A model of inter-organizational network formation*

Shweta Gaonkar and Angelo Mele

May 17, 2023

Abstract

How do inter-organizational networks emerge? We address this question using an equilibrium framework where firms' decisions to form links with other firms are modeled as a strategic game. In this game, firms weigh the costs and benefits of establishing a relationship with other firms and form ties if their net payoffs are positive. We characterize the equilibrium networks as exponential random graphs (ERGM), and we estimate the firms' payoffs using a Bayesian approach. We apply the framework to a co-investment network of venture capital firms in the medical device industry, providing economic interpretations ERGM parameter estimates. We learn that firms rely on their joint partners (transitivity) and prefer to form ties with firms similar to themselves (homophily). These results hold after controlling for the interdependence among ties. Our structural approach allows us to simulate the effects of economic shocks or policy counterfactuals. We test two such policy shocks, namely, firm entry and regulatory change. We show that while entry of few firms does not have meaningful effects on the shape of the network, a regulatory shock of minimum capital requirements increases the co-investment network's density and clustering.

*Address: Johns Hopkins University, Carey Business School, 100 International Dr, Baltimore, MD 21202. Email: gaonkar@jhu.edu; angelo.mele@jhu.edu

1   Introduction

In a highly competitive environment, organizations rely on their partners to navigate markets and access resources crucial for their success (Gulati 1995a, Powell et al. 1996, Lavie 2007, Gaonkar and Moeen 2023). One such organization is Medtronics that employs a plethora of deals to navigate competitive and emerging markets. For instance, in 2007, Medtronic partnered with Shandong Weigao Group Co. Ltd to open a research and development center in China to develop orthopedic technologies and devices for the local market. The then-president and CEO of Medtronics, Bill Hawkins, described the deal as follows: "*Weigao has a broad orthopedic and trauma product line that complements Medtronic's offerings, but even more importantly, we feel we can generate synergies with their very strong presence and reputation in China. We view Weigao as an ideal strategic partner* ".[1] This quote highlights how organizations come together to share complementary assets and knowledge, gaining access to local markets and reputation through partners' network effects. Our goal is to provide a theoretical and empirical framework to analyze these strategic linking decisions among firms.

In light of the importance of inter-organizational ties (Ahuja et al. 2012), organizational scholars have focused on examining how these relationships are established (Ahuja 2000, Ahuja et al. 2009, Chung et al. 2000, Garcia-Pont and Nohria 2002, Gulati 1999, Kogut et al. 2007, Li and Rowley 2002, Rothaermel and Boeker 2008, Hallen 2008, Reuer and Lahiri 2013, Stern et al. 2014, Lomi and Pattison 2006). In particular, research on tie formation have indicated that links are path-dependent, and firms are more likely to seek future relationships from the pool of their current direct or indirect partners (Gulati 1995a). Consequently, the process of network formation is endogenous, generating interdependence among ties (Ahuja et al. 2012).[2] Organizational network scholars have attempted to control for the endogeneity and interdependence by using the perspective of social network analysis and recommend a structural approach to study tie formation (Borgatti and Halgin 2011).

---

[1]https://www.businesswire.com/news/home/20071217006232/en/Medtronic-Weigao-Announce-Joint-Venture-China

[2]These empirical findings suggest that the prior ties (Gulati 1995a) are important in shaping future relationships, indicating that there is an interdependence among ties.

In this study, we also adopt a structural approach to examine how firms establish relationships by developing a game-theoretical model of network formation, whose strategic equilibrium characterizes the interdependence of firms' linking decisions highligthed in previous studies (Ahuja et al. 2012, Gulati 1995a). Furthermore, an equilibrium model provides a disciplined approach to understanding the network formation process, allowing the researchers to simulate counterfactual policy experiments that predict the changes in network structure in response to any change in the environment surrounding the firms, such as entry and exit of firms, or changes in regulation.

We demonstrate the structural model's use by applying this approach to a network of co-investments among venture capital firms in the medical devices industry, using data from 1940 to 2013. Two or more venture capital firms are linked if they have co-invested in the same medical devices start-up. The network exhibits a standard core-periphery structure, where a few firms have many connections among themselves (the core), while the rest of the firms have few links (or ties) to the main core (the periphery). The structural model allows us to understand what factors lead to this network structure. The estimated results strongly support the homophily argument established by the extant literature on VC syndication networks (Gulati and Gargiulo 1999, Kogut et al. 2007, Zhang et al. 2017). The results suggest that firms tend to form ties with firms similar to them in age, geographic location, and managed capital.

A key advantage of the structural equilibrium approach is that it allows researchers to run policy counterfactual experiments. When there is a policy change or an exogenous shock to the economy, firms will make an optimal decision regarding their new partnerships and alliances, thus changing their network's shape. However, firms will also respond optimally to other firms' decisions due to the policy change or exogenous shock, triggering an additional secondary adjustment of their strategy following these equilibrium effects. Our equilibrium model takes these feedback effects into account, while other partial equilibrium models are not able to incorporate the equilibrium adjustments (Lucas 1976). We focus on two types of policy changes; 1) entry of new firms in the market; 2) imposition of a minimum capital to operate. We show how these policies change the equilibrium network configuration. While

3

entry of ten new firms in New York has a small effect on the new stationary equilibrium, we show that minimum capital requirements increase network density and clustering, as well as homophily in equilibrium. We provide intuition about the mechanisms that lead to new stationary equiibria, showing how we can disentangle the effect of homophily and externalities on the degree distribution. Our analysis further support the need for both channels to explain the equilibrium degree distribution.

We contribute to several strands of literature. The management literature is interested in understanding how network structures emerge (Ahuja et al. 2012, Kim et al. 2015), and scholars have examined tie formation as a consequence of firms attempting to fulfill their strategic or resource needs through their partners. However, the opportunity to form ties is constrained by social opportunities (Eisenhardt and Schoonhoven 1996). In other words, ties established to fulfill a firm's needs are with firms within their current or past relationships. Empirical research that has examined tie formation has focused on the aspect of a firms' resource needs that will be fulfilled through their partners (Chung et al. 2000, Hallen et al. 2014) or their opportunities to establish relationships with other firms (Brennecke 2020, Gulati et al. 2012, Kang and Zaheer 2018, Zaheer and Soda 2009). Few studies have examined network formation as a strategic outcome of all firms' seeking partners.

Empirical works that have studied network formation have explored these relationships as dyadic or triadic level outcomes. Accordingly, these studies have used binary choice models to analyze these dyadic- or triadic-level relationships (Shipilov and Li 2012), wherein firms select partners based on observable attributes, such as their resources, specialization (Chung et al. 2000, Rothaermel and Boeker 2008), trust (Gulati 1995a) or their network positions (Ahuja et al. 2009, Gulati and Gargiulo 1999). However, existing research on embeddedness has shown that network ties are correlated (Gulati 1995a), making each link formation decision endogenous. This stream of research has focused on how the firm's position in a preexisting network structure determines the formation of inter-organizational relationships (Gulati 1995a, Powell et al. 1996, Gulati et al. 2012). These preexisting ties create path dependence in establishing new relationships with other firms because repeated interaction reduces uncertainty while increasing interdependence (Gulati and Gargiulo 1999). In other words, when firms

face competition and uncertainty, they rely on familiar partners, they can trust (Beckman et al. 2004, Sorenson and Stuart 2008). Hence, the decision to form ties is a consequence of the resources its potential partner can offer and the underlying network structure of the focal and partner firm (Eisenhardt and Schoonhoven 1996).

The challenge of estimating network formation using standard binary choice models is that they fail to account for interdependence among ties. For example, when a firm forms several links, a discrete choice model like logit or probit would consider those decisions to be independent.[3] But, the formation of several links consists of interrelated decisions if one assumes the company behaves strategically, and each link is costly. As a consequence, the links should not be considered independent in statistical analysis. In the network economics literature, a network is interpreted as the equilibrium outcome of a strategic game (Jackson and Wolinsky 1996, Mele 2017, Jackson 2008), thus formalizing the interdependence of links as a by-product of strategic decisions. Management and economics researchers who study networks have adopted new modeling and estimation techniques to overcome these challenges. One such approach involves the use of matching models (Mindruta et al. 2016, Fox et al. 2018). However, matching models are better suited for analyzing bipartite networks, where the market can be divided into two sides. Several papers have examined the network formation process using ERGM (Ghosh et al. 2016, Brennecke 2020, Lomi and Pallotti 2012, Lomi and Pattison 2006, Madhavan et al. 2004). Our work contributes to this literature by proposing an equilibrium model of strategic network formation among firms, whose stationary equilibrium is a ERGM. This allows the researcher to interpret the parameters of the ERGM in terms of the underlying strategic game payoffs, and to provide the economic intuition behind the emergence of the firm network (Mele and Zhu 2023, Mele 2017). The mix of utility-based and equilibrium restrictions produces testable implications for homophily, transitivity, and other network features, providing researchers with new insights regarding how firms form ties in equilibrium. This approach also allows scholars to formulate

---

[3]For more details regarding the limitation and current methodology used in management literature, please refer to (Kim et al. 2015, Mindruta et al. 2016).

and test policy counterfactuals leading to a deeper understanding of how these ties are established and change with policy shocks.[4] Finally, additional sparsity can be introduced in the network model using ideas similar to Krivitsky and Kolaczyk (2015) that have been shown to have a behavioral foundation in Butts (2019) and can be easily incorporated in our payoff functions.

2   Theoretical model of network formation

In this section we provide an abridged model description to explain the mechanics of network formation among firms.[5] We model an economy populated by n firms. Each firm is characterized by some observable characteristics. Firm i has attributes x_i, such as firm size, industry, and location. For ease of exposition, we will consider only one discrete attribute in this section (e.g., industry). We describe the network among firms as a $n \times n$ matrix g of zeros and ones, whose entry at row i and column j is g_{ij}; if firms i and j have a link, then $g_{ij} = 1$; otherwise $g_{ij} = 0$.

The network is formed sequentially over time, according to the phases shown in Figure 1. In each period of the network formation game:

1. A pair of firms is randomly chosen, and they have the opportunity to update their mutual link. Updating a link involves either creation of a new tie or ending a pre-existing relationship.

2. Before firms decide to update their link, they observe their matching quality. The matching quality is a random variable that models unobservable characteristics affecting the outcome generated by the link. For example, it may depend on unobservable (to the researcher) long-term strategic goals of the firms, or complementarity of their research and development programs.

---

[4]Our approach differs from Stochastic Actor-Oriented Models (SAOM) (Snijders et al. 2010) because of our focus on a stationary equilbrium, thus allowing us to estimate the model using only one single network. Our approach also differs from Separable Temporal ERGMs (STERGMs), because our assumption on the meeting process among firms allow us to estimate the stationary distribution of the model without estimating the transition probabilities (Krivitsky and Handcock 2014). Furthermore, the networks that maximize the likelihood and are most likely to be observed in the data correspond to our model's game-theoretical equilibrium, where no pair of firms is willing to modify their linking strategy.

[5]For a detailed description of the formal model, refer to the technical appendix.

3. Firms decide jointly how to update the links, on the basis of both firms' payoffs and their matching quality. And, the economy moves to the next period.

Figure 1: Visualization of a time period for the network formation game
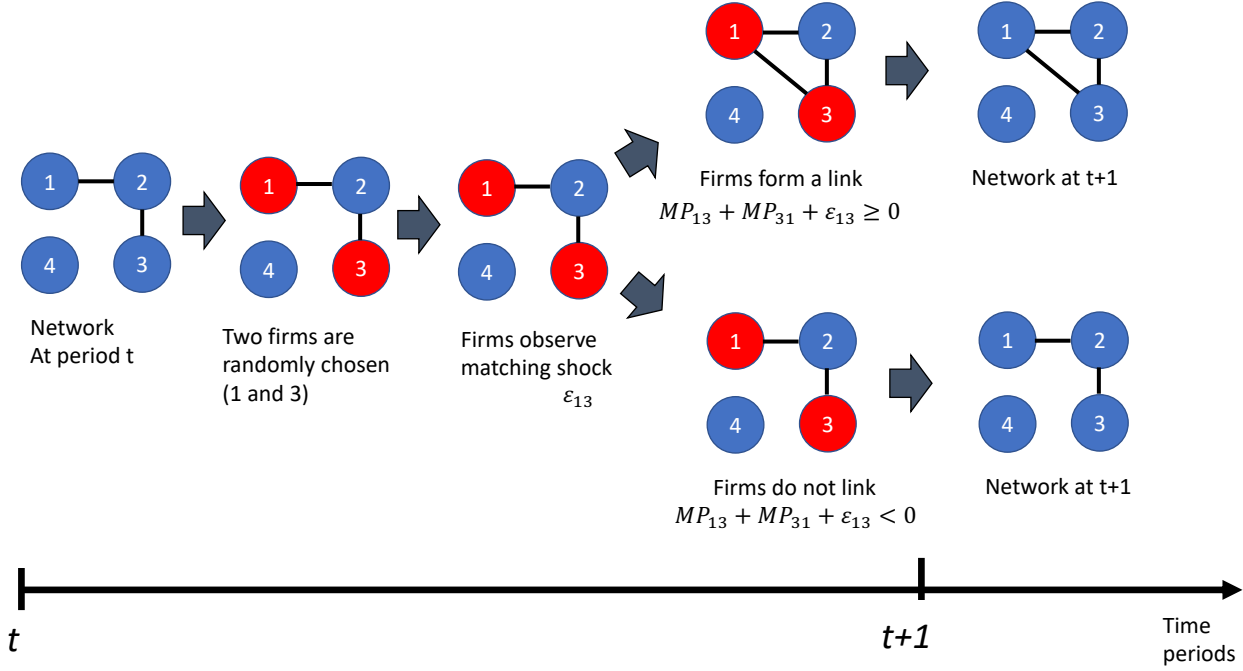
In the example we have 4 firms, represented as a blue dot. At the beginning of period t, the starting network is given by the left-most network, where firm 2 is linked to firm 1 and 3. The period evolves as follows: 1) two firms are randomly chosen, for example firms 1 and 3 (, highlighted as red dots); 2) firms 1 and 3 are able to observe their matching value ϵ_{13}; 3) the link between 1 and 3 is formed if the surplus generated by the link is positive ($MP_{13} + MP_{31} + \varepsilon_{13} \geq 0$), otherwise the network does not change. The network is then updated and becomes the initial network at $t + 1$. The game then enters next period $t + 1$ and the process continues, repeating steps 1), 2) and 3).

The above mentioned process yields a time series of networks. Before showing how this process reaches a stationary equilibrium, we discuss the structure of the meeting process, payoffs, matching quality and firms bargaining power.

The probability that two firms i and j meet and have a chance to change their link at a particular period is $\rho_{ij} > 0$. This probability is *positive for any pair of nodes*, so that each pair of firms has an opportunity to revise their strategies. A low probability means that the pair i and j has rare opportunities to revise their linking strategy; on the other hand, a high probability means that i and

7

j get to revise their linking strategy often over time. The probability ρ is quite general, and it may depend on the structure of the network; for example, firms with many links and are popular may have a higher probability of revisions. The probability ρ_{ij} can also depend on the firms' observable characteristics x_i and x_j. For example, firms that are similar in their observable characteristics may have a higher probability of meeting.

We assume that firms evaluate different networks structures and the composition of their partners using a payoff function. The payoff has two components, deterministic and random, as in standard random utility models (Heckman 1978). The deterministic part of payoff for firm i with a network configuration g, firm attributes x, and the parameters $\theta = (\alpha_0, \alpha_1, \beta, \gamma)$ is given by the sum of net benefits of each link

$$U_i(g, x; \theta) = \sum_{j=1}^{n} g_{ij} \left[\underbrace{\alpha_0}_{\text{cost}} + \underbrace{\alpha_1 \cdot sametype_{ij}}_{\text{homophily/heterophily}} + \underbrace{\beta \cdot pop_j}_{\text{popularity of } j} + \underbrace{\gamma \cdot common_{ij}}_{\text{common partners}} \right]. \tag{1}$$

This payoff is a weighted sum of all links formed by the ego – firm i – where the weight is the function in the bracket and corresponds to the costs and benefits generated by each link. Each additional link has a marginal cost α_0, and benefits accrued by homophily, popularity and shared partners. The variable $sametype_{ij}$ is an indicator equal to one if both firms – i and j – are of the same type ($x_i = x_j$), and zero otherwise. And α_1 captures the payoff for homophily, or how much the firm values homophily. The variable pop_j is the number of links of the alter j, or its popularity; parameter β is the value of each additional link of the alter, or the marginal benefit of the alter's popularity. The variable $common_{ij}$ is the number of shared partners for ego and alter; and the parameter γ is the payoff for each additional common partner, or the marginal benefits of having a shared partner.

