates”, *American Sociological Review*, 39(3), 393–406.

Amstrong, T.B. (2014): “Weighted KS Statistics for Inference on Conditional Moment Inequalities”, *Journal of Econometrics*, 181(2), 92–116.

Andrews, D.W.K., and X. Shi (2013): “Inference Based on Conditional Moment Inequalities”, *Econometrica*, 81(2), 609–666.

Andrews, D.W.K., and X. Shi (2017): “Inference Based on Many Conditional Moment Inequalities”, *Journal of Econometrics*, 196(2), 275–287.

Andrews, D.W.K., and G. Soares (2010): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection”, *Econometrica*, 78(1), 119–157.

Aradillas-Lopez, A., and A.M. Rosen (2016): “Inference in Ordered Response Games with Complete Information”, Working Paper, available *here*.

Atalay, E. (2013): “Sources of Variation in Social Networks”, *Games and Economic Behavior*, 79, 106–131.

Bajari, P., H. Hong, and S.P. Ryan (2010): “Identification and Estimation of a Discrete Game of Complete Information”, *Econometrica*, 78(5), 1529–1568.

Badev, A. (2014): “Discrete Games with Endogenous Networks: Theory and Policy”, Working Paper, available *here*.

Bajari, P., H. Hong, and S.P. Ryan (2010): “Identification and Estimation of a Discrete Game of Complete Information”, *Econometrica*, 78(5), 1529–1568.

Bandiera, O., and I. Rasul (2006): "Social Networks and Technology Adoption in Northern Mozambique", *The Economic Journal*, 116(514), 869–902.

Barabási, A.L., and R. Albert (1999): "Emergence of Scaling in Random Networks", *Science*, 286(5439), 509–512.

Barbi, V. (2000): "Interlocking Directorhip Networks: What is Relevant for the Evolution and Change of the Networks?", *Quaderni del Dipartimento di Economia Politica*, 278, Universitá di Siena, available *here*.

Bayer, P., S. Ross, and G. Topa (2008): "Informal Hiring Networks and Labor Market Outcomes", *Journal of Political Economy*, 116(6), 1150–1196.

Baysinger, B.D., and H.N. Butler (1985): "Corporate Governance and the Board of Directors: Performance Effects of Changes in Board Composition", *Journal of Law, Economics, & Organization*, 1(1), 101–124.

Bellenzier, L., and R. Grassi (2014): "Interlocking Directorates in Italy: Persistent Links in Network Dynamics", *Journal of Economic Interaction and Coordination*, 9(2), 183–202.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification in Models with Convex Moment Predictions", *Econometrica*, 79(6), 1785–1821.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, 76(4), 763–814.

Bernstein, L.A. (1978): *Financial Statement Analysis*, Georgetown, Ontario: Irwin-Dorsey.

Berry, S. (1992): "Estimation of a Model of Entry in the Airline Industry", *Econometrica*, 60(4), 889–917.

Bertoni, F., and P. Randone (2006): "The Small-World of Italian Finance: Ownership Interconnections and Board Interlocks amongst Italian Listed Companies", Working Paper, available *here*.

Besag, J. (1975): "Analysis of Non-Lattice Data", *The Statistician*, 24(3), 179–195.

Bianco, M., and E. Pagnoni (1997): "Board interlocks Across Listed Companies in Italy: The Case of Banks", *BNL Quarterly Review*, Special Issue on Property, Control and Corporate Governance of Banks.

Bisin, A., A. Moro, and G. Topa (2011): "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions", NBER Working Paper 17196, available *here*.

Bloch, F., and M. Jackson (2005): "The Formation of Networks with Transfers among Players", *Journal of Economic Theory*, 133(1), 83–110.

Bloch, F., and M. Jackson (2006): "Definitions of Equilibrium in Network Formation Games", *International Journal of Game Theory*, 34(3), 305–318.

Bjorn P., and Vuong Q. (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: a Game Theoretic Formulation with an Application to Labor Force Participation", Caltech Working Paper 537, available *here*.

Booth, J.R., and D.N. Deli (1996): "Factors affecting the number of outside directorships held by CEOs", *Journal of Financial Economics*, 40(1), 81–104.

Boucher, V. (2016): "The Estimation of Network Formation Games with Positive Spillovers", Working Paper, available *here*.

