Venture Capital: A Tale of Three Networks

Ke Shi*

Caltech

Feb 2024

Preliminary. Please do not quote.

Abstract

Networks play a key role in reducing risk and uncertainty and improving fund performances in the venture capital (VC) industry. However, the often-used coinvestment networks do not reflect the true social connections, i.e., the informal and personal ties between VC partners. In this paper, I connect three VC networks–coinvestment, past, and social–and study their impact on VC performances with a structural network model. To address the major endogeneity concerns in this setting, I exploit exogenous variations in VC partners' past connections through professional and alumni networks. Furthermore, to incorporate social networks, I endogenize network formations and structurally recover the unobserved, underlying social connections from VCs' equilibrium performance outcomes. I find that social networks have a significant effect on VC performances. Counterfactual experiments suggest that the industry suffers in terms of both welfare and equality from this reliance on personal connections.

*California Institute of Technology. Email: keshi@caltech.edu. I thank Michael Ewens, Gabriel Lopez-Moctezuma, Matthew Shum, Yi Xin, and seminar participants at Caltech, Workshop on Social Networks in Economics (Duke Kunshan), and Western Economics Association International (WEAI) Annual Conference for their valuable feedback.

1 Introduction

Networks are a prominent feature of the venture capital (VC) industry. Despite intense competition for the most promising entrepreneurial projects, VCs also collaborate extensively through coinvestment (or syndication). In a typical round of startup funding, several VCs jointly invest in the target company to spread the risk and share the responsibility of nurturing the startup. The VC industry is characterized by the *coinvestment networks* where, in the case of the US market, almost all but a handful VCs are path-connected through coinvestments. Underlying and parallel to this rich web of formal engagements are the *social networks*–personal and informal ties among VC partners. Anecdotal evidence suggests that the VC industry is a small and tight-knit community with frequent and meaningful interactions beyond coinvestments. Two VCs may have never coinvested in a startup, but could be well-acquainted and collaborative, offering each other resources and contacts. These social connections, in turn, are based on long-standing *alumni and professional networks* well before partners have entered the VC industry. Most VC partners have graduated from top colleges, gone to the handful prestigious business schools, and worked at large financial institutions and business corporations. The VC industry is often called the old boys' club because these past connections carry on to shape the industry today.[1]

Despite the importance of VC networks, their impact has not been adequately studied (Da Rin, Hellmann, and Puri 2013). Researchers generally agree that better-connected VCs also perform better (Tian 2011), yet the causal identification faces at least three empirical challenges. First, the study of networks in any empirical setting is plagued by endogeneity concerns. The key problem is simultaneity: Do networks cause superior performance or are they caused *by* superior performance (Da Rin, Hellmann, and Puri 2013)? Causality in both directions is plausible. The endogeneity issue is further complicated by omitted variables such as a VC's ability to source better deal flow and add value to its portfolio companies. More able VCs may be better at attracting other VCs to coinvest. Second, the existing empirical approaches generally lack micro-foundations, unable to pin down the marginal effect of social connections or attribute the performance of VCs to any specific ties. Building on graph theory and network theory in sociology, the literature on VC networks typically works with some measures of *network centrality*. This approach offers many powerful insights, but the difficulty lies in the interpretation of some centrality

[1]https://www.forbes.com/sites/oliversmith/2019/02/03/new-industry-report-exposes-british-vc-industry-as-an-old-boys-club

measures: We could only say something about the effect of network *position*, rather than connections themselves. For example, the betweenness and the eigenvector centralities are highly abstract measures, and estimates on these measures lack direct and intuitive interpretations. Third, systematic data on social networks are not available because we cannot observe the personal and informal ties between VCs. Although coinvestments are a good measure of connections, they could understate network effects because informal and personal ties also contribute to VC performance. The unobserved social networks could be denser than the observed coinvestment networks, giving rise to an inaccurate estimate of network spillover effects.

Before introducing my approach to addressing these challenges, it is helpful to first understand how networks help improve a VC's performance. The main mechanism in this study is the diffusion of information. Information is critical in the financial intermediary market that is permeated with uncertainty, risk, and information asymmetry. There are two channels through which information helps improve VC performance. The first pertains to the information that a VC acquires. VCs perform two important roles in entrepreneurship finance, screening (selecting companies with high growth potential) and value-adding (nurturing startups and helping them grow), both of which require significant informational inputs. First, startups are characterized by high degrees of risk and uncertainty. To alleviate this problem, VCs can exchange information and pool correlated signals to select better investments. Social connections also help diffuse information across sector and geographic boundaries and expand the reach of investments, allowing VCs to diversify their portfolio (Hochberg, Ljungqvist, and Lu 2007). Afterward diligent screening, VCs actively add value to their portfolio companies and help startups grow (Sørensen 2007) by sharing contacts, expertise, and resources. Having connections expands a VC's informational toolkit to guide its portfolio companies. In addition, strong relationships with other VCs improve the chances of securing follow-on funding for the startup. The second channel of information flow pertains to what VCs signal to startups in what is effectively a matching market. Entrepreneurship finance is best described as a two-sided matching market in which VCs and startups mutually select each other (Sørensen 2007). Promising startups are cautious in accepting capital investment due to the concern of control. Reputation of a VC is critical in the financial intermediary market because startups are forward-looking and care about the long-term implications when it accepts funding from a VC (Nahata 2008). Being well-connected is a signal to entrepreneurs that the VC is competent and reliable.