This payoff structure allows us to derive the marginal payoff for the ego firm i. The marginal payoff is the additional value received when i forms a link to j, *keeping the rest of the network fixed.*

8

We denote the marginal payoff of i for a link with j as MP_{ij},

$$MP_{ij} = \alpha_0 + \alpha_1 \cdot sametype_{ij} + \beta \cdot pop_j + 2\gamma \cdot common_{ij} \tag{2}$$

The first term of the marginal payoff – α_0 – is the marginal cost of a link. The direct marginal benefit of a partner j of the same type is α_1; if it's positive ($\alpha_1 > 0$), then we have homophily as being of the same type will increase the marginal payoff and therefore the willingness of i to form a link; viceversa if it's negative ($\alpha_1 < 0$) we have heterophily. Notice that in the more general version of the model we may have multiple characteristics and these can be continuous or discrete (see Appendix). [6]

The third term β is the marginal payoff from popularity of firm j. Each additional link that j has formed increases i's marginal payoff by β. If β is positive, firm i gets higher payoffs when it forms a link to a firm with many partnerships. For example, firm j may have many links because of particular resources, such as unique technology or market power. This is a particular form of preferential attachment, where nodes are more likely to link other nodes that are popular (Albert and Barabasi 2002). On the other hand, if the parameter β is negative, then firm i will get lower payoff from linking to a firm with many links. This may reflect a competition effect, where a firm j with many links can devote fewer resources to each of its partner. Ultimately, the sign of β is an empirical question.

The last term, 2γ, is the marginal payoff of triadic closure (transitivity). A new link to j increases (or decreases) i's payoff by an amount 2γ, if j has a common partner. If γ is positive, then having common partners gives higher payoffs. This term may be related to trust, cooperation, or social capital explanations, facilitating collaborations and enforcement of contracts and behaviors. It also captures the idea of repeated ties (Gulati 1995b), wherein firms seek partners among their current relationships. If γ is negative, having common partners will decrease the marginal payoff. The negative payoff is indicative of the fact that the current partners may lock a firm ability to seek partners outside their

---

[6]The general model that we use in the empirical analysis allows for homophily in some characteristics and heterophily in others.

current network, leading to a negative impact of pre-existing ties (Ahuja et al. 2012). As for β, the sign of γ is an empirical matter.

The third and fourth term of the payoff correspond to the *strategic* part of our model, because this part of the payoff depends on linking decisions made by other firms in the market. Indeed, the popularity of the alter is the sum of alter's links, $pop_j = \sum_{r=1}^{n} g_{jr}$; this is a function of linking decisions g_{jr} made by firm j. Analogously, the number of shared partners is $common_{ij} = \sum_{r=1,r\neq i,j}^{n} g_{jr}g_{ri}$ and reflects decisions about linking made by j and other firms r different than i. Therefore these payoff terms describe how firm i responds to the strategic decisions of other firms.

The random component of the payoffs corresponds to a matching quality. We assume that when two firms have the opportunity to revise their mutual link, they observe the value of the matching quality, which is a random variable, modeling the compatibility of the firms along attributes unobserved by the researcher. For example, the matching quality may depend on the long term strategic goals of the firms, their research and development objectives or the quality of their recent hires. We assume that deterministic and random components are additive, therefore the total payoff is the sum of deterministic and random component (Heckman 1978).

We model the formation or deletion of a link as a *cooperative decision, requiring both firms agreement.* Upon being selected with probability $\rho_{ij} > 0$, firms i and j decide how to update their link g_{ij}. We assume that such a decision maximizes the joint surplus generated by the link, i.e., the sum of their marginal payoffs. The marginal surplus generated by adding a link between firms i and j is $MP_{ij} + MP_{ji}$.

Furthermore, *we allow firms to have different bargaining power to share the surplus of a link.* Usually, smaller and less established firms have less bargaining power than more established large firms in negotiations, for example. We model the differential bargaining power by allowing firms to transfer part of their payoff to the other firm, through side payments. Firms with higher bargaining power will receive a higher share of the surplus, through a transfer from the firm with lower bargaining power.

10

This means that for a link between i and j the sum of payoffs is $MP_{ij}+transfer_{ij}+MP_{ji}-transfer_{ij}$, where $transfer_{ij}$ is the payoff transfer from firm i to firm j. Notice that the transfers cancel out since the amount $transfer_{ij}$ transferred from i to j is the same amount received by j from i. Therefore the relevant quantities that determine whether two firms a form link are MP_{ij} and MP_{ji}.

Finally, because the payoff includes an unobserved matching quality associated with creating or deleting a link, we assume that firms observe the random matching quality ε_{ij} before deciding to form or delete a link, respectively. The matching quality models unobserved characteristics of the firms that the researcher cannot observe, but can affect firms' willingness to form a link—for example, compatibility of research programs among two firms, and long-term strategic agendas. We note that the match quality can be positive or negative.

According to this framework, i and j will form a link if the sum of their payoffs and matching quality when they form the link is higher than the sum of payoffs and matching quality when the link is not created, that is

$$g_{ij} = 1 \quad \text{if} \quad MP_{ij} + MP_{ji} + \varepsilon_{ij} \geq 0 \tag{3}$$

This structure is the same as most random utility models in econometrics and statistics and if the matching quality ε_{ij} are independent and identically distributed according to a logistic distribution, we can compute the probability that i and j form a link, conditioning on their characteristics and the shape of the network, as

$$P(g_{ij} = 1|g, x, \theta) = \frac{\exp\left[MP_{ij} + MP_{ji}\right]}{1 + \exp\left[MP_{ij} + MP_{ji}\right]} \tag{4}$$

The structure of this game of network formation is such that only one link gets updated in each period. Furthermore, with high probability, the update will increase the surplus generated by that link.

Summarizing, the model satisfies 3 assumptions:[7] 1) each pair of firm meets with positive proba-

---

[7]See Appendix for the formal statements of each assumption.

bility that does not depend on the existence of a link among them; 2) the matching quality is i.i.d. and logistic; 3) the payoff satisfies the restrictions needed to obtain a potential game. Mele (2017) shows that under these assumptions, if we observe this process of link updates for a large number of periods, the long-run probability of observing a particular network configuration g is[8]

$$\pi(g, x; \theta) = \frac{\exp\left[Q(g, x; \theta)\right]}{c(\theta, x)} \tag{5}$$

where $\theta = (\alpha_0, \alpha_1, \beta, \gamma)$ is the vector of parameters to estimate and $Q(g, x; \theta)$ is the so-called *potential function*,

$$Q(g, x; \theta) = \alpha_0 \cdot links + \alpha_1 \cdot sametype + \beta \cdot twostars + \gamma \cdot triangles \tag{6}$$

where *links* is the total number of links in the network; *sametype* is the number of links between firms of the same type; *twostars* is the number of 2-stars in the network; and *triangles* is the number of triangles. The function $c(\theta, x)$ is a normalizing constant guaranteeing that the likelihood (5) is a proper distribution.

For a set of given values of α_0, α_1, β, and γ, the networks with the highest probability of being observed in the data are the ones that maximize the potential function $Q(g, x, \theta)$.[9] These networks correspond to *pairwise stable equilibrium networks with transfers*, where no pair of firms are willing to mutually form or delete a link (Monderer and Shapley 1996, Mele 2017, Jackson 2008, Jackson and

---

[8]Technically, the sequence of graphs generated by the network formation game is a Markov chain (Levin et al. 2008, Meyn and Tweedie 2009), converging to a unique stationary equilibrium distribution over networks, which we can characterize in closed-form.

[9]However, the networks that maximize the potential function (6) are not necessarily welfare-maximizing, as shown in the appendix. This means that the strategic equilibrium networks do not correspond to the optimal network architecture that maximizes the sum of the firms' payoffs in the economy. Indeed, when forming an additional link, each firm only considers the private costs and benefits of its decision, but does not take into account *all* the externalities created for other firms. In practice, when a firm creates a new direct link, it is also creating an additional *indirect* link for other firms, affecting their payoffs through the popularity and common partner components. In turn, this can increase the odds that those firms will form additional links in future periods. Our strategic equilibrium model captures this mechanism of cascading behavior through the potential function. However, the potential function does not include all the relevant externalities.

Watts 2001, Mele 2022). Thus, the networks with the highest probabilities correspond to a well-defined game-theoretical equilibrium.

We assume that the observed network from the data corresponds to a long-run equilibrium of the model to operationalize estimation. Under this assumption, we can use the likelihood function (5) to estimate the model's parameters. However, this is challenging, because the normalizing constant $c(x, \theta)$ is intractable and cannot be evaluated even in small networks. The next section explains how we circumvent this problem.

Finally, we note that *the model's equilibrium distribution corresponds to a particular instance of an exponential random graph (ERGM)*, with homophily, 2-stars, and triangles (Snijders 2002, Caimo and Friel 2011, Kim et al. 2015, Mele 2017). While ERGMs are widely used by researchers and practitioners to estimate network formation in many applications, our model provides a game-theoretical equilibrium foundation that allows us to interpret the ERGM parameters as marginal payoffs of the firms.

2.1 Understanding the mechanics of the model

To understand the mechanics of the model and how the process of link updates lead to the equilibrium, we consider the model for several configurations of parameters' values. To hihglight the strategic aspect of our modeling strategy, we start from a model without strategic terms in the payoffs and use it as a baseline; we then introduce the strategic terms of the payoffs and see how this changes the possible network shapes in equilibrium.

Model with no strategic terms. To understand the role of the payoffs from popularity and common partners (β and γ), let us first analyze the model without those effects, corresponding to a case where $\beta = 0$ and $\gamma = 0$. In this case, when two firms i and j meet, they have a probability $p_{ij}^s = \frac{\exp(2\alpha_0 + 2\alpha_1)}{1 + \exp(2\alpha_0 + 2\alpha_1)}$ of linking if they are of the same type ($sametype_{ij} = 1$); and a probability $p_{ij}^d = \frac{\exp(2\alpha_0)}{1 + \exp(2\alpha_0)}$ of linking if they are of different types ($sametype_{ij} = 0$). We notice that in this model, the parameter α_0 drives the density of the network. If $\alpha_0 < 0$ and relatively large, then the

probability of a link between two firms is small, leading to a sparse network in equilibrium with few links. However, when $\alpha_0 > 0$, the network is relatively dense. Accordingly, if $\alpha_1 > 0$, the network in equilibrium will have most links among firms of the same type. But, if $\alpha_1 < 0$, most links are formed among firms of a different type. These are the basic elements of the model.

Popularity effects. Let's now consider the model including the effect of popularity, with $\gamma = 0$, but $\beta \neq 0$. In this case, when two firms i and j meet, they have a probability $p_{ij}^s = \frac{\exp(2\alpha_0 + 2\alpha_1 + \beta(pop_j + pop_i))}{1 + \exp(2\alpha_0 + 2\alpha_1 + \beta(pop_j + pop_i))}$ of linking if they are of the same type. If the two firms are of different types, then, their linking probability is $p_{ij}^d = \frac{\exp(2\alpha_0 + \beta(pop_j + pop_i))}{1 + \exp(2\alpha_0 + \beta(pop_j + pop_i))}$. From these probablities we can deduce that there are several cases that would lead to different network outcome. First, suppose that $\alpha_1 > 0$ and $\beta > 0$. Here, firms will tend to form more links to firms of the same type and firms that are popular (have high degree), given all other things being equal. However, in some cases firms of different types may decide to form a link; as long as they are sufficiently popular. To be concrete, consider the case $\alpha_1 = \beta$; then, the probability of a link between i and j when they are of the same type and have both zero links ($pop_i = pop_j = 0$) is $\frac{\exp(2\alpha_0 + 2\alpha_1)}{1 + \exp(2\alpha_0 + 2\alpha_1)}$; if they are of different types but their degrees sum to 2 ($pop_i + pop_j = 2$), the probability of linking is the same as in the previous case, $\frac{\exp(2\alpha_0 + 2\beta)}{1 + \exp(2\alpha_0 + 2\beta)} = \frac{\exp(2\alpha_0 + 2\alpha_1)}{1 + \exp(2\alpha_0 + 2\alpha_1)}$. So in this example, firms may be willing to form links with different types of firms, as long as those firms are popular enough. Third, consider the case $\alpha_1 > 0$ and $\beta < 0$. A similar argument applies, as two firms of different type with no links ($pop_i = pop_j = 0$) may have higher probability of linking than two firms of same type with 4 links ($pop_i + pop_j = 4$), depending on the magnitude of β. Finally, the equilibrium with $\alpha_1 < 0$ and $\beta < 0$ consists of a relatively sparse network where most firms will link to firms of different type, and have few links, generating a relatively sparse network with heterophily.

Transitivity effects. Let's consider the model with $\beta = 0$ and $\gamma \neq 0$. In this case, we only have transitivity effects and no popularity effects. When two firms i and j meet, they have a probability $p_{ij}^s = \frac{\exp(2\alpha_0 + 2\alpha_1 + 4\gamma common_{ij})}{1 + \exp(2\alpha_0 + 2\alpha_1 + 4\gamma common_{ij})}$ of linking if they are of the same type; they have a probability $p_{ij}^d = \frac{\exp(2\alpha_0 + 4\gamma common_{ij})}{1 + \exp(2\alpha_0 + 4\gamma common_{ij})}$ of linking if of different type. If $\alpha_1 > 0$ and $\gamma > 0$, then most links will occur among firms with common partners and the same type. However, if i and j are of different type

14

but have many common partners, it may still be profitable to form a link, since common partners' payoff may be larger than zero. For instance, if $\alpha_1 = \gamma > 0$ two firms of different types with one friend in common will have a higher probability of linking than two firms of the same type without any common partner. The case of $\alpha_1 > 0$ and $\gamma < 0$ implies that having common partners decreases payoffs, thus the probability of linking. The equilibrium networks will be relatively sparse, as most links will occur among firms with no common partners, and the network will also have low clustering. More cases can be analyzed using the same reasoning.

This short discussion shows that even a simplified version of our model is rich enough to provide quite an array of possible network configurations with varying degrees of density, homophily, clustering, and the economy's firms' popularity. In the empirical section, we use a much richer specification of the payoffs to capture more realistic linking behavior and match the observed network properties.

2.2 The importance of the strategic equilibrium for policy analysis

The distinction between the homophily part of the payoff – $\alpha_0 + \alpha_1 \cdot sametype_{ij}$ – and the strategic payoffs – β and γ – has important *policy-relevant implications*. Indeed, many observed networks display clustering, where nodes organize in more densely connected subnetworks while maintaining fewer connections across clusters (Graham 2017, Ahuja et al. 2012). There are at least two possible explanations for high clustering levels. First, the clusters may be the consequence of homophily, a preference to form links to similar nodes. Second, the clusters may arise because of a positive payoff from transitivity or popularity. If most of the clustering is due to payoffs from transitivity or popularity, then an economic shock that affects one of the nodes will also impact other nodes through equilibrium adjustment. However, if the clustering is mostly due to homophily, a shock that affects a node will not spread to other nodes.

The adjustment of equilibrium to potential policy shocks is handy to understand the "what if"

15

questions, as in our policy counterfactuals. For example, suppose that a firm receives a negative economic shock. As a result, the firm may need to end some collaborations resulting in a change in the network's shape. If the payoffs do not include the strategic equilibrium terms for popularity and common partnerships ($\beta = 0, \gamma = 0$), the effect of the economic shock stops here. No other firms will update their linking strategy. On the other hand, if firms' payoffs are affected by the strategic terms ($\beta \neq 0, \gamma \neq 0$), then the network will change and adjust towards a new equilibrium. Going back to the policy shock (mentioned above), where firms decide to sever some links, the firms' partners will suffer a change to their payoff due to the changes in popularity and shared partners. This change means that once firms have elected to revise their strategy, they will update some of their links, either deleting or forming new ones. These updates to the links will result in other firms experiencing a change in their popularity or common partners, creating an incentive to modify their linking strategy in future periods. This chain of adjustments of firms strategies will converge to a new long-run equilibrium network, where with a high probability no firms are willing to change their links.

A model that does not include strategic or equilibrium considerations will not display these feedback effects and, therefore, would predict small effects of policy changes.[10] In the terminology used by Heckman and Vytlacil (2007) our strategic equilibrium model describes the payoff parameters $(\alpha_0, \alpha_1, \beta, \gamma)$, that are *policy invariant*. The idea of policy invariance implies that the payoff parameters do not change when economic conditions change by the effect of a policy or shock to the economy. Furthermore, our model includes *endogenous* quantities – such as the number of transitive triples – that will change after an economic shock because of the equilibrium adjustments.

The model and our approach are useful in two ways. First, it allows us to *test* whether the data support the strategic considerations we postulate. Indeed, after estimating the model, we can test whether the parameters are equal to zero, that is $\beta = 0$ or $\gamma = 0$ or both, according to standard statistical hypothesis testing. This test will inform the researcher whether the strategic considerations

---

[10]This challenge is similar to the Lucas (1976) critique of structural equation models that do not incorporate equilibrium considerations when estimating the parameters.

are essential or not. Second, our model allows us to predict how the network will reach an equilibrium. Therefore it allows us to study how the economy transitions from equilibrium to another after a policy shock.