Bramoullé, Y., and B.W. Rogers (2009): "Diversity and Popularity in Social Networks", Cahier de recherche/Working Paper 09-03, available *here*.

Bramoullé, Y., S. Currarini, M.O. Jackson, P. Pin, and B.W. Rogers (2012): "Homophily and Long-Run Integration in Social Networks", *Journal of Economic Theory*, 147(5), 1754–1786.

Braun, M., and J. McAuliffe (2010): "Variational Inference for Large-Scale Models of Discrete Choice", *Journal of the American Statistical Association*, 105(489), 324–335.

Brouwer, A.E., and J.H. Koolen (2009): "The Vertex-Connectivity of a Distance-Regular Graph", *European Journal of Combinatorics*, 30(3), 668–673.

Bugni, F. (2010): "Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set", *Econometrica*, 78(2), 735–753.

Bugni, F., M. Caner, A.B. Kock, and S. Lahiri (2016): "Inference in Partially Identified Models with Many Moment Inequalities Using Lasso", Working Paper, available *here*.

Bugni, F., I. Canay, and X. Shi (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models", Working Paper, available *here*.

Bunting, D. (1976): "Corporate interlocking. Part III. Interlocks and return on investment", *Directors & Boards*, 1(3), 4–11.

Burt, R.S. (1983): *Corporate Profits and Cooptation*, New York: Academic.

Canay, I. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity", *Journal of Econometrics*, 156(2), 408–425.

Calvó-Armengol, A. (2004): "Job Contact Networks", *Journal of Economic Theory*, 115(1), 191–206.

Calvó-Armengol, A., and R. Ilkiliç (2009): "Pairwise-Stability and Nash Equilibria in Network Formation", *International Journal of Game Theory*, 38(1), 51–79.

Calvó-Armengol, A., and M.O. Jackson (2004): "The Effects of Social Networks on Employment and Inequality", *American Economic Review*, 94(3), 426–454.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): “Peer Effects and Social Networks in Education”, *Review of Economic Studies*, 76(4), 1239–1267.

Calvó-Armengol, A., and Y. Zenou (2004): “Social Networks and Crime Decisions: The Role of Social Structure in Facilitating Delinquent Behavior”, *International Economic Review*, 45(3), 939–958.

Candelaria, L.E. (2016): “A Semiparametric Network Formation Model with Multiple Linear Fixed Effects”, Working Paper, available *here*.

Carrington, P.J. (1981): *Horizontal Co-optation Through Corporate Interlocks*, PhD thesis, Dept. Sociol. Univ. Toronto, available *here*.

Chandrasekhar, A.G. (2016): “Econometrics of Network Formation”, in *The Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoullé, A. Galeotti, and B. Rogers, Chapter 21, Oxford University Press.

Chandrasekhar, A.G., and M. O. Jackson (2014): “Tractable and Consistent Random Graph Models”, NBER Working Paper 20276, available *here*.

Chaney, T. (2014): “The Network Structure of International Trade”, *American Economic Review*, 104(11), 3600–3634.

Chatterjee, S., and P. Diaconis (2013): “Estimating and Understanding Exponential Random Graph Models”, *Annals of Statistics*, 41(5), 2428–2461.

Chernozhukov, V., D. Chetverikov, and K. Kato (2016): “Testing Many Moment Inequalities”, CeMMAP Working Paper 42/16, available *here*.

Chernozhukov, V., H. Hong, and E. Tamer (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models”, *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): “Intersection Bounds: Estimation and Inference”, *Econometrica*, 81(2), 667–737.

Chesher, A., and A.M. Rosen (2012): "Simultaneous Equations Models for Discrete Outcomes, Coherence, Completeness, and Identification", CeMMAP Working Paper 21/12, available *here*.

Chetverikov, D. (2017): "Adaptive Test of Conditional Moment Inequalities", *Econometric Theory*, forthcoming, available *here*.

Christakis, N., J. Fowler, G. Imbens, and K. Kalianaraman (2010): "An Empirical Model of Strategic Network Formation", NBER Working Paper 16039, available *here*.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets", *Econometrica*, 77(6), 1791–1828.