Given the information mechanism proposed above, I adopt a structural approach to address the aforementioned challenges and to identify the effect of network spillover. Following the network models by Battaglini, Leone Sciabolazza, and Patacchini (2020) and Battaglini, Patacchini, and Rainone (2021), I present three empirical models to deal with these problems progressively. The key idea of all three models is that given information flow through networks, the performance of a VC should be a function of the performance of other VCs with whom it has a connection. Being connected with better performing VCs improves one's own performance through the flow of information.

In the *baseline network model*, I focus on the micro-foundations and disregard the endogeneity problem. In what follows, I take VC performance as a proxy for the informational content that a VC possesses. That is, a VC performs better if and only if it has superior information. Given the mechanism, the structural model starts with a simple production function that relates a VC's performance to its peers' performances, in addition to the effort exerted by the VC and its characteristics. The production function reflects the assumption that if two VCs are connected, then they would share information and mutually enhance each other. VCs solve their optimal effort problem in rational expectation of the equilibrium performance of other VCs. It can be shown that there exists a unique Nash equilibrium in which the performances of all VCs are jointly determined in a system of equations, with performances on both sides of the equation mediated through networks. Equivalently, VC performances "solve" VCs' connections in equilibrium. Further, the performance of a VC corresponds to its alpha centrality in the network, lending to a nice interpretation that VC performances are direct measures, rather than the consequences, of network positions. To implement this model, I estimate the network spillover effect using the observed coinvestment networks directly. VCs are assumed to be endowed with these connections rather than actively choosing them. This approach is similar to much of the existing literature as far as data is concerned, essentially relating VC performances to coinvestment networks. The difference is my approach is supplemented with a micro-foundation that models how networks can improve performance.

To address the endogeneity problem in this empirical setting, I adopt a two-step estimation procedure that corrects for selection biases in a *Heckman-corrected network model*. I use the alumni and professional networks extracted from the LinkedIn profiles of VC partners to control for endogeneity. In the first step, pairwise coinvestment connections are explained by past alumni and professional connections as well as the distances between VCs in terms of their characteristics under the assumption of homophilly. The residuals from the first step are used as a regressor in the second step which is the same as the baseline network model, capturing individual heterogeneity that is unexplained by the coinvestment networks. This approach solves the simultaneity problem because the alumni and professional networks have taken shape before the coinvestment networks do and, therefore, cannot be affected by the latter. In addition, the alumni and professional networks control for individual unobservable factors such as the ability and socioeconomic background of VC partners. This approach enables a causal identification between VC networks and performance.

The analysis so far has ignored one important structural aspect of networks, namely that equilibrium performance should also induce connection formation. Coinvestment connections not only depend on past connections, but also VCs' rational expectation of their peer's performance. To address this concern, in the *endogenous network formation model*, I enrich the model with an explicit account for network formation and recover the unobserved social networks from the data. In this slightly more complicated formulation of a two-period game, VCs also choose their connections in rational expectation of the impact of these choices on their performances, taking others' performances as given in equilibrium. This structural model not only accounts for the source of endogeneity due to simultaneity, but also enables a direct recovery of the elasticity of network formation, that is, how performance induces network formation. Since the networks are now equilibrium objects recoverable from the data, I interpret them as the underlying social networks among VCs. The intuition is that the underlying social connections can be inferred from the variations in the observed performances, the past professional and alumni networks, as well as the distance between characteristics in network formation, assuming that the network model is correctly specified. For example. if two VCs are similar in their characteristics, have extensive past connections, and both perform well, then it is likely that they share a strong social tie; conversely, if one excels but the other falls behind, then it is likely that they have a weak connection. This approach has the advantage that it does not rely on coinvestment networks at all and allows us to distinguish the difference between the coinvestment networks and the social networks.

I present three sets of results following the outline above. First, in the baseline model, controlling for other determinants of VC fund performance such as fund size, I find that better-connected VCs also perform better in terms of the rate of successful portfolio exits. If a VC has a coivestor's exit rate increasing by 10%, or if it makes a new connection with an exit rate of 10%, the VC would enjoy a 0.1 percentage increase in its own exit rate. The magnitude of this estimates is