3   Estimation, Data and Model Specification

3.1   Application: Venture Capital Syndication

We apply this methodology in the context of venture capital investments. Wilson (1968) defines a syndicate as a group of individual decision-makers who must make a common decision under uncertainty, and who, as a result, will receive a joint payoff to be shared among them. Venture capitalists operate in highly uncertain environments and tend to work alongside one another (Gu and Lu 2014, Lerner 1994). These venture syndicates allow venture capitalists to deal with risks in an uncertain environment (Manigart et al. 2006).

Syndication networks provide access to future investment opportunities (Hochberg et al. 2007), thus shaping future co-investment ties (Zhang et al. 2017). As a result, venture capitalists seek new partners based on firm attributes (Sorenson and Stuart 2008), which is guided by their prior investments. There is an interdependence among links in this context (Zhang et al. 2017), creating an estimation challenge that can be dealt with using our structural model and estimation method. Hence, venture capital syndication is an apt empirical setting to use our model to examine network formation.

We built a dataset of all venture capital firms with headquarters in the United States that invested in the medical device industry between 1940 and 2013. Data on venture capitalists come from the Venture Xpert database maintained by Thompson One. Our sample includes all the firms in this database that declare their investment type as "venture capital." This sample gives us a total of 833 firms. In our data, we have information, for each firm, about age, total capital under management,

Figure 2: The network of venture capital co-investments in the medical device industry (1940-2013)
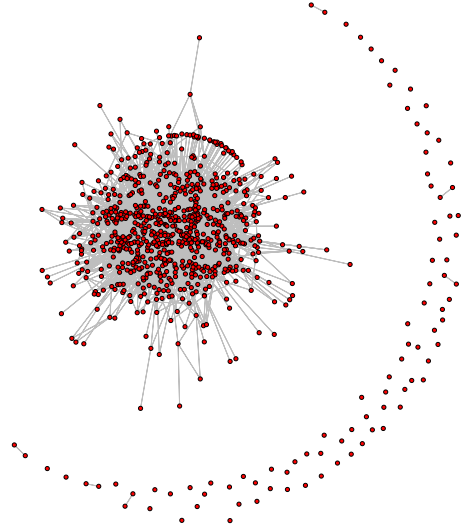
The network consists of 833 venture capital firms that have invested at least once in a company in the medical device industry between 1940 and 2013. There are 6997 co-investment links.

firm type, fund size, and address of headquarters.

The network we examine in our empirical analysis is the co-investment network. If two venture capital firms have invested together in a medical device firm, then two firms are linked or have a tie. The resulting network is undirected, containing 833 nodes and 6997 links.

Figure 2 shows the network from the venture capitalists' data. We observe a typical core-periphery structure, where some firms are extremely connected (the core), and some firms only form a few links to the densely connected core (the periphery). We notice that some firms are isolated, as they did not form syndicates with other firms during the period considered.

3.2 Model specification

We have tried several specifications of the marginal payoff function, and we will focus on the following in the empirical results

$$
\begin{aligned}
MP_{ij} \;=\; & \alpha_0 + \alpha_1 sametype_{ij} + \alpha_2 |capital_i - capital_j| \\
+ \; & \alpha_3 |age_i - age_j| + \alpha_4 samestate_{ij} \\
+ \; & \beta \cdot pop_j + 2\gamma \cdot common_{ij}
\end{aligned}
\tag{7}
$$

The variable $sametype_{ij} = 1$ if the venture capital firms ij belongs to the same type, and $sametype_{ij} = 0$ otherwise. Firms belonging to the following categories: {Angel Group, Bank Affiliated, Corporate PE/Venture, Endowment/Foundation or Pension Fund, Government Affiliated Program, Incubator/Development Program, Individuals, Insurance Firm Affiliate, Investment Management Firm, Private Equity Advisor or Fund of Funds, Private Equity Firm, SBIC, Service Provider, University Program}. The variable $capital_i$ is the (log of) capital under management at venture capital firm i; age_i is the firm's age in 2013; and $samestate_{ij} = 1$ if firms i and j belong to the same state, and $samestate_{ij} = 0$ otherwise. The terms pop_j and $common_{ij}$ are the number of links of firm j (or j's popularity) and the number of common partners of i and j, respectively.

We interpret the first term α_0 of payoff (7) as firm i cost of forming a link. The remaining terms of (7) represent firm's i direct benefit of forming a link with firm j. [11] The sign of the coefficients $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, \gamma)$ is left unconstrained. The interpretation is that if we find $\alpha_1 > 0$, then we have homophily in types, and firms of the same type will be more likely to form links, other things being equal. Vice versa, if we find a $\alpha_1 < 0$, we have heterophily in firm types; thus, firms of the same type will have a lower probability of linking. The same reasoning applies to α_4. However, for α_2 and

---

[11]Our implicit assumption is that the cost of forming links is the same across all the firms, but their benefits may vary according to their characteristics. As explained in Mele (2022) we can only identify the *net benefit* of a direct connection. An alternative specification could assume that the benefit of forming a link is constant and that the cost varies with observable characteristics.

α_3, the reasoning is reversed. If $\alpha_2 > 0$, we have heterophily in the capital, and firms with similar capital levels will have a lower probability of forming a syndicate than firms with large differences in managed capital. On the other hand, when $\alpha_2 < 0$, firms tend to seek partners with similar capital (homophily).

The model parameters to estimate are therefore $\theta = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta, \gamma)$, and we denote the parameter space as Θ, that is $\theta \in \Theta$.

3.3 Estimation

We adopt a Bayesian approach and estimate the posterior distribution of parameters, using the simulation methods developed in (Mele 2017) and (Caimo and Friel 2011). All Bayesian analysis starts from a prior distribution $p(\theta)$ that summarizes the researcher's prior knowledge regarding the parameters of the model. We then use Bayes rule to compute the posterior distribution of the payoff parameters, $p(\theta|g, x)$, defined as the conditional probability distribution of the parameters $\theta = (\alpha, \beta, \gamma)$, given the data, the likelihood of the model and the prior distribution

$$p(\theta|g, x) = \frac{\pi(g, x; \theta)p(\theta)}{\int_{\Theta} \pi(g, x; \eta)p(\eta)d\eta} \tag{8}$$

where $\pi(g, x; \theta)$ is the likelihood distribution (5) and Θ is the set of all parameter values.

Intuitively, the posterior distribution answers the question: what parameter vectors are more likely to generate the network observed in the data, according to our strategic model, and given our prior knowledge about the parameters of the phenomenon under study? The answer to this question is a distribution of parameters that assigns a higher probability for the parameter vectors that are most likely to generate the data.

The Bayesian approach is well suited to estimating this class of models for several reasons. First, the literature on network estimation has established that some exponential random graphs models suffer from *degeneracy*. Degeneracy implies that the model puts the highest probability on a small

20

network set, usually the empty or full network. Degeneracy is problematic because, empirically, it is very unusual to observe empty or complete networks. Researchers have resolved this issue by providing alternative specifications of the model (Snijders 2002, Robins et al. 2007), by including or excluding some terms (like triangles), and by "curving" the exponential distribution (Geyer 1992).

The Bayesian approach allows the researcher to specify priors that consider how some parameter vectors imply degenerate models. A simple solution is to impose null prior probability for regions of the parameter space that generate almost empty or almost complete networks. A strength of our estimation strategy is that it does not rely on this ultimate prior choice. Indeed, the estimation algorithm that we use in this paper, called the *exchange* algorithm (Murray et al. 2006, Mele 2017, Caimo and Friel 2011), simulates from the posterior distribution of the parameters, giving more weight to those parameters that generate networks similar to the observed data. Therefore, the algorithm estimates a posterior where parameters that lead to degeneracy have a small or null probability. This feature of the algorithm is a remarkable property that mitigates the issue of degeneracy explained above.

Second, we bypass the problem of computing the normalizing constant $c(\theta, x)$ of the likelihood, because our algorithm recovers the posterior distribution of the parameters without evaluating the likelihood of the model. While recent advances in computational methods have allowed researchers to perform estimation in large networks (Byshkin et al. 2018, Stivala et al. 2016, Mele and Zhu 2023), the computation of the normalizing constant is still the main challenge in estimating this class of models. The ERGM literature has developed alternative estimators, like the Markov Chain Monte Carlo Maximum Likelihood Estimator (MCMC-MLE) or the Maximum Pseudo-Likelihood Estimator (MPLE). The MCMC-MLE relies on a Monte Carlo approximation of the likelihood, through a long simulation of the model (Snijders 2002, Geyer and Thompson 1992), which estimates the parameters by maximizing the approximated likelihood. As the number of simulations grows large, the parameters estimated with this method converge to the maximum likelihood estimate (Geyer and Thompson 1992). However, practitioners have demonstrated several difficulties with these approximations, which may be

inadequate if the simulation's chosen starting value lies too far from the (exact) maximum likelihood estimate. Besides, there are cases in which the maximum likelihood estimate does not exist; when this is the case, the simulation output is unstable, leading to unreliable inference (Geyer and Thompson 1992, Snijders 2002, Butts 2009).

On the other hand, the Maximum Pseudo-Likelihood Estimator simplifies the estimation by considering the linking choices as independent (Besag 1974, Wasserman and Pattison 1996). This simplification corresponds to a logit estimation with endogenous regressors. One can show that under some assumptions and as the network size becomes large, such estimates are consistent (Boucher and Mourifi 2017, Besag 1974). However, in many practical applications, the estimates are flawed and imprecise (Caimo and Friel 2011, Mele 2017, Snijders 2002). Furthermore, standard errors are underestimated and need adjustment.

To estimate the model, we assume that the network we observe in the data is a stationary network drawn from the likelihood of the theoretical model (5). In many applications, the posterior distribution is available in closed-form; for example, it could be a normal distribution. However, in our case the posterior (8) is a *doubly intractable distribution* because it contains two intractable normalizing constants. Indeed, we can rewrite the posterior as

$$p(\theta|g,x) = \frac{\frac{\exp[Q(g,x;\theta)]}{c(\theta,x)}p(\theta)}{\int_\Theta \frac{\exp[Q(g,x;\eta)]}{c(\eta,x)}p(\eta)d\eta} = \frac{\frac{\exp[Q(g,x;\theta)]}{c(\theta,x)}p(\theta)}{p(g,x)} \tag{9}$$

The first intractable constant is $p(g,x) = \int_\Theta \frac{\exp[Q(g,x;\eta)]}{c(\eta,x)}p(\eta)d\eta$, which corresponds to the normalizing constant of the posterior distribution. This is also called *marginal likelihood* or *model evidence* in the Bayesian literature (Gelman et al. 2003). The marginal likelihood is the probability of observing the network g given our specific model (that is, our model's ability to explain the network in the data). This quantity is intractable because it involves a high-dimensional integration over the parameter space Θ. The computation of this constant is circumvented using Metropolis-Hastings Markov Chain Monte

Carlo (MCMC) algorithms (Liang et al. 2010), simulating parameters from the posterior distribution in an iterative way. However, in this model, we have an additional intractable normalizing constant in the likelihood

$$c(\theta, x) = \sum_{\omega \in \mathcal{G}} \exp\left[Q(\omega, x; \theta)\right].\tag{10}$$

This is a sum over all possible networks with n firms – the set of $\mathcal{G}$ – containing a total of $2^{n(n-1)/2}$ terms. Even a small network with $n = 20$ firms would imply a sum over $2^{90} \approx 10^{27}$ terms, each term involving the computation of the potential function $Q(\omega, x; \theta)$. This is infeasible for most network sizes, implying that the likelihood and the posterior cannot be computed directly.[12]

We use an *approximate algorithm* that circumvents the need to compute both constants $p(g, x)$ and $c(\theta, x)$. Murray et al. (2006) developed the original algorithm, later adapted to networks in Caimo and Friel (2011) and Mele (2017). The methods is implemented in the open source package `Bergm` in the statistical software `R`. Our estimation code is available in Appendix.[13] We provide a brief description of the algorithm and its useful properties. The readers interested in technical details can find a formal description in our appendix [14].

We start the simulation from parameter vector $\theta^{(0)}$ and network observed in the data g. The simulations proceed according to the following steps:

STEP 1 propose a new vector of parameters θ';

STEP 2 given the proposed parameters θ', simulate the network formation process using the model for R steps; collect the network from last step g';

---

[12]With some algebra, we can show that first and second derivatives of the likelihood also depend on the normalizing constant $c(\theta, x)$, thus making maximum likelihood estimation quite challenging. This estimation challenge is an additional reason to prefer the Bayesian approach to this class of models' frequentist approach. However, recent computational approaches and approximations allow researchers to scale estimation to much larger networks (Byshkin et al. 2018, Stivala et al. 2016, Mele and Zhu 2023).

[13]Our data are proprietary, therefore we are not able to share them. The code we provide uses data generated by simulation of the model. This documentation should provide enough guidance to an empirical researcher that wants to implement this method on different data-sets.

[14]And, the proofs of convergence in (Mele 2017)

STEP 3 check if the simulated network g' is similar to the network in the data g;

- if simulated and observed network are similar enough, accept the proposed parameter with high probability;

- otherwise, reject the parameter;

STEP 4 Repeat steps 1 to 3 for S times.

This algorithm samples parameter vectors that are most likely to generate the data. In STEP 1, we propose a parameter vector. Then, we simulate R networks from the model in STEP 2. We need to choose R large enough to be sure that the last simulated network g' is approximately drawn from the distribution (5);[15] we then compare this *simulated* network g' with the observed network g, using a likelihood ratio.[16] If the simulated and observed networks are similar, that is, their likelihood ratio is close to 1, then the parameter vector we have proposed is very likely to generate the network in our data and should receive high probability in the posterior. That implies that we should accept this proposed parameter with high probability. Vice versa, if the simulated and observed networks look very different and the likelihood ratio is very different from 1, the parameter we have proposed is unlikely to generate our data and thus should receive low probability in the posterior. Accordingly, we would reject this parameter vector with a high probability.

This simulation will converge to the posterior distribution of the parameters, provided that both the number of network simulations R, and parameter simulations S are extensive. In setting the number of simulations, we follow suggestions from the applied probability literature that has established convergence speed for this class of algorithms (Bhamidi et al. 2011, Caimo and Friel 2011, Mele 2017).

Our algorithm attenuates degeneracy problems because parameters that generate almost-empty or almost-complete networks will receive very low or null probability in the estimated posterior. Such parameters will not generate simulated networks that resemble our network data. Thus, the

---

[15]A simple rule of thumb is to use $R = n^2 \ln(n)$ where n is the number of nodes of the network. However, we suggest to try different values of R and show an implementation in our sample code.

[16]Mele (2017) (Appendix B) shows that using the likelihood ratio eliminates the constant $c(\theta, x)$ and speeds up computations, thus avoiding the main computational bottleneck.

algorithm will reject them in favor of parameter vectors that generate networks similar to the observed data. Furthermore, the algorithm does not require a careful choice of starting value, unlike in the MCMC-MLE method. Indeed, the exchange algorithm converges to the correct posterior distribution independently of the starting value.[17] By contrast, the MCMC-MLE is very sensitive to the initial parameter value, sometimes providing very poor (local) approximations to the likelihood. Finally, the software for estimation is open source and available in the statistical package R. We use the package `Bergm`, developed by Caimo and Friel (2011), to perform our analysis. We also provide a sample code with guidance in the Appendix. Finally, we believe that the combination of recent advances in computation (Byshkin et al. 2018, Stivala et al. 2016, Mele and Zhu 2023) and Bayesian methods is a promising avenue for estimation of this class of models.