Ciocca, P. (2007): *Ricchi per Sempre? Una Storia Economica d'Italia (1796-2005)*. Torino, Bollati-Boringhieri.

Collins, N.R., and L.E. Preston (1969): "Price-Cost Margins and Industry Structure", *Review of Economics and Statistics*, 51(3), 271–286.

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

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

Conley, T.G., and C.R. Udry (2010): "Learning about a New Technology: Pineapple in Ghana", *American Economic Review*, 100(1), 35–69.

Core, J.E, R.W. Holthausen, D.F. Larcker (1999): "Corporate Governance, Chief Executive Officer Compensation, and Firm Performance", *Journal of Financial Economics*, 51(3), 371–406.

Currarini, S., M. Jackson, and P. Pin (2009): "An Economic Model of Friendship: Homophily, Minorities and Segregation," Econometrica, 77(4), 1003–1045.

Orbanz, P., and Roy, D.M. (2015): "Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures", *IEEE Trans. Pattern Anal. Mach. Intelligence*, 37(2), 437–461.

de Paula, A. (2013): "Econometric Analysis of Games with Multiple Equilibria", *Annual Review of Economics*, 5(1), 107–131.

de Paula, A. (2016): "Econometrics of Network Models", *Advances in Economics and Econometrics: Theory and Applications*, Eleventh World Congress, forthcoming, available *here*.

de Paula, A., S. Richards-Shubik, and E. Tamer (2016): "Identifying Preferences in Networks with Bounded Degree", CeMMAP Working Paper 54/16, available *here*.

Dooley, P.C. (1969): "The Interlocking Directorate", *American Economic Review*, 59(3), 314–323.

Dzemski, A. (2014): "An Empirical Model of Dyadic Link Formation in a Network with Unobserved Heterogeneity", Working Paper, available *here*.

Dorogovtsek, S.N, J.F. Mendes, and A.N. Samukhin (2000): "Structure of Growing Networks with Preferential Linking", *Phys. Rev. Lett*, 85(21), 4633-4636.

Epstein, L.G., H. Kaido, and K. Seo (2016): "Robust confidence regions for incomplete models", Econometrica, 84(5), 1799–1838.

Erdös, P., and A. Rényi (1959): "On Random Graphs I", *Publicationes Mathematicae*, 6, 290–297.

Erdös, P., and A. Rényi (1960): "On the Evolution of Random Graphs', *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 17–61.

Fafchamps, M., and F. Gubert (2007): "The Formation of Risk Sharing Networks", *Journal of Development Economics*, 83(2), 326–350.

Farrell, J., and M. Rabin (1996): “Cheap Talk”, *Journal of Economics Perspectives*, 10(3), 103–118.

Fligstein, N., and P. Brantley (1992): “Bank Control, Owner Control, or Organizational Dynamics: Who Controls the Large Modern Corporation?”, *American Journal of Sociology*, 98(2), 280–307.

Frank, O. (1991): “Statistical Analysis of Change in Networks”, *Statistica Neerlandica*, 45(3), 283–293.

Frank, O., and D. Strauss (1986): “Markov Graphs”, *Journal of the American Statistical Association*, 81(395), 832–842.

Fronczak, A. (2014): “Exponential Random Graph Models”, in *Encyclopedia of Social Network Analysis and Mining*, ed. by R. Alhajj, and J. Rokne, 500–517, Springer.

Gabrielsen, T.S., E. Hjelmeng, and L. Sørgard (2011): “Rethinking Minority Share Ownership and Interlocking Directorships: The Scope for Competition Law Intervention”, *European Law Review*, 6, 837–860.

Galichon, A., and M. Henry (2011): “Set Identification in Models with Multiple Equilibria”, *Review of Economic Studies*, 78(4), 1264–1298.

Garfield, E. (2006): “Histcite Data”, available *here*.

Gayle, G.-L., L. Golan, and R.A. Miller (2015): “Interlocked Executives and Insider Board Members: An Empirical Analysis”, Federal Reserve Bank of St. Louis Working Paper 2015-040A, , available *here*.

Gaviria, A., and S. Raphael (2001): “School-Based Peer Effects and Juvenile Behavior”, *Review of Economics and Statistics*, 83(2), 257–268.

Gilbert, E.N (1959): “Random Graphs”, *Annals of Mathematical Statistics*, 30(4), 1141–1144.

Gilles, R.P., and S. Sarangi (2005): "Stable Networks and Convex Payoffs", Louisiana State University Working Paper.

Goyal, S., and S. Joshi (2006): "Unequal links", *International Journal of Game Theory*, 34(3), 319–349.

Graham, B. (2015): "Methods of Identification in Social Networks", *Annual Review of Economics*, 7(1), 465–485.

Graham, B. (2016): "An econometric model of network formation with degree heterogeneity", CeMMAP Working Paper 08/17, available *here*.

Hallock, K.F. (1997): "Reciprocally Interlocking Boards of Directors and Executive Compensation", *Journal of Financial and Quantitative Analysis*, 32(3), 331–344.

Helmers, C., M. Patnam, and P.R. Rau (2015): "Do Board Interlocks Increase Innovation? Evidence from a Corporate Governance Reform in India", Working Paper, available *here*.

Hermalin, B.E., and M.S. Weisbach (1991): "The Effects of Board Composition and Direct Incentives on Firm Performance", *Financial Management*, 20(4), 101–112.

Hong, H., and J. Xu (2015): "Inferring Latent Social Networks from Stock Holdings", AFA 2015 Boston Meetings Paper.

Jackson, M. (2009): "*Social and Economic Networks*", Princeton University Press.

Jackson, M.O., and B.W. Rogers (2007): "Meeting Strangers and Friends of Friends: how Random are Social Networks?", *American Economic Review*, 97(3), 890–915.

Jackson, M.O. and J. Wolinsky (1996): "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, 71(1), 44–74.

Jia, P. (2008): "What Happens When Wal-Mart Comes to Town: an Empirical Analysis of the Discount Retail Industry", *Econometrica*, 76(6), 1263–1316.

Jochmans, K. (2017): "Semiparametric Analysis of Network Formation", *Journal of Business & Economics Statistics*, forthcoming, available *here*.

Jochmans, K., and M. Weidner (2016): "Fixed-Effect Regressions on Network Data", CeMMAP Working Paper 32/16, available *here*.

Jordan M.I. (2004): "Graphical Models", *Statistical Science*, 19(1), 140–155.

Jordan, M. I., and M. J. Wainwright (2008): "Graphical Models, Exponential Families, and Variational Inference", *Foundations and Trends in Machine Learning*, 1(1-2), 1–305.

Kaido, H., F. Molinari, and J. Stoye (2016): "Confidence intervals for projections of partially identified parameters", CeMMAP Working Paper 02/16, available *here*.

Kaplan, S.N., and D. Reishus (1990): "Outside directorships and corporate performance", *Journal of Financial Economics*, 27(2), 389–410.

Kline, B., and E. Tamer (2016): "Bayesian Inference in a Class of Partially Identified Models", *Quantitative Economics*, 7(2), 329–366.

Koenig, T., and R. Gogel (1981): "Interlocking Corporate Directorships as a Social Network", *American Journal of Economics and Sociology*, 40(1), 37–50.

Kolaczyk, E. (2009): *Statistical Analysis of Network Data*, Springer-Verlag.

Kooreman, P. (1994): "Estimation of Econometric Models of Some Discrete Games", *Journal of Applied Econometrics*, 9(3): 255–268.

Lalanne, M. (2016): "Job networks and Independent Board Appointments: Evidence from US Publicly Quoted Companies", Working Paper, available *here*.

Lee, R. S., and K. Fong (2013): "Markov-Perfect Network Formation: An Applied Framework for Bilateral Oligopoly and Bargaining in Buyer-Seller Networks" Working Paper, , available *here*.

Lee, S., K. Song, and Y. Whang (2013): "Testing Functional Inequalities", *Journal of Econometrics*, 172(1), 14–32.

Lee, S., K. Song, and Y. Whang (2014): "Testing for a General Class of Functional Inequalities", CeMMAP Working Paper 09/14.

Leslie, C. (2004): "Trust, Distrust, and Antitrust", *Texas Law Review*, 83(3), 515–680.

Leung, M. (2015): "Two-Step Estimation of Network Formation Models with Incomplete Information", *Journal of Econometrics*, 188(1), 182–195.