4   Empirical Results

4.1   Descriptive statistics

Our network contains 833 venture capital firms and 6997 co-investment partnerships. As shown in Figure 3, most of our firms are located in California, Massachussets, and New York (left). Most companies in our data are private equity firms (right).
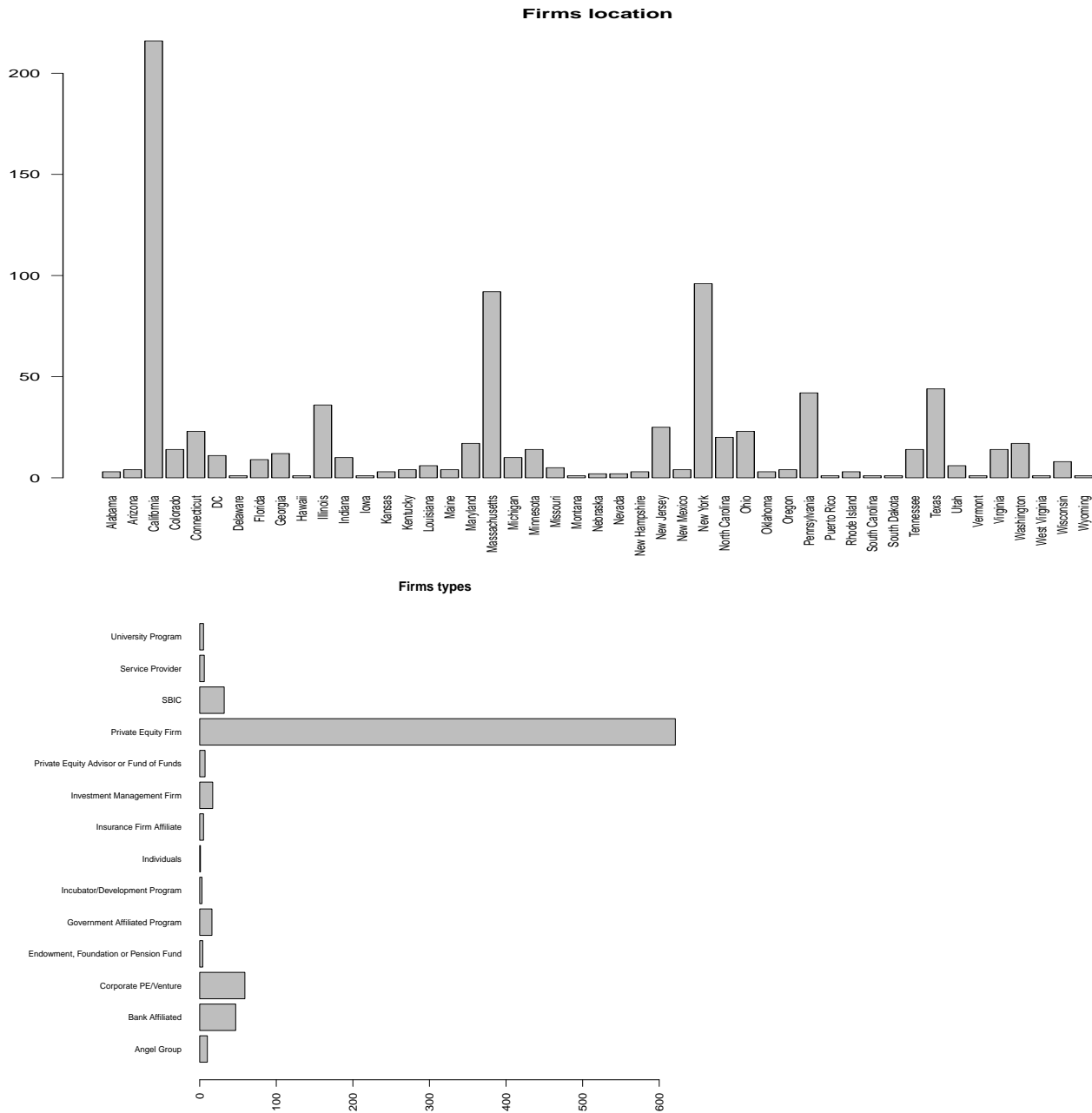
We show in Figure 4 the distribution of capital managed by the firm in logs (left) and the age of the company (right). The largest company manages the capital of 97,700 million dollars. Most companies are relatively young. However, there are a few older firms that have been around for over 50 years.

The average firm has 34 links, while the median is 14. This substantial difference between mean and median is typical of core-periphery networks. In Figure 5 we show the whole degree distribution. The histogram shows the typical pattern of core-periphery networks; a few firms have most connections

---

[17]For a formal proof of convergence, see Appendix B of (Mele 2017).

Figure 3: Distribution of location and types of firms in the sample

(core firms), while most companies have very few links to the core of the network (periphery firms). In the core, we find firms like Delphi Ventures, Johnson & Johnson Development Corp, and Cdib

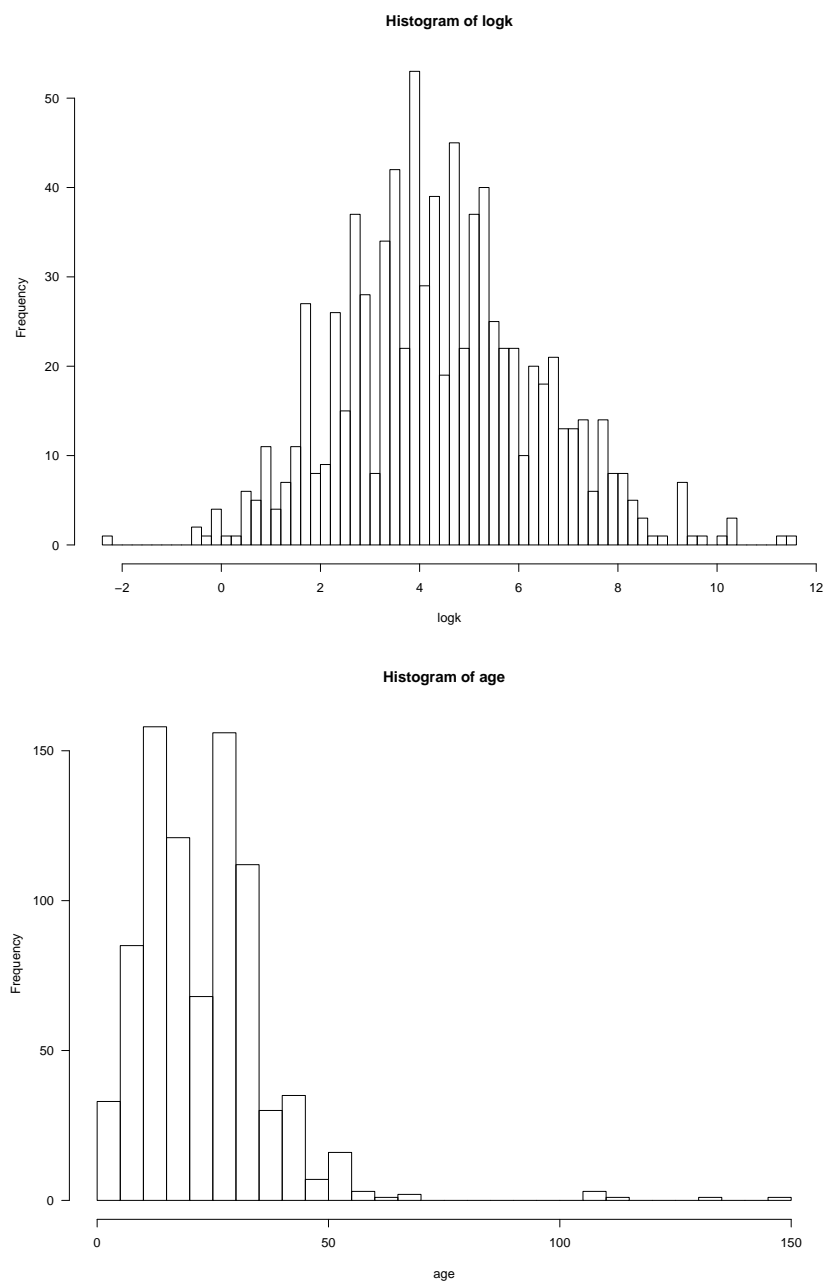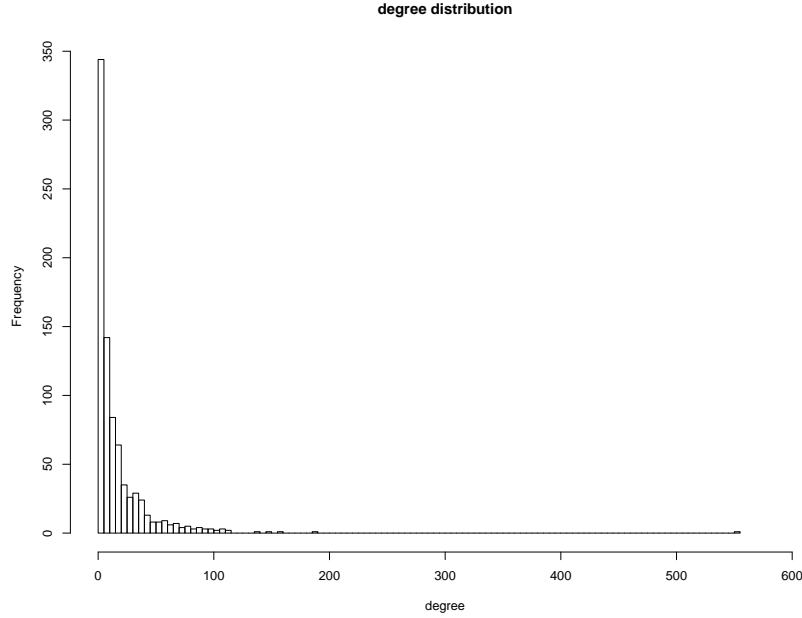Figure 4: Distribution of capital managed (in logs) and age in the sample

Histogram of logk

Histogram of age

Venture Management.

Figure 5: Degree distribution of the co-investment network

4.2 Estimated structural parameters

The posterior estimates for the model appear in Tables 1 and 2. We compare the specification with endogenous network effects (Table 1) to a standard logit model (Table 2) that does not include any endogenous network term ($\beta = \gamma = 0$). While the estimated posterior distributions are similar, below, we show that the endogenous network effects model performs better in matching crucial network properties.

In Table 1, the first column reports the posterior mean, while the second column is the standard deviation of the posterior. The third column shows the standard error of the posterior mean estimate. The fourth and fifth columns show the 95% credible interval. Credible intervals represent the possible values that the model's parameters can take, given the data observed for a given significance level (usually 95%).

Table 1: Model with endogenous network variables

Variable	Posterior Mean	Posterior Std. Dev.	Standard Error	Credible Interval 2.5	Credible Interval 97.5
Cost of link (α_0)	-5.897	0.096	0.0001	-6.083	-5.709
Popularity (β)	0.006	0.0006	0.000006	0.005	0.007
Common partners (γ)	0.526	0.016	0.0001	0.494	0.558
Same firm type (α_1)	-0.054	0.074	0.0008	-0.200	0.090
Abs. Difference Capital (log) (α_2)	-0.023	0.021	0.0002	-0.064	0.018
Abs. Difference Age (α_3)	-0.017	0.003	0.00003	-0.023	-0.012
Same State (α_4)	0.641	0.096	0.003	0.452	0.829

Acceptance rate is 0.132

In Table 1, the estimated cost of forming a link is negative, as expected, and it appears relatively high, with a posterior mean of -5.987. This result implies that syndication links are costly to maintain. The 95% credible interval does not include zero; therefore, we can credibly state that this estimate is negative. The estimate is precise, as shown by the posterior's small standard deviation and the posterior mean's small standard error. The economic interpretation is that forming an additional link will cost an average decrease of 5.897 in payoffs.

The estimated cost seems relatively high, so we need to consider the benefits of partnerships. These estimates are in the rest of Table 1, and include both endogenous network benefits (popularity and common partners) and payoffs from exogenous attributes. The first endogenous network term is the popularity (β), which is determined by the number of 2-stars in the network. The credible interval for this parameter is positive, with a posterior mean of 0.006. Because β is interpreted in our structural model as the marginal payoff, the economic implication is that forming a link to a partner with an additional link would increase payoff by 0.006, on average.

The second endogenous network term is the number of partners shared by the two companies (γ). This term has quite a substantial effect on the payoff; an additional shared partner would increase payoffs by an average of 0.526 (credible interval positive). This value is the marginal payoff of a shared partner. The estimate shows that shared partners have a substantial effect on payoffs, driving the

clustering observed in our venture capital firms' network and explaining the aggregate core-periphery structure.

Not surprisingly, the remaining variables in Table 1 show strong evidence of homophily for two firm attributes. Two venture capital firms belonging to the same firm type do not seem to have a higher-than-random probability of forming a link. This result is shown by the credible interval that includes zero and estimates a quite large posterior probability around zero. In a standard frequentist analysis, this would correspond to saying that the coefficient is not significant.

One would expect that the amount of capital managed by the venture capital firm plays a vital role in determining link formation. Indeed our estimate shows that a 1% increase in the difference (in absolute value) of partners' capital decreases payoffs by 0.023 on average.[18] However, the credible interval includes zero, and therefore this conclusion is weakly supported.

There is strong evidence of homophily in the age of firms. The estimated posterior for the age difference shows that a 1-year difference decreases payoffs by 0.017, providing support to age homophily. The location is critical. Payoffs increase by 0.641 on average if partnered venture capital firms are in the same state.

In summary, venture capital firms tend to prefer syndication with firms of a different type, a similar level of managed capital, and similar age. There is a strong bias towards firms in the same state. However, the homophily effects for the type of firm and capital managed by the firm are not very precisely estimated, so these effects are ambiguous. Focusing on the structural network terms, we see a preference for syndication with firms with a higher number of partners and shared partners. This result is most likely a consequence of the reputation effect. Firms with many partners signal that they have had successful syndications in the past; firms with shared partners provide a screening device, as the joint partner can certify the quality of the previous syndications.

In Table 2, we report the posterior estimates for the standard logit model, where the specification excludes the endogenous structural network terms β and γ. While most of the effects have a similar

---

[18]The difference in capital managed by the venture capital firms is logged to facilitate interpretation.

Table 2: Model with no endogenous network variables

Variable	Posterior Mean	Posterior Std. Dev.	Standard Error	Credible Interval 2.5	97.5
Cost (α_0)	-3.756	0.048	0.0004	-3.849	-3.661
Same firm type (α_1)	0.009	0.041	0.0003	-0.070	0.089
Abs. Difference Capital (log) (α_2)	-0.035	0.012	0.0001	-0.059	-0.012
Abs. Difference Age (α_3)	-0.018	0.002	0.00002	-0.022	-0.014
Same State (α_4)	1.049	0.059	0.0005	0.933	1.166

Acceptance rate is 0.189

sign, we notice that the estimated cost is of lower magnitude; and a higher coefficient for homophily by the state. This result means that homophily imputes the equilibrium network effects for popularity and transitivity estimated in Table 1, rather than the clustering. Therefore, it is essential to determine which specification provides the best fit for our data because this will determine whether the observed network's core-periphery structure is generated by pure homophily effects or by the transitivity and reputation effects implied by the effect of clustering in the payoff function. Furthermore, as explained in the theory section, the effect of external economic shocks are different in a model with network effects.

4.3 Goodness of Fit

In the alliance literature, researchers have few ways to check the fit of a model. One of the advantages of our structural Bayesian approach is that we can check whether the estimated model posterior can replicate the network (Caimo and Friel 2011, Mele 2017, Caimo and Lomi 2015, Kim et al. 2015).

We take a sample of 1000 parameter vectors from the posterior distribution estimated in the previous section to implement goodness-of-fit tests. For each of these parameter vectors, we simulate our model and generate a network. We then compare the simulated networks to the observed network. A good fit should generate networks similar to the one observed in the data, for example, in terms of similar degree distribution and similar triangle counts. As a robustness check, we compare our simu-

lated networks' distribution, and the observed one for network features not included in the estimation. In our payoff specification, we explicitly include transitivity and popularity effects. Showing that our estimated model can replicate the transitivity and popularity observed in the network would not be surprising nor proof of good fit because we target these network statistics directly in our specification.

Therefore, we provide tests for network statistics that are endogenous, but not included in the payoffs. The geodesic distance is the minimum shortest path in the network between two nodes. The number of edge-wise shared partners counts the number of pairs of nodes with exactly k common partners, where k varies from 1 to the maximum number of nodes. Our goodness of fit test will focus on these statistics.[19]

We show the results in the Top panel of Figure 6 for the model estimated in Table 1. This figure shows three goodness-of-fit statistics: the degree distribution, the geodesic distance, and the edge-wise shared partners. The red line represents the observed data. The boxplot is drawn at the interquartile range (25% and 75%). The grey lines represent 95% confidence bands. If our estimated model can replicate the observed network well, then the simulated networks should have degree distributions that are "close" to the observed degree distribution. In the picture, we expect to see most of the simulations falling within the boxplots, and certainly within the 95% bands. This is the case in the Top Panel of Figure 6 for degree distribution, geodesic distance, and edge-wise shared partners.

We compare these test results to the goodness-of-fit test for the model estimated in Table 2. This model does not include the endogenous network statistics (stars and triangles), and it is equivalent to a logistic regression model. The fit of this specification is inferior, as shown in the Bottom Panel of Figure 6. This result implies that it is essential to include the endogenous network effects in the payoff functions. Therefore our equilibrium model captures essential features of the data.
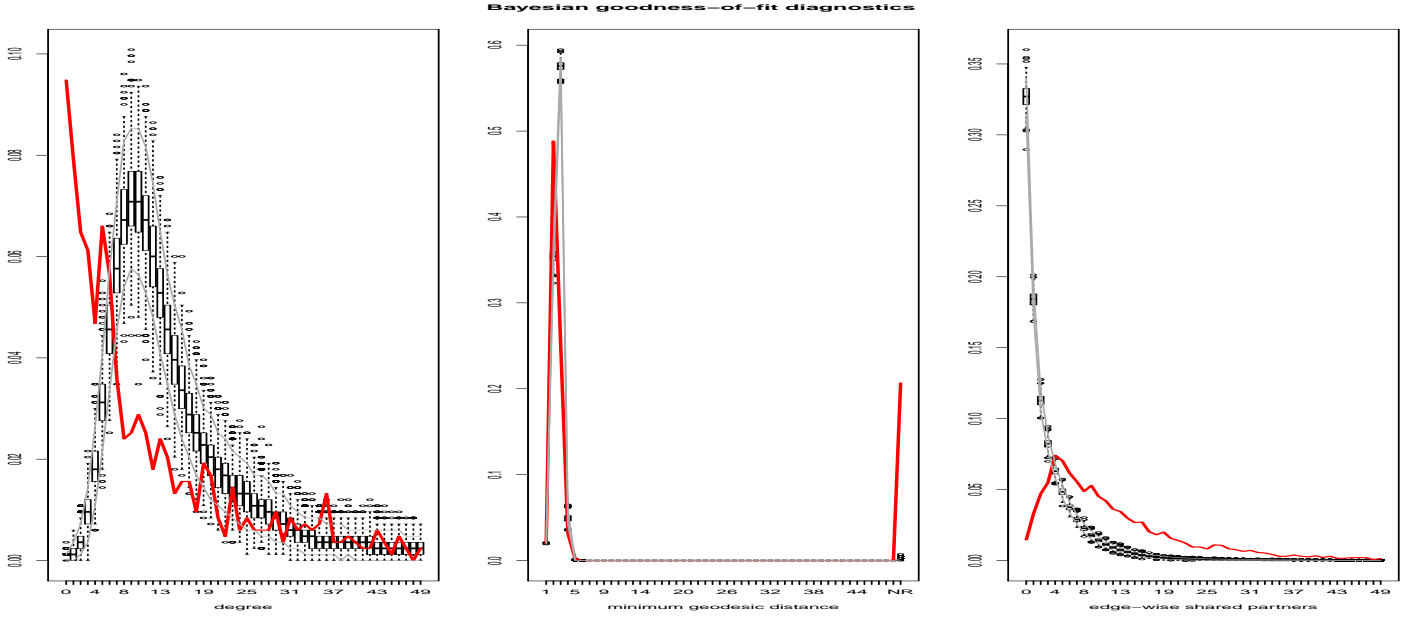
---

[19]This is also the default output of the goodness of fit test in the R package `Bergm` that we use for our empirical analysis.

Figure 6: Goodness of fit tests
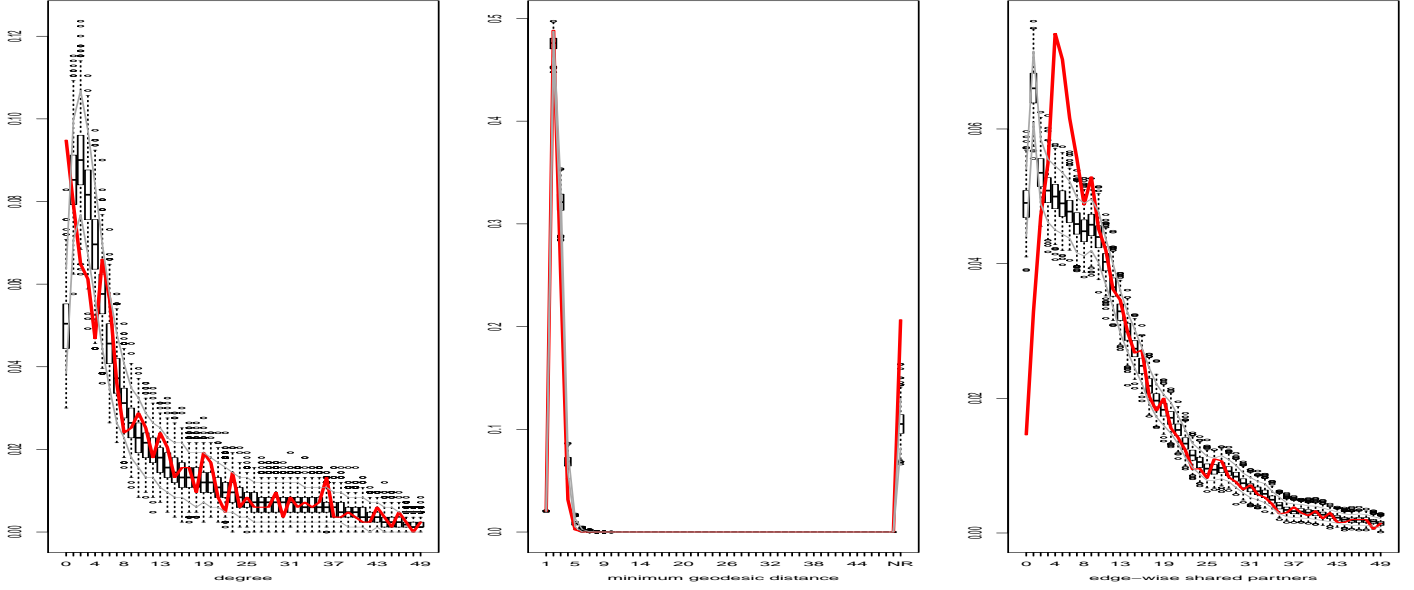
Equilibrium Model (Table 1)

No endogenous effects (Table 2)

We report goodness of fit test for the equilibrium model estimated in Table 1 (Top Panel) and the model without endogenous network terms in Table 2 (Bottom Panel). Each figure is generated through a simulation of 1000 networks using the estimated posterior distribution. For each network, we compute degree distribution, geodesic distance, and edge-wise shared partners. The red line represents the observed network's value, while the black boxplots summarize the simulated networks. A good fit entails simulations such that the red lines fall within the boxplots limits. We conclude that our equilibrium model fits the data better.

33

5 Counterfactual policy experiments

The main advantage of a structural model compared to the reduced-form models is the ability to run counterfactual experiments. For example, what would be the effect of several firms entering the market in a particular period on the equilibrium network? Answering this question using a reduced-form approach will incur in the so-called Lucas Critique (Lucas 1976). This inability to run policy counterfactuals arises since a logit or probit model is a partial equilibrium model, where the aggregate equilibrium effects are not incorporated. In practice, the parameters estimated using a reduced-form approach are not policy-invariant (Heckman and Vytlacil 2007). Therefore, they cannot be used to make predictions on the effect of policy changes. On the other hand, our model includes the equilibrium feedback directly in the payoff function and models the equilibrium in the strategic network formation game.

An additional advantage of our structural model is that quantities like centrality, clustering, and network density are considered equilibrium quantities. When there is a policy change, we expect those network statistics to change as well. This change will impact how our firms make decisions about alliances and links, therefore further impacting the value of centrality, clustering, and density. This feedback effect generated by the equilibrium model is absent in a standard logit or probit model.

We use our estimated model to see how the equilibrium networks would change in two different scenarios: 1) Entry of 10 Venture capital firms in New York; 2) Regulation that requires minimum capital to enter the market.[20] For each policy experiment, we simulate 1000 equilibrium networks using the posterior estimates of the structural parameters.[21] We compute density, clustering, homophily by firm type and state, and the degree distributions for each of these networks. We then compare the distribution of each of these structural equilibrium features with the stationary distribution before the

---

[20]In the last counterfactual policy simulation, we assume that firms that do not fulfill the minimum capital requirement exit the market. We assume that firms do not merge to avoid the new policy's consequences because the decision to merge is not modeled explicitly in our framework. In principle, we could simulate the merging decisions using an auxiliary model, but we keep things simple in our example.

[21]Each network is simulated using a large number of steps, to make sure that the network is an approximate draw from the new equilibrium of the model, after the policy change.

policy change. This comparison allows us to determine the effect of a shock or policy change on the stationary equilibrium network architecture.

In Figure 7, we show the results for the entry of 10 private equity firms in the New York area. In the simulation, the (log) capital managed by the entrants is randomly drawn from a normal distribution with mean and standard deviation equal to the observed (log) capital mean and standard deviation. Therefore, the entrants are, on average, similar to the incumbent firms.

We focus on the new equilibrium network's four structural features: (A) the density of links,[22] (B) the clustering,[23] (C) the fraction of links between firms of the same type, and (D) the fraction of links between firms in the same state. The statistics (C) and (D) are measures of homophily. In panels (A) through (D) in Figure 7, the red lines represent the stationary distribution before entry, while the blue densities are the distribution of the simulated equilibrium networks after entry.

Our results show that when these ten private equity firms enter the New York Market, the effect is relatively small. A t-test rejects the equality of the means for all four network statistics.

To gain some intuition about the mechanisms leading to this outcome, we ran auxiliary simulations, whose output is reported in Figure 8. In red we report the observed degree distribution and in blue the average degree distribution generated in Figure 7. The black line is the average degree distribution implied by the counterfactual assuming no externalities ($\beta = \gamma = 0$). Figure 8 shows that in this case the network is mostly driven by homophily, but because the cost of forming links is relatively high, most firms create just a few links to similar firms. In green we report the average degree of the same counterfactual assuming there is no effect of covariates ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$). The results show that this leads to more clustering and higher density. When we combine both mechanisms we obtain the blue line. In short, to obtain the density, clustering and homophily levels observed in the stationary distribution after entry, we need both effects of homophily and externalities. This further validates the equilibrium approach.
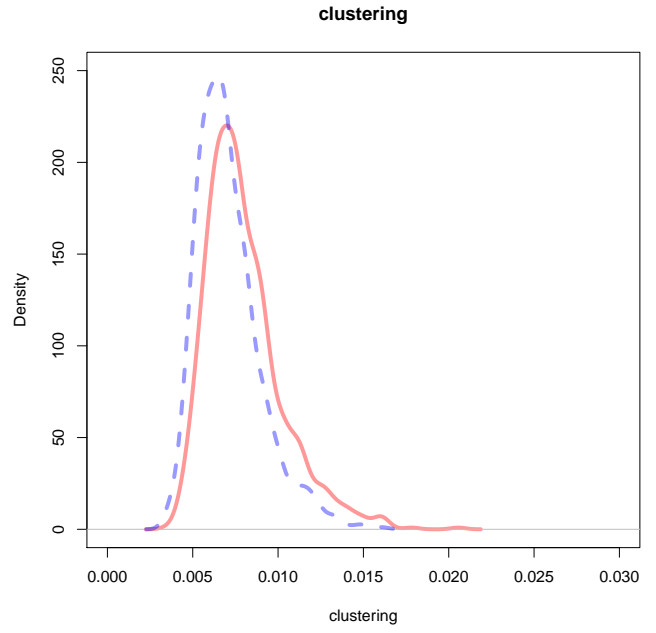
---

[22]We measure the ratio of the total number of links over the maximum possible number of links.

[23]Our measure of clustering is the total number of triangles divided by $n(n-1)(n-2)/6$. See (Jackson 2008) for alternative ways to measure clustering.
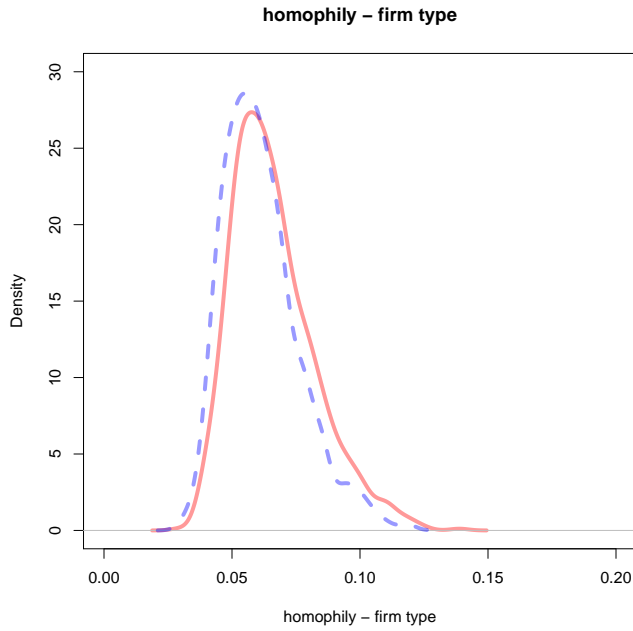
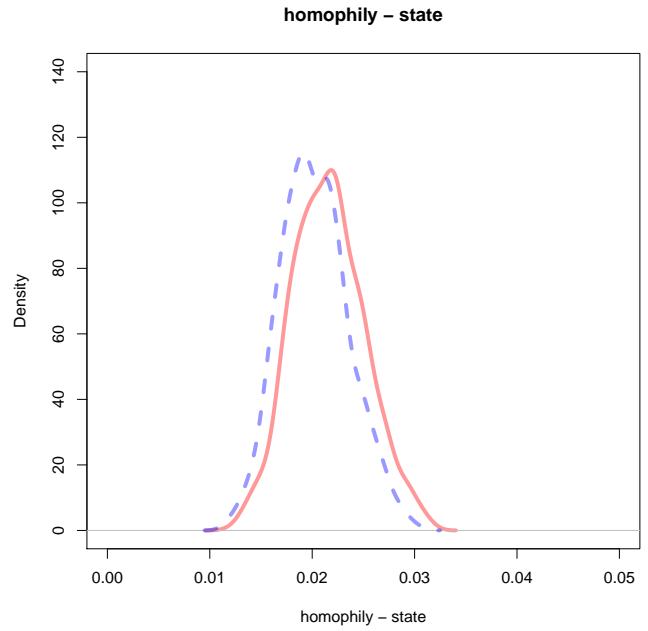Figure 7: Policy counterfactual: Entry of 10 firms in New York
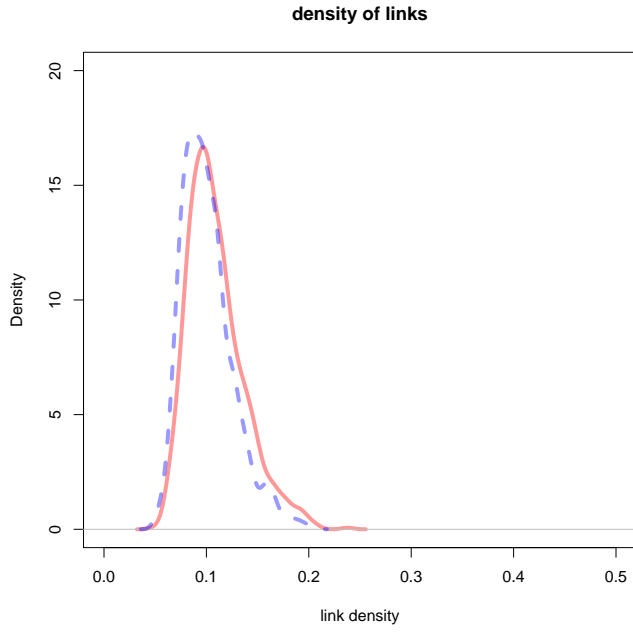
(A) Link density

(B) Clustering

(C) Homophily - Firm type

(D) Homophily - State

The red line represent the marginal densities of the network statistics according to the estimated stationary distribution. The dotted blue line shows the same marginal distributions according to the new stationary distribution, after the entry of 10 firms in New York state. All the figures are obtained from 1000 network simulations from the posterior distribution.

Figure 8: Entry of 10 firms in New York, relative magnitudes of homophily and externalities effects on the degree distribution

degree distibution
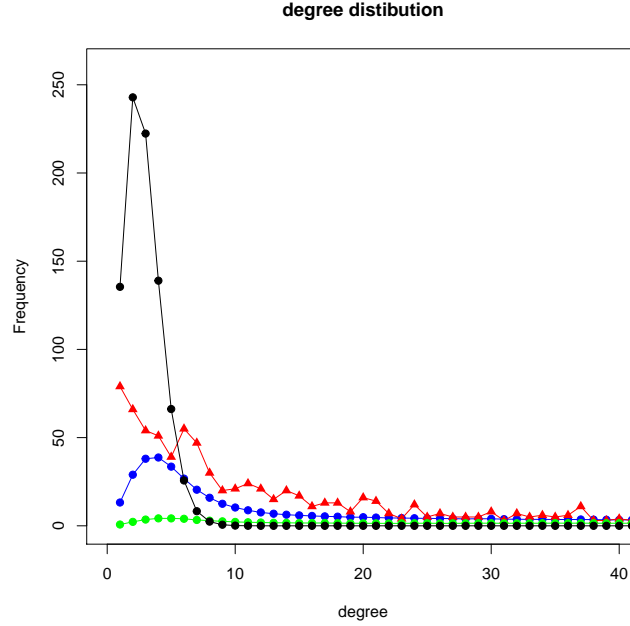
The lines show the degree distribution of a counterfactual simulation where 10 firms enter the market in New York. The red line represent the observed degree distribution in the data. The blue line is the average degree distribution from the simulated counterfactual. The black line is the average degree distribtuion when there is no externality ($\beta = \gamma = 0$). The green line is the average degree distribution when the effect of covariates is set to zero ($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$).

The second counterfactual is a policy in which the regulator imposes a minimum capital require-ment for venture capital firms. In our simulation, we impose a minimum capital that is the observed 25% quantile of capital managed by the firms in the market. To keep the simulation simple, we as-sume that the firms that do not have enough capital will exit the market. We, therefore, abstract from the possibility of mergers to fulfill the new capital requirements. The main reason is that we have not modeled the merger decision in our model, and therefore this would not follow the spirit of the structural analysis.

Figure 9 shows the results. After the policy, there are only 625 firms in the market with the minimum required capital. The new equilibrium configuration (blue dotted lines) involves a denser network and a much more clustered network. These effects are relatively large. In the new equilibrium,

the number of firms with no syndication links drops on average by 70. These firms' exit contributes to the increase in density. In the new equilibrium with fewer firms, it becomes even more profitable to form a link, as the effect of externalities pushes more firms to create new links, thus increasing clustering and magnifying the effect of transitivity, further incentivizing the creation of new ties. The figure shows a significant change in the shape of the network, which is driven by equilibrium effects.
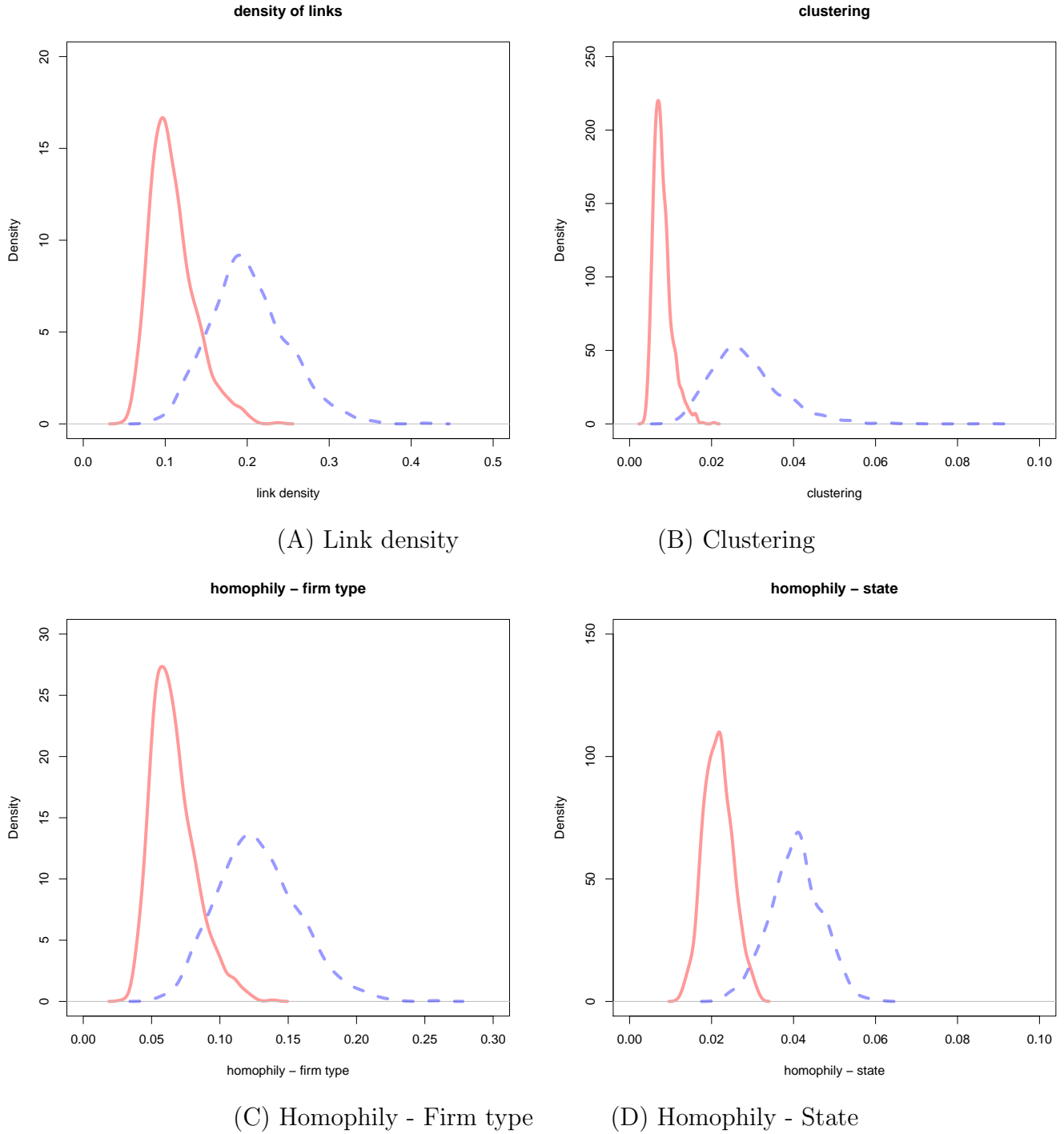
6   Conclusions and Discussion

Several studies have analyzed how firms establish ties to fulfill their resource needs (Chung et al. 2000, Rothaermel and Boeker 2008), access knowledge (Baum and Cowan 2010), and gain reputation (Gu and Lu 2014). The most salient challenge in this research area addresses the network formation process's endogenous nature stemming from the interdependence among ties. Management scholars have empirically examined tie formation as a dyadic Ahuja (2000), Ahuja et al. (2009), Chung et al. (2000), Garcia-Pont and Nohria (2002), Gulati (1999), Li and Rowley (2002), Rothaermel and Boeker (2008), Hallen (2008), Reuer and Lahiri (2013), Stern et al. (2014) or triadic relationship (Davis 2016) using a reduced form regression. These studies have provided valuable insight into the tie formation's micro-mechanisms and highlighted how dynamic it is. However, a reduced-form estimation using the binary choice model to understand link formation has two fundamental limitations. First, reduced form regression is lacking in its ability to account for interdependence among ties (Mindruta et al. 2016, Kim et al. 2015).[24] Second, a reduced form regression does not enable scholars to run policy counterfactuals. This study addresses both these challenges by developing a general equilibrium model to describe the formation of ties among firms.

We develop a strategic network formation model that identifies a firm's net benefit from forming a link in equilibrium. This approach controls the endogeneity of network effects, such as transitivity, by

---

[24]More recently, studies have tackled this challenge using Exponential Random Graph Models or ERGMs (Snijders 2002, Brennecke 2020). Despite the progress, using ERGM to estimate tie formation is that the economic interpretations based on the parameter estimates remain challenging.

Figure 9: Policy counterfactual: Minimum capital requirement

density of links

Density

link density

(A) Link density

clustering

Density

clustering

(B) Clustering

homophily – firm type

Density

homophily – firm type

(C) Homophily - Firm type

homophily – state

Density

homophily – state

(D) Homophily - State

The red line represent the marginal densities of the network statistics according to the estimated stationary distribution. The dotted blue line shows the same marginal distributions according to the new stationary distribution, after the minimum capital requirement is enforced. All the figures are obtained from 1000 network simulations from the posterior distribution.

imposing strategic equilibrium conditions on the linking decisions. Consequently, we obtain a coherent economic and statistical framework to estimate the effect of endogenous network characteristics, such as clustering or popularity. Furthermore, we show that our theoretical model's equilibrium corresponds to an exponential random graph model (ERGM), and our Bayesian estimation strategy improves existing methods by attenuating degeneracy problems [25] and guarantees convergence to the structural parameters' posterior distribution.

Our structural equilibrium approach's key advantage is that the estimated parameters correspond to meaningful economic concepts, such as the net marginal benefits of syndication for each firm, including marginal costs, benefits and externalities. Additionally, the network structural features, such as density, clustering, or homophily, are strategic equilibrium quantities. This focus on the equilibrium implies that the model can predict how the structural features will change when there is an external shock such as policy change, entry, or exit of firms. Therefore, using a utility-based equilibrium model allows us to perform analysis of policy counterfactuals. These counterfactual experiments allow managers or organizations to understand the implication of their strategic decisions on the emergence of inter-organizational network structure and decode the "what if" questions.

Our application to Venture Capital syndication networks provides a simple illustration of how the methodology can be implemented, from the specification of the marginal payoff functions, estimation of the model, the economic interpretation of the structural estimates, and simulation of counterfactual policy experiments. Our results confirm previous findings that venture capital firms form syndicates due to their preference for firms similar to them regarding age, managed capital, and geographic location, after controlling for the interdependence among ties and the equilibrium restrictions imposed by our model (Gulati and Gargiulo 1999, Sorenson and Stuart. 2001, Kogut et al. 2007). Our policy counterfactuals show how the system responds to shocks and quantitatively assess the effect on the clustering, density and homophily in the aggregate.

---

[25]The simulation of models suffering from degeneracy problems tend to sample degenerate networks, either almost empty or almost complete. This feature of the model makes it challenging to fit the observed data and makes estimators unstable and unreliable.

While our method has several advantages, we need to acknowledge some minor limitations. First, the estimation is computationally intensive. In particular, the simulation of the posterior distribution may be more computationally intensive than the commonly used Markov Chain Monte Carlo Maximum Likelihood Estimator (MCMC-MLE). However, we note that this problem will decrease in a few years with faster hardware and better computational methods (Bhamidi et al. 2011, Byshkin et al. 2018, Stivala et al. 2016, Mele and Zhu 2023). Second, we model payoffs as functions of observable firms' characteristics, though unobserved heterogeneity may play a role in forming links. Introducing unobserved heterogeneity in our model is possible at the cost of a substantial increase in computational burden. While this is an essential direction of research, we leave the detail of this extension to future work. Third, in the empirical analysis, we use data from 1940 to 2013, and we include all the firms and links appearing in our data during that period, abstracting from entry and exit. This is a limitation of the stationary approach, because it only focuses on the network existing at a single point in time; but it is also a limitation of our data lacking information on entry and exit. We also abstract from collaborations among VCs in other industries, because we have data on the links within the medical device industry only. This may affect our results as it will not take into account additional links and externalities from other sectors. We conjecture that including unobserved heterogeneity in the model may help attenuate these effects.

This approach can be adopted in management research to study other kinds of networks. Several empirical networks studied in the management literature display a similar core-periphery structure. For example, Fleming et al. (2004) examines the emergence of a dense cluster in the network of patents' co-authors in the early 1990s in Silicon Valley. This model would be applicable to study the emergence of co-author networks and predict changes in the network structure based on competition or any other policy change. Our model can be used as a framework to understand how endogenous ties create knowledge spillovers in equilibrium.

Furthermore, our approach allows the simulation of counterfactuals, such as the effect of subsidies on research and innovation on the network structure, or how public policies affect knowledge diffusion.

We believe that the equilibrium approach to networks used in this paper, combined with recent computational advances in estimation using large networks (Byshkin et al. 2018, Stivala et al. 2016, Mele and Zhu 2023), provides a robust framework for the analysis of networks in management. We hope that our work will stimulate more researchers to apply this framework to new questions and datasets.

References

Ahuja G (2000) The duality of collaboration: Inducements and opportunities in the formation of interfirm linkages. *Strategic Management Journal* 21(3):317343.

Ahuja G, Polidoro F, Mitchell W (2009) Structural homophily or social asymmetry? the formation of alliances by poorly embedded firms. *Strategic Management JOURNAL* 30(9):941–958.

Ahuja G, Soda G, Zaheer A (2012) The genesis and dynamics of organizational networks. *Organization Science* 23(2):434–448.

Albert R, Barabasi AL (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74:47–97.

Badev A (2021) Nash equilibria on (un)stable networks. *Econometrica* 89(3):1179–1206.

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

Baum JAC, Cowan R (2010) Network-independent partner selection and the evolution of innovation networks. *Management Science* 56(11):20942110.

Beckman CM, Haunschild PR, Phillips DJ (2004) Friends or strangers? firm-specific uncertainty, market uncertainty, and network partner selection. *Organization Science* 15(3):259275.

Besag J (1974) Spatial interaction and the statistical analysis od lattice systems. *Journal of the Royal Statistical Society Series B (Methodological)* 36(2):192–236.

Bhamidi S, Bresler G, Sly A (2011) Mixing time of exponential random graphs. *The Annals of Applied Probability* 21(6):2146–2170.

Blume LE (1993) The statistical mechanics of strategic interaction. *Games and Economic Behavior* 5(3):387–424.

Borgatti SP, Halgin DS (2011) On network theory. *Organization Science* 22(5):1168–1181, URL http://dx.doi.org/10.1287/orsc.1100.0641.

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

Brennecke J (2020) Dissonant ties in intraorganizational networks: Why individuals seek problem-solving assistance from difficult colleagues. *Academy of Management Journal* 63:743–778.

Butts C (2009) Using potential games to parameterize erg models, working paper.

Butts CT (2019) A dynamic process interpretation of the sparse ergm reference model. *The Journal of Mathematical Sociology* 43(1):40–57.

Byshkin M, Stivala A, Mira A, Robins G, Lomi A (2018) Fast maximum likelihood estimation via equilibrium expectation for large network data. *Scientific Reports* 8(11509).

Caimo A, Friel N (2011) Bayesian inference for exponential random graph models. *Social Networks* 33(1):4155.

Caimo A, Lomi A (2015) Knowledge sharing in organizations: A bayesian analysis of the role of reciprocity and formal structure. *Journal of Management* 41(2):665–691.

Chandrasekhar A, Jackson M (2014) Tractable and consistent exponential random graph models, stanford University.

Chung S, Singh H, Lee K (2000) Complementarity, status similarity and social capital as drivers of alliance formation. *Strategic Management Journal* 21(1):1–22.

Conley TG (1999) Gmm estimation with cross sectional dependence. *Journal of Econometrics* 92(1):1–45.

Conley TG, Topa G (2007) Estimating dynamic local interactions models. *Journal of Econometrics* 140(1):282–303.

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

Currarini S, Jackson MO, Pin P (2010) Identifying the roles of race-based choice and chance in high school friendship network formation. *the Proceedings of the National Academy of Sciences* 107(11):4857–4861.

Davis JP (2016) The group dynamics of interorganizational relationships: Collaborating with multiple partners in innovation ecosystems. *Administrative Science Quarterly* 61(4):621–661.

Eisenhardt KM, Schoonhoven CB (1996) Resource-based view of strategic alliance formation: strategic and social effects in entrepreneurial firms. *Organization Science* 7(2):136150.

Fleming L, Colfer L, Marin A, McPhie J (2004) Why the valley went first: Agglomeration and emergence in regional inventor networks. Working paper.

Fox J, Hsu D, Yang C (2018) Unobserved heterogeneity in matching games. *Journal of Political Economy* 126(4):1339–1373.

Gaonkar S, Moen M (2023) Standing on the parent's shoulder or in its shadow? alliance partner overlap between employee spinouts and their parents. *Strategic Management Journal* 44(2):415–440.

Garcia-Pont C, Nohria N (2002) Local versus global mimetism: The dynamics of alliance formation in the automobile industry. *Strategic Management Journal* 23(4):307–321.

Gelman A, Carlin J, Stern H, Rubin D (2003) *Bayesian Data Analysis, Second Edition* (Chapman & Hall/CRC).

Geyer C, Thompson E (1992) Constrained monte carlo maximum likelihood for depedendent data. *Journal of the Royal Statistical Society, Series B (Methodological)* 54(3):657–699.

Geyer CJ (1992) Practical markov chain monte carlo. *Statistical Science* 7:473–511.

Ghosh A, Ranganathan R, Rosenkopf L (2016) The impact of context and model choice on the determinants of strategic alliance formation: Evidence from a staged replication study. *Strategic Management Journal* 37(11):2204–2221.

Graham B (2017) An empirical model of network formation: with degree heterogeneity. *Econometrica* 85(4):1033–1063.

Gu Q, Lu X (2014) Unraveling the mechanisms of reputation and alliance formation: A study of venture capital syndication in china. *Strategic Management Journal* 35(5):739–750.

Gulati R (1995a) Does familiarity breed trust? the implications of repeated ties for contractual choice in alliances. *Academy of Management Journal* 38(1):85–112.

Gulati R (1995b) Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly* 40(619-652).

Gulati R (1999) Network location and learning: The influence of network resources and firm capabilities on alliance formation. *Strategic Management Journal* 20(5):397–420.

Gulati R, Gargiulo M (1999) Where do interorganizational networks come from? *American Journal of Sociology* 104(5):1439–1493.

Gulati R, Sytch M, Tatarynowicz A (2012) The rise and fall of small worlds: Exploring the dynamics of social structure. *Organization Science* 23(2):449–471.

Hallen BL (2008) The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments? *Administrative Science Quarterly* 53(4):685–718.

Hallen BL, Katila R, Rosenberger JD (2014) How do social defenses work? a resource-dependence lens on technology ventures, venture capital investors, and corporate relationships. *Academy of Management Journal* 57(4):1078–1101.

Heckman JJ (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4):931–959.

Heckman JJ, Vytlacil EJ (2007) Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. Heckman J, Leamer E, eds., *Handbook of Econometrics*, volume 6 of *Handbook of Econometrics*, chapter 70 (Elsevier).

Hochberg YV, Ljungqvist A, Lu Y (2007) Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance* 62(1):251–301.

Jackson M, Watts A (2001) The existence of pairwise stable networks. *Seoul Journal of Economics* 14(3):299–321.

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

Jackson MO (2008) *Social and Economics Networks* (Princeton).

Kang R, Zaheer A (2018) Determinants of alliance partner choice: Network distance, managerial incentives, and board monitoring. *Strategic Management Journal* 39:2745–2769.

Kim JY, Howard M, Pahnke EC, Boeker W (2015) Understanding network formation in strategy research: Exponential random graph models. *Strategic Management Journal* 37(1):22–44.

Kogut B, Urso P, Walker G (2007) Emergent properties of a new financial market: American venture capital syndication,1960-2005. *Management Science* 53(7):1181–1198.

Krivitsky PN, Handcock MS (2014) A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1):29–46.

Krivitsky PN, Kolaczyk ED (2015) On the question of effective sample size in network modeling: An asymptotic inquiry. *Statist. Sci.* 30(2):184–198.

Knig MD (2016) The formation of networks with local spillovers and limited observability. *Theoretical Economics* 11(3):813–863.

Lavie D (2007) Alliance portfolios and firm performance: A study of value creation and appropriation in the u.s. software industry. *Strategic Management Journal* 28:11871212.

Lerner J (1994) The syndication of venture capital investments. *Financial management* 16–27.

Levin DA, Peres Y, Wilmer EL (2008) *Markov Chains and Mixing Times* (American Mathematical Society).

Li SX, Rowley TJ (2002) Inertia and evaluation mechanisms in interorganizational partner selection: Syndicate formation among us investment banks. *Academy of Management JOURNAL* 45(6):1104–1119.

Liang F, Liu C, Carroll RJ (2010) *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples* (John Wiley and Sons, Ltd).

Lomi A, Pallotti F (2012) Relational collaboration among spatial multipoint competitors. *Social Networks* 34(1):101 – 111.

Lomi A, Pattison P (2006) Manufacturing relations: An empirical study of the organization of production across multiple networks. *Organization Science* 17(3):313–332.

Lucas RJ (1976) Econometric policy evaluation: A critique. Brunner K, Meltzer A, eds., *The Phillips Curve and Labor Markets*, volume 1 of *Carnegie-Rochester Conference Series on Public Policy*, 19–46.

Madhavan R, Gnyawali DR, He J (2004) Twos company, threes a crowd? triads in cooperative-competitive networks. *Academy of Management Journal* 47:918–927.

Manigart S, Lockett A, Meuleman M, Wright M, Landström H, Bruining H, Desbrières P, Hommel U (2006) Venture capitalists' decision to syndicate. *Entrepreneurship Theory and Practice* 30(2):131–153.

Marjoram P, Molitor J, Plagnol V, Tavar S (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* 100(26):15324–15328.

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

Mele A (2022) A structural model of homophily and clustering in social networks. *Journal of Business and Economic Statistics* 40(3):1377–1389.

Mele A, Zhu L (2023) Approximate variational estimation for a model of network formation. *The Review of Economics and Statistics* 105(1):113–124.

Meyn S, Tweedie RL (2009) *Markov Chains and Stochastic Stability* (Cambridge University Press).

Mindruta D, Moeen M, Agarwal R (2016) A two-sided matching approach for partner selection and assessing complementarities in partners' attributes in inter-firm alliances. *Strategic Management JOURNAL* 37(1):206–231.

Monderer D, Shapley L (1996) Potential games. *Games and Economic Behavior* 14:124–143.

Murray IA, Ghahramani Z, MacKay DJC (2006) Mcmc for doubly-intractable distributions. *Uncertainty in Artificial Intelligence* .

Nakajima R (2007) Measuring peer effects on youth smoking behavior. *Review of Economic Studies* 74(3):897–935.

Powell WW, Koput KW, Smith-Doerr L (1996) Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly* 41:116–145.

Reuer JJ, Lahiri N (2013) Searching for alliance partners: Effects of geographic distance on the formation of r&d collaborations. *Organization Science* 25(1):283–298.

Robins G, Snijders T, Wang P, Handcock M, Pattison P (2007) Recent developments in exponential random graph (p*) models for social networks. *Social Networks* 29:192–215.

Rothaermel FT, Boeker W (2008) Old technology meets new technology: Complementarities, similarities, and alliance formation. *Strategic Management JOURNAL* 29(1):47–77.

Shipilov AV, Li SX (2012) The missing link: The effect of customers on the formation of relationships among producers in the multiplex triads. *Organization Science* 23(2):472–491.

Snijders TA (2002) Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure* 3(2).

Snijders TA, Van de Bunt GG, Steglich CE (2010) Introduction to stochastic actor-based models for network dynamics. *Social networks* 32(1):44–60.

Sorenson O, Stuart T (2001) Syndication networks and the spatial distribution of venture capital investments. amer. *American Journal of Sociology* 106:1546–1588.

Sorenson O, Stuart T (2008) Bringing the context back in: settings and the search for syndicate partners in venture capital investment networks. *Administrative Science Quarterly* 53:266298.

Stern I, Dukerich JM, Zajac E (2014) Unmixed signals: How reputation and status affect alliance formation. *Strategic Management JOURNAL* 35(4):512–531.

Stivala AD, Koskinen JH, Rolls DA, Wang P, Robins GL (2016) Snowball sampling for estimating exponential random graph models for large networks. *Social Networks* 47:167–188.

Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge University Press).

Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*. *Psychometrika* 61(3):401–425.

Wilson R (1968) The theory of syndicates. *Econometrica: journal of the Econometric Society* 119–132.

Zaheer A, Soda G (2009) Network evolution: The origins of structural holes. *Administrative Science Quarterly* 54(1):1–31.

Zhang L, Gupta AK, Hallen BL (2017) The conditional importance of prior ties: A group-level analysis of venture capital syndication. *Academy of Management Journal* 60(4):1360–1386.

A   Technical details for the equilibrium model of network formation

In this appendix we provide more theoretical detail regarding the theoretical model of network formation used in the empirical analysis.

A.1   Setup

The economy consists of n firms and time is discrete ($t = 1, 2, 3, ...$). A generic firm i has a set of observable characteristics that we denote x_i. For example, x_i could include the size of the firm, its industry and location. We assume that each firm has M observable attributes, that is $x_i = \{x_{i,1}, x_{i,2}, ..., x_{i,M}\}$, and we denote as x the $n \times M$ matrix that contains all the vectors of firms' observable attributes.

The network of firms is represented by an $n \times n$ adjacency matrix g. The element at row i and column j will be denoted g_{ij}; we follow the convention in the literature and set $g_{ij} = 1$ if there is a link between firms i and j; otherwise $g_{ij} = 0$. In our application the network is *undirected* and each link requires mutual consent; therefore the adjacency matrix g is symmetric (that is, $g_{ij} = g_{ji}$ for all $i, j = 1, ..., n$). Most of the theoretical results below can be easily extended to directed networks.[26]

A.2   The network formation game

We assume that the network is formed sequentially over time and that firms maximize the surplus generated by each link. We distinguish between *opportunity* and *willingness* to form a link.

We model opportunity through a stochastic meeting process. In each period, two randomly selected firms, i and j can form (or delete) a link. This opportunity occurs with probability $\rho(g, x_i, x_j)$, which can depend on the existing network g. For example, firms with shared partners may have more

---

[26]See (Mele 2017) for a treatment of directed networks.

frequent chances to form alliances. Furthermore, the probability ρ can depend on the observable firms' characteristics x_i and x_j; for instance, firms with similar observable characteristics may have more opportunities to form partnerships.[27]

Upon receiving the opportunity to modify a link, firms i and j decide whether they want to update their connection g_{ij}. If the link does not exist, they decide whether to form a new link; if the link already exists, they choose to eliminate it. When choosing whether to update the link, companies behave myopically, maximizing the current surplus generated by their link.[28]

To characterize the equilibrium of the model and obtain a tractable and estimable likelihood, we make some assumptions on the payoffs and the probabilities $\rho(g, x_i, x_j)$ that govern the rate at which firms receive opportunities to create and delete links. Let g_{-ij} denote the network g with the exclusion of link g_{ij}.

Assumption 1 *The meeting process is i.i.d. over time, the probability that firms i and j meet is*

$$\rho(g, x_i, x_j) = \rho(g_{-ij}, x_i, x_j) > 0 \tag{11}$$

and the sum of these probabilities over all possible pairs of firms is one.

Assumption 1 guarantees that any pair of firms have an opportunity to update their links. The probability $\rho(g_{-ij}, x_i, x_j)$ can be very small, but it is necessarily positive. The main implication is that the network formation process can reach any equilibrium network with positive probability (Mele (2017)). The simplest probability model for ρ that satisfy Assumption 1 is a discrete uniform distribution. More generally, the probability ρ can depend on the network g_{-ij}. The crucial part of Assumption 1 is that the probability for any pair is positive.

---

[27]See Currarini et al. (2009) and Currarini et al. (2010) for a model where meetings are biased in favor of agents of the same group. Mele (2017), Badev (2021) and Chandrasekhar and Jackson (2014) also consider variants of this "meeting" technology.

[28]This modeling approach has been used in previous work by Nakajima (2007), Mele (2017), Mele and Zhu (2023), Badev (2021), Bala and Goyal (2000) and Jackson and Watts (2001) among others.

Firms' payoffs are across the networks, g, and observable characteristics x. We will denote as $U_i(g, x; \theta)$ the payoff of firm i from network g, observable characteristics x, and parameters $\theta = \{\alpha, \beta, \gamma\}$.

Assumption 2 *The payoff of firm i is*

$$U_i(g, x; \theta) = \sum_{j=1}^{n} g_{ij} \left[u(x_i, x_j; \alpha) + \beta \sum_{r \neq i, j}^{n} g_{jr} + \gamma \sum_{r \neq i, j}^{n} g_{jr} g_{ri} \right] \tag{12}$$

where α, β and γ are parameters.

The payoff $U_i(g, x; \theta)$ has three components. First, when firm i forms a link with firm j, it receives a net payoff $u(x_i, x_j; \alpha)$ that depends on characteristics x_i and x_j, and a parameter α. For example, a firm may find companies in the same industry more attractive for a partnership. The payoff $u(x_i, x_j; \alpha)$ includes both *costs and benefits* of direct connections, so it should be interpreted as *net direct benefit* of forming a link. We follow the convention in the strategic network literature and assume that *firms pay a cost for direct links, but that indirect connections are free* (Jackson and Wolinsky (1996), Jackson (2008)). In the empirical section we provide an explicit functional form for $u(x_i, x_j; \alpha)$.

Second, when firm i connects to a firm j, it receives an additional payoff β for each firm that has formed a link to j in previous periods. If firm j has 2 links, then firm i will receive 2β; if firm j has 5 links, then firm i will receive 5β. If $\beta > 0$, firm i prefers to link to a "popular" firm; viceversa, there could be a competition effect, and i may receive less utility from a "popular" firm since this firm has to share resources with other partners. The sign and magnitude of β is ultimately an empirical question. Third, firm i receives a payoff of γ for each partner in common with j. The term $\sum_{r \neq i, j}^{n} g_{jr} g_{ri}$ corresponds to the number of common partners between i and j.

This part of the payoff captures the clustering effects or triadic closure process. When γ is positive, a firm receives more surplus from companies with which it shares many partners. Viceversa, when γ is negative, a firm receives a negative surplus when linking to a company with many shared partners. In

the network literature, there is an empirical regularity: if two nodes have a common neighbor, there is a high chance that they form a link (Wasserman and Faust (1994), Wasserman and Pattison (1996), Jackson (2008)). This model accounts for this property through the payoff structure, generating the triadic closure property as an equilibrium outcome. On the other hand, we do not impose a positive parameter γ and instead let the data determine the sign. Therefore our model can accommodate both transitivity and intransitivity.

Finally, we assume that firms receive a joint matching shock $\varepsilon_{ij} = (\varepsilon_{0,ij}, \varepsilon_{1,ij})$ before choosing whether to update a link. The random shock models idiosyncratic reasons that could affect the decision to link. For example, in some periods, two firms may be a bad match (negative matching shock) for reasons that are unobservable to the researcher, such as misaligned long-term strategies or incompatible risk profiles. On the other hand, there are periods in which those companies may be a good match (positive matching shock), increasing both firms' willingness to create a partnership.

Assumption 3 *Firms receive a logistic matching shock before updating their links, which is i.i.d. over time and across pairs.*

Assumption 3 is standard in discrete choice models and random utility models in the empirical literature. In our model, this assumption is crucial to derive the likelihood of the network in closed-form,[29] allowing us to perform maximum likelihood estimation or Bayesian estimation.[30]

A.3   Characterization of Equilibrium Networks

We focus on equilibrium networks that satisfy *pairwise stability with transfers*, one of the most common equilibrium notions used in the network literature in economics.[31] This equilibrium notion requires that both firms consent to form a new link. However, the surplus generated by a new link is allowed

---

[29]See also Mele (2017), Mele and Zhu (2023), Chandrasekhar and Jackson (2014) and Heckman (1978).

[30]As an alternative to assumptions 1-3, we could use the spatial GMM (Conley (1999), Conley and Topa (2007)) or the Approximate Bayesian Computation (ABC). These techniques do not require knowing the likelihood in closed-form (Marjoram et al. (2003), Knig (2016)).

[31]See Jackson (2008), Mele and Zhu (2023) and Chandrasekhar and Jackson (2014) for examples.

to differ between the partners because firms can shift part of the other party's payoff. This is a way to model asymmetric bargaining power or different investments of resources in the partnership.

We will denote the transfer from firm i to firm j as τ_{ij}. By definition, $\tau_{ij} = -\tau_{ji}$. Conditional on being randomly selected, firms i and j will form (or keep) a link if the surplus generated by forming the link (including transfers and matching shocks) is larger than the surplus without the links, that is if

$$U_i(g_{ij} = 1, g_{-ij}, x_i, x_j; \theta) + U_j(g_{ij} = 1, g_{-ij}, x_j, x_i; \theta) + \varepsilon_{1,ij} \geq$$
$$U_i(g_{ij} = 0, g_{-ij}, x_i, x_j; \theta) + U_j(g_{ij} = 0, g_{-ij}, x_j, x_i; \theta) + \varepsilon_{0,ij} \tag{13}$$

In equation (13) the transfers do not appear; if i tranfers a positive amount τ_{ij} to j, then i's payoff will be $U_i(g_{ij} = 1, g_{-ij}, x_i, x_j; \theta) - \tau_{ij}$ while for j the payoff will be $U_j(g_{ij} = 1, g_{-ij}, x_j, x_i; \theta) + \tau_{ij}$. When we sum the payoffs to compute the total surplus of the link, the transfers cancel out. Therefore, the condition (13) can be re-stated as follows

$$U_i(g_{ij} = 1, g_{-ij}, x_i, x_j; \theta) + U_j(g_{ij} = 1, g_{-ij}, x_j, x_i; \theta) -$$
$$[U_i(g_{ij} = 0, g_{-ij}, x_i, x_j; \theta) + U_j(g_{ij} = 0, g_{-ij}, x_j, x_i; \theta)] \geq$$
$$\varepsilon_{0,ij} - \varepsilon_{1,ij}$$

Let's define

$$\Delta_{ij} := U_i(g_{ij} = 1, g_{-ij}, x_i, x_j; \theta) + U_j(g_{ij} = 1, g_{-ij}, x_j, x_i; \theta) -$$
$$[U_i(g_{ij} = 0, g_{-ij}, x_i, x_j; \theta) + U_j(g_{ij} = 0, g_{-ij}, x_j, x_i; \theta)]$$

Our Assumption 3 implies that the matching shocks are logistic. Therefore the difference of matching shocks $\varepsilon_{0,ij} - \varepsilon_{1,ij}$ is also logistic. The main consequence is that the probability that firms i and j

form (or keep) a link in any period has the standard logit form (Heckman (1978))

$$P(g_{ij} = 1|g_{-ij}, x_i, x_j, \theta) = \frac{\exp[\Delta_{ij}]}{1 + \exp[\Delta_{ij}]} \tag{14}$$

Equation (14) is a *conditional* probability. It describes the probability of a link between i and j; given the rest of the network's links g_{-ij}, the observable characteristics x_i and x_j, and the parameters of the model θ. This probability is also implicitly conditioned on the event that firms i and j have an opportunity to revise their link, which happens with probability $\rho(g_{-ij}, x_i, x_j)$. The probability (14) shows how the link g_{ij} depends on the other links in the network g_{-ij}, thus relaxing the independence assumption implicit in a standard logit model. Indeed, this is what strategic network formation implies: there is dependence among linking decisions, induced by the strategic equilibrium.

The model generates a sequence of networks as a result of link creation or deletion. This sequence is Markovian and converges to a unique stationary distribution that characterizes the probability of observing a specific network architecture in the long run, as shown in Mele (2017).

Let's define the aggregate function

$$Q(g, x; \theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} u(x_i, x_j; \alpha) + \frac{\beta}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r \neq i,j}^{n} g_{ij} g_{jr} + \frac{2\gamma}{3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r \neq i,j}^{n} g_{ij} g_{jr} g_{ri} \tag{15}$$

Equation (15) is called a *potential function*, and it summarizes the incentives of each firm to form links, net of the matching shock. The crucial property of the potential is

$$Q(g, x; \theta) - Q(g', x; \theta) = U_i(g, x; \theta) + U_j(g, x; \theta) - \left[U_i(g', x; \theta) + U_j(g', x; \theta) \right] \tag{16}$$

where g is a network where firms i and j have a link (that is $g_{ij} = 1$), and g' is the same network g, excluding the link between firms i and j (that is $g'_{ij} = 0$ and $g'_{-ij} = g_{-ij}$). Notice that the right-hand side of (16) represents the incentive of firms i and j to form the link; if the sum of their payoffs when they form the link is greater than the sum of their payoffs when they do not have a link, then they

will form the link (excluding the stochastic matching shock). The left-hand side of (16) shows that this difference in payoffs can be retrieved using the potential function. This property holds for any pair of firms i and j and for any link g_{ij}, $i, j = 1, ..., n$.[32]

This relationship means that the difference in potential functions (the left-hand side of equation (16)) can be computed using the profitable deviations of firms i and j. Therefore all the equilibrium networks can be found using the potential.

A network is *pairwise stable with transfers* if no two firms want to form a link or delete a link. In our model, all the *pairwise stable networks (with transfers)* correspond to the (local) maxima of the potential function (15).[33] Intuitively, let's consider a network that maximizes the potential function. Suppose we delete the link between firms i and j. This event will decrease the potential because the network was a maximizer of the potential. A decrease in potential means that the sum of payoffs of i and j is higher when the link exists than when we delete their link. Therefore, they will not be willing to delete their partnership. The same will hold if we consider an additional link between two firms in the network. Adding a link will decrease the potential and therefore implies that the firms involved in this relationship are better off not forming the new link. We can repeat this reasoning for any pair of firms, showing that indeed the network that maximizes the potential function is a pairwise equilibrium network with transfers. This result is important, and it facilitates the computation of the equilibria. Furthermore, at least a pairwise stable equilibrium with transfers is guaranteed by the potential function's existence, as shown in Monderer and Shapley (1996).

We should note that the potential function is different from a welfare function. Indeed if we assume

---

[32]The formal proof is contained in Mele (2017) and Mele and Zhu (2023).

[33]See Monderer and Shapley (1996), Mele (2017), Jackson and Watts (2001) and Badev (2021).

a classical welfare function, which is the sum of the firms' payoffs, we have

$$W(g, x; \theta) = \sum_{i=1}^{n} U_i(g, x; \theta) \tag{17}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} u(x_i, x_j; \alpha) + \beta \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r \neq i,j}^{n} g_{ij} g_{jr} + \gamma \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r \neq i,j}^{n} g_{ij} g_{jr} g_{ri} \tag{18}$$

As a consequence, the pairwise stable networks are not necessarily efficient. This is a standard result in the economics of networks literature. The decentralized equilibrium is inefficient because there are externalities in link formation; a firm that forms a direct link pays a cost and receives some benefits, while an indirect link has no cost but brings additional benefits (positive or negative). Therefore, forming a new link creates an externality for other firms that could benefit from a new indirect connection without paying any cost. This externality can be positive or negative, depending on the sign of the coefficients β and γ.

Using the potential function characterization we can write the conditional probability of linking as

$$P(g_{ij} = 1 | g_{-ij}, x_i, x_j, \theta) = \frac{\exp[\Delta_{ij}]}{1 + \exp[\Delta_{ij}]} =$$

$$= \frac{\exp[Q(g_{ij} = 1, g_{-ij}, x; \theta) - Q(g_{ij} = 0, g_{-ij}, x; \theta)]}{1 + \exp[Q(g_{ij} = 1, g_{-ij}, x; \theta) - Q(g_{ij} = 0, g_{-ij}, x; \theta)]} \tag{19}$$

$$= \frac{\exp\left[u(x_i, x_j; \alpha) + u(x_j, x_i; \alpha) + \beta \sum_{r \neq i,j}^{n} (g_{jr} + g_{ir}) + \gamma \sum_{r \neq i,j}^{n} (g_{jr} g_{ri} + g_{ir} g_{rj})\right]}{1 + \exp\left[u(x_i, x_j; \alpha) + u(x_j, x_i; \alpha) + \beta \sum_{r \neq i,j}^{n} (g_{jr} + g_{ir}) + \gamma \sum_{r \neq i,j}^{n} (g_{jr} g_{ri} + g_{ir} g_{rj})\right]} \tag{20}$$

Therefore, the potential fully characterizes the conditional choice probabilities of the firms.

The sequence of graphs generated by the network formation game is a Markov chain (Levin et al.

57

(2008), Meyn and Tweedie (2009)), converging to a unique stationary equilibrium distribution over networks, which we can characterize in closed-form as

$$\pi(g, x; \theta) = \frac{\exp\left[Q(g, x; \theta)\right]}{c(\theta, x)} \tag{21}$$

where $Q(g, x; \theta)$ is the potential function in (15) and $c(\theta, x)$ is a normalizing constant that sums over all possible networks with n nodes (we call this set $\mathcal{G}$)

$$c(\theta, x) = \sum_{\omega \in \mathcal{G}} \exp\left[Q(\omega, x; \theta)\right] \tag{22}$$

The normalizing constant (22) guarantees that equation (21) is a proper distribution, in other words, that it sums to one when we sum over all possible networks.[34]

Equation (21) is the likelihood of observing a particular network structure in the long run. Notice that this likelihood has peaked at the maxima of the potential function. Therefore, in the long run, we are most likely to observe the networks with high potential. From the discussion above, we know that all the pairwise stable equilibrium networks are maxima of the potential function; as a consequence, the model predicts that, in the long run, we observe equilibrium networks with very high probability.

To perform the estimation, we will assume that the network of firms observed in our data realizes the model's long-run stationary equilibrium. Accordingly, the distribution (21) is the likelihood of observing a particular network.

Finally, our model has a very important property. Let's consider the potential function (15) and its elements. The sum

$$t_S(g, x) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r \neq i, j}^{n} g_{ij} g_{jr} \tag{23}$$

---

[34]The proof of convergence and the computation of (21) can be found in Mele (2017) and Mele and Zhu (2023).

corresponds to the number of 2-stars in network g; while the sum

$$t_T(g,x) = \frac{2}{3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r \neq i,j}^{n} g_{ij} g_{jr} g_{ri} \tag{24}$$

is the number of triangles in network g. Thus the likelihood of the network formation game (21) corresponds to the likelihood of an ERGM with 2-stars and triangles. We can easily obtain alternative ERGM specifications by changing the structure of the payoffs in Assumption 2.

A.4 Extensions of the Network Formation Model

While this exposition has focused on a dynamic network formation model, characterizing the long-run distribution of networks in equilibrium, we can interpret this model in an alternative fashion. Consider a static network formation game, where each firm *simultaneously* chooses its portfolio of links. In such a game, multiple equilibrium networks are pairwise stable (with transfers). The stationary equilibrium of our model, described in the previous section, corresponds to a refinement of the static equilibrium, called *stochastic best-response dynamics* (Blume (1993), Jackson and Watts (2001)). According to this equilibrium refinement, firm pairs randomly encounter the opportunity to revise one of their link choices, but they make "mistakes" (modeled through random matching shocks ε_{ij}). The iteration of this stochastic best-response procedure generates the long-run distribution of networks (21). This refinement of the equilibrium concept can also be considered an equilibrium selection device.

The network formation game presented in this paper is a model for a single network observation. We can estimate the parameters (α, β, γ) by observing only a single network. However, if we have multiple networks or observe the network over time, we can exploit the information in the dynamics to identify richer sets of payoffs.

We can generalize the payoff functions used in this paper to include more externalities from link formation and accommodate externalities that also depend on each firm's observable characteristics. However, to precisely estimate such payoffs, we need a more extensive network or multiple network

observations.

B   Estimation details

We estimated all the models using the package `Bergm` in R, developed by Caimo and Friel (2011). All the computations have been performed on a desktop Dell Precision T7620 with 2 Intel Xeon CPUs E5-2697 v2 with 12 Dual-core processors at 2.7GHZ each and 64GB of RAM.

Fifty thousand parameter simulations yield each table of parameter estimates after discarding 5000 simulations as burn-in. For each parameter proposal, we simulate the model for 10000 iterations and pick the last simulated network to compute the exchange algorithm's acceptance ratio. We use the snooker algorithm with 10 parallel simulations to improve convergence as implemented in the package (see Caimo and Friel (2011).

Codes for estimation are available upon request.

B.1   Implementation of the exchange algorithm

The exchange algorithm shown in the estimation section is computationally intensive and requires some tuning.

The algorithm proceeds as follows. At each iteration $s = 1, 2, ...$ and current parameter θ_s

1. Propose a new parameter vector θ'

$$\theta' \sim q_\theta(\cdot|\theta_s)$$

2. Given the proposed parameter θ', simulate a network g^* as follows. At iteration r

(a) Propose a new network g'

$$g' \sim q_g(\cdot|g_r)$$

(b) Then update the network at iteration $r+1$

$$g_{r+1} = \begin{cases} g' & \text{with prob. } \alpha_g \\ g_r & \text{with prob. } 1 - \alpha_g \end{cases}$$

where α_g is

$$\alpha_g = \min\left\{1, \frac{\exp\left[Q(g', x; \theta')\right]}{\exp\left[Q(g_r, x; \theta')\right]} \frac{q_g(g_r|g')}{q_g(g'|g_r)}\right\}$$

(c) Iterate this process for $r = 1, ..., R$ and collect the last network $g_R = g^*$.

3. Update the parameter at iteration $s+1$

$$\theta_{r+1} = \begin{cases} \theta' & \text{with prob. } \alpha_{ex} \\ \theta_r & \text{with prob. } 1 - \alpha_{ex} \end{cases}$$

where the probability α_{ex} is given by

$$\alpha_{ex} = \min\left\{1, \frac{\exp\left[Q(g^*, x; \theta')\right]}{\exp\left[Q(g_r, x; \theta')\right]} \frac{\exp\left[Q(g_r, x; \theta_s)\right]}{\exp\left[Q(g^*, x; \theta_s)\right]} \frac{p(\theta')}{p(\theta_s)} \frac{q_\theta(\theta_s|\theta')}{q_\theta(\theta'|\theta_s)}\right\} \tag{25}$$

Notice that the probability (25) does not contain any normalizing constant. Neither κ nor $c(\theta, x)$ appear in the formulas; therefore, our simulations are feasible. The simulations of parameters converge to the posterior distribution of the structural parameters. All the technical details and the proofs of convergence are in Appendix B of Mele (2017).

The proposal distribution $q_\theta(\cdot|\theta_s)$ is a normal centered at the current parameter. To obtain the optimal variance for this proposal distribution, we estimate the model several times and adjust the

proposal distribution variance to obtain better acceptance rates. The exchange algorithm has low acceptance rates compared to a standard Metropolis-Hastings algorithm. Our estimates have acceptance rates of 13% and 18%, compared with an optimal (asymptotic) rate of 25% for the Metropolis-Hastings.

The proposal distribution for networks $q_g(\cdot|g_r)$ selects a random pair of nodes and proposes to end their link (if it exists) or form the link (if it does not exist). With a small probability, the proposed network swaps all the entries of the adjacency matrix. This step allows the sampler to reach other modes, extremely useful if the likelihood has multiple modes. For example, this is the case in several models analyzed in Mele (2017).

B.2   Additional estimation results

This appendix reports alternative estimates and goodness of fit tests for the model. These results were obtained using Maximum Pseudolikelihood estimation and Monte Carlo Maximum Likelihood estimation. The MPLE results are shown in Table 3. Notice that this is a frequentist approach and it delivers estimated parameters and standard error. We notice that the estimates are quite different from the posterior means shown in Table 1, especially the parameters for Number of common partners.

Table 3: Model with endogenous network variables, estimated with Maximum Pseudolikelihood (MPLE)

Variable	Estimate	Std. Error
Cost	-5.078	0.039
Number of partners	0.007	0.0001
Common partners	0.268	0.002
Same firm type	-0.154	0.031
Abs. Difference Capital (log)	-0.016	0.009
Abs. Difference Age	-0.016	0.001
Same state	0.510	0.037

The main advantage of the MPLE is that it is quite fast and usually converges relatively fast to the final estimate. This advantage is due to the assumption for the Maximum Pseudolikelihood

estimator that the conditional choice probabilities of forming links factorize into the pseudolikelihood: the underlying assumption is that conditional on the rest of the network, the links are independent. This assumption is not satisfied if we believe that the actors (firms) are strategic in their decision to form links. Theoretical results are showing that MPLE estimates are consistent. On the other hand, many practitioners complain of this estimator's poor performance, and the standard errors are imprecise and underestimated.

To give an idea of this estimator's poor performance, we provide the goodness of fit tests similar to those performed after the Bayesian estimation. We report our tests in Figure 10. We simulate 100 networks from the model, using the parameters in Table 3 and compare the observed network to the distribution of the simulated ones. We focus on three network statistics: the degree distribution, the edge-wise shared partners, and the geodesic distance. The solid line represents the observed values in our data, while the dotted lines are the 95% confidence levels for the simulated networks. The estimated model can replicate the observed network's feature if the solid line is within the dotted lines.

The picture shown in Figure 10 provide evidence that the MPLE is inadequate to fit these data, and performs worse than our Bayesian approach.

Figure 10: Goodness of fit, model with network effects, Maximum Pseudolikelihood estimates

Goodness−of−fit diagnostics
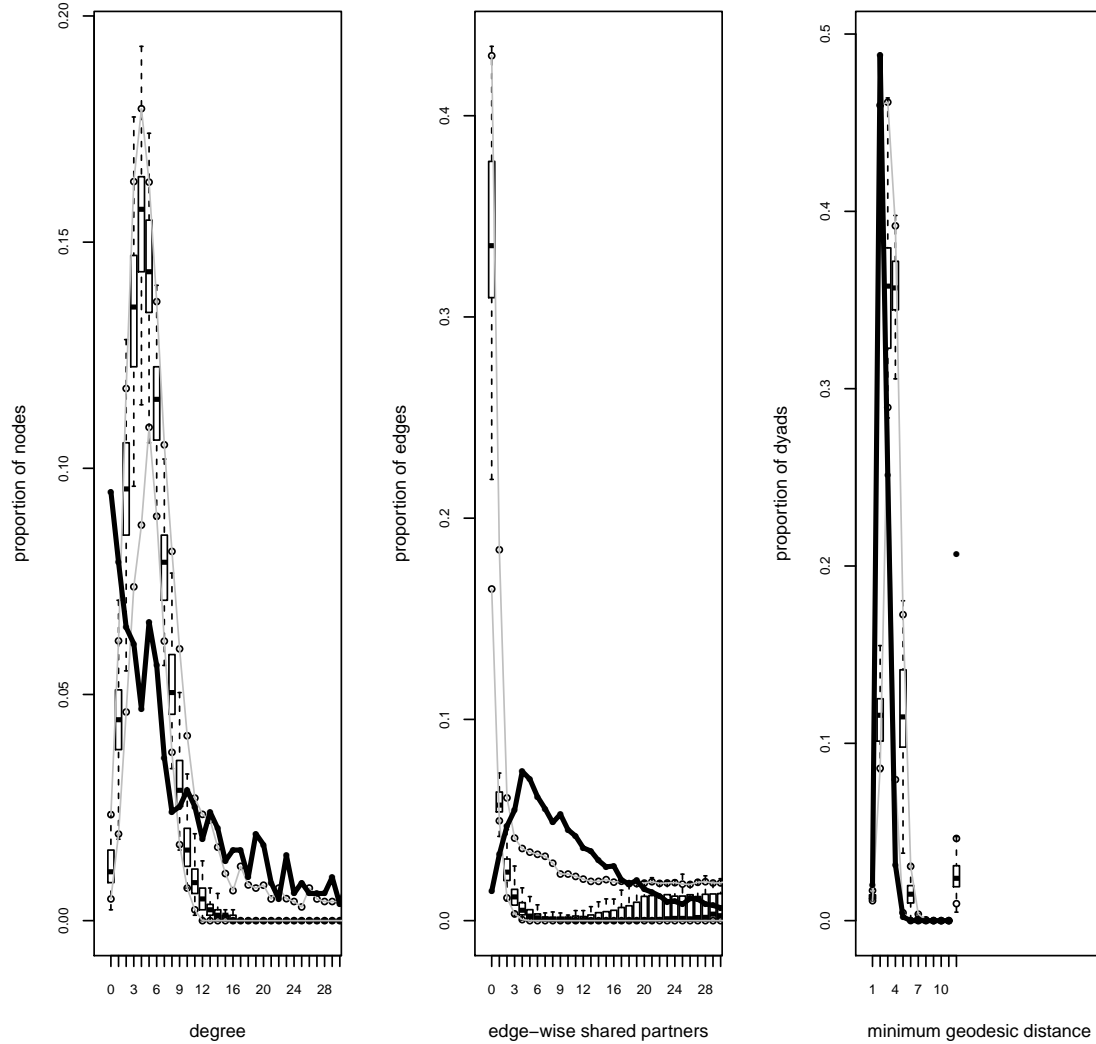
The black solid line represents the observed values of the degree distribution, distribution of edge-wise shared partners and minimu geodesic distance, respectively. The dotted lines represent the 95% confidence interval for the simulated model. The test is performed by simulating 100 networks from the equilibrium model. The parameters for the simulations are in Table 3. Each network is simulated with a MCMC run of 100000 steps.