Luzzatto Fegiz, P. (1928): "Il Consiglio d'Amministrazione e l'Interdipendenza delle Imprese", *Giornale degli Economisti e Rivista di Statistica*, 68(3), 197–231.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, 57(2), 995–1026.

Meeusen, W., and L. Cuyvers (1985): "The Interaction between Interlocking Directorships and the Economic Behaviour of Companies", in *Networks of Corporate Power*, ed. by N. Stokman, R. Ziegler, and J. Scott, Cambridge, England: Polity.

Mele, A. (2017): "A Structural Model of Dense Network Formation", *Econometrica*, forthcoming, available *here*.

Menzel, K. (2016): "Strategic Network Formation with Many Agents", Working Paper, available *here*.

Mintz, B., and M. Schwartz (1981): "The Structure of Intercorporate Unity in American Business", *Social Problems*, 29(2), 87–103.

Miyauchi, Y. (2016): "Structural Estimation of Pairwise Stable Networks with Nonnegative Externality", *Journal of Econometrics*, 195(2), 224–235.

Mizruchi, M.S. (1996): "What Do Interlocks Do? An Analysis, Critique, and assesssment of Research on board interlocks", *Annual Review of Sociology*, 22, 271–298.

Mizruchi, M.S, and D. Bunting (1981): "Influence in Corporate Networks: An Examination of Four Measures", *Administrative Science Quarterly*, 26(3), 475–489.

Mizruchi, M.S., and L.B. Stearns (1988): "A Longitudinal Study of the Formation of board interlocks", *Administrative Science Quarterly*, 33(2), 194–210.

Mizruchi, M.S., and L.B. Stearns (1994): "A Longitudinal Study of Borrowing by Large American Corporations", *Administrative Science Quarterly*, 39(1), 118–140.

Myerson, R. (1991): *Game Theory: Analysis of Conflict*. Harvard University Press.

Nakajima, R. (2007): "Measuring Peer Effects on Youth Smoking Behaviour", *Review of Economic Studies*, 74(3), 897–935.

OECD (2008): *Antitrust Issues Involving Minority Shareholding and Interlocking Directorates*, available *here*.

OECD (2010): *Information Exchanges Between Competitors Under Competition Law*, available *here*.

Ornstein, M. (1984): "Board interlocks in Canada: Intercorporate or Class Alliance?", *Administrative Science Quarterly*, 29(2), 210–231.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57(5), 1027–1057.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application", *Econometrica*, 83(1), 315–334.

Palmer, D. (1983): "Broken Ties: board interlocks and Intercorporate Coordination", *Administrative Science Quarterly*, 28(1), 40–55.

Patnam, M. (2013): "Corporate Networks and Peer Effects in Firm Policies", Working Paper, available *here*.

Pattison, P., and S. Wasserman (1999): "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations", *The British journal of mathematical and statistical psychology*, 52(2), 169–193.

Pennings, J.M. (1980): *Board Interlocks*, San Francisco, Calif.: Jossey-Bass.

Pennock, D.M., G.W. Flake, S. Lawrence, E.J. Glover, and C.L. Giles (2002): "Winners Don't Take All: Characterizing the Competition for Links on the Web", *Proceedings of the National Academy of Sciences*, 99(8), 5207–5211.

Pfeffer J. (1972): "Size and Composition of Corporate Boards of Directors: The Organization and its Environment", *Administrative Science Quarterly*, 17(2), 218–228.

Pfeffer J., and G.R. Salancik (1978): *The External Control of Organisations: A Resource Dependence Perspective*, New York: Harper & Row.

Rinaldi, A., and M. Vasta (2005): "The Structure of Italian Capitalism, 1952-1972: New Evidence Using the Interlocking Directorates Technique", *Financial History Review*, 12(2), 173–198.

Rinaldi, A., and M. Vasta (2012): "The Italian Corporate Network After the Golden Age (1972-1983): From Centrality to Marginalization of State-Owned Enterprises", *Enterprise and Society*, 13(2), 378–413.

Rinaldo, A., S. E. Fienberg, and Y. Zhou (2009): "On the Geometry of Discrete Exponential Families with Application to Exponential Random Graph Models", *Electronic Journal of Statistics*, 3, 446–484.

Richardson, R.J. (1987): "Directorship Interlocks and Corporate Profitability", *Administrative Science Quarterly*, 32(3), 367–386.

Robins, G., P. Pattison, and S. Wasserman (1999): "Logit models and logistic regressions for social networks: III. Valued relations", *Psychometrika*, 64(3), 371–394.

Robins, G., T. Snijders, P. Wang, M. Handcock, and P. Pattison (2007): "Recent Developments in Exponential Random Graph (p^*) Models for Social Networks", *Social Networks*, 29(2), 192–215.

Romano, J.P., and A. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Planning and Statistical Inference*, 138(9), 2786–2807.

Romano, J.P., and A. M. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models", *Econometrica*, 78(1), 169–211.

Romano, J., A. Shaikh, and M. Wolf (2014): "A practical two-step method for testing moment inequalities with an application to inference in partially identified models", *Econometrica*, 82(5), 1979–2002.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities", *Journal of Econometrics*, 146, 107–117.

Sacerdote, B. (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, 116(2), 681–704.

Sacerdote, B. (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far", in *Handbook of the*

Economics of Education, ed. by E.A. Hanushek, S.J. Machin, and L. Woessmann, vol.3, 249–277, Elsevier.

Santella, P., C. Drago, and A. Polo (2009): "The Italian Chamber of Lords Sits on Listed Company Boards: An Empirical Analysis of Italian Listed Company Boards from 1998 to 2006", Working Paper, available *here*.

Santella, P., C. Drago, A. Polo, and E. Gagliardi (2009): "A Comparison Among the Director Networks in the Main Listed Companies in France, Germany, Italy, and the United Kingdom", MPRA Paper 16397, available *here*.

Schervish, O. (1995): *Theory of Statistics*, Springer.

Sheng, S. (2016): "A Structural Econometric Analysis of Network Formation Games", Working Paper, available *here*.

Snijders, T. (2002): "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models", *Journal of Social Structure*, 3(2), 1–40.

Solomonoff, R., and A. Rapoport (1951): "Connectivity of Random Nets", *The bulletin of mathematical biophysics*, 13(2), 107–117.

Stokman, F.N, J. Van Der Knoop, and F.W. Wasseur (1988): "Interlocks in the Netherlands: Stability and careers in the period 1960-1980", *Social Networks*, 10(2), 183–208.

Stokman, F.N, F.W. Wasseur, and D. Elsas (1985): "The Dutch Network. Types of Interlocks and Network Structure", in *Networks of Corporate Power. A comparative Analysis of Ten Countries*, ed. by R. Ziegler, and J. Scott, Oxford: Polity Press.

Stoye, J. (2009): "More on Confidence Regions for Partially Identified Parameters", *Econometrica*, 77(4), 1299–1315.

Strauss, D., and M. Ikeda (1990): "Pseudolikelihood Estimation for Social Networks", *Journal of the American Statistical Association*, 85(409), 204–212.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria", *Review of Economic Studies*, 70(1), 147–167.

Topa, G. (2001): "Social Interactions, Local Spillovers and Unemployment", *Review of Economic Studies*, 68(235), 261–295.

Thompson, J.D., and W. McEwen (1958): "Organizational Goals and Environment: Goal-Setting as an Interaction Process", *American Sociological Review*, 23(1), 23–31.

Vasta, M., and A. Baccini (1997): "Banks and industry in Italy, 1911-36: New Evidence Using the Interlocking Directorates Technique", *Financial History Review*, 4(2), 139–159.

Waller, S.W. (2011): "Corporate Governance and Competition Policy", *George Mason Law Review*, 18(4), 833–887.

Warner, W.L., and D. Unwalla (1967): "The System of Interlocking Directorates", in *The Emergent American Society. Large Scale Organizations*, ed. by L. Warner, D. Unwalla, and J. Trimm, 121–157, New Haven.

Wasseman, S., and P. Pattison (1996): "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and $p^{\star}$", *Psychometrika*, 61(3), 401–425.

Zheng, T., M.J. Salganik, and A. Gelman (2006): "How Many People Do You Know in Prison?: Using Overdispersion in Count Data to Estimate Social Structure in Networks", *Journal of the American Statistical Association*, 101(474), 409–423.

Zimmerman, D.J. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *Review of Economics and Statistics*, 85(1), 9–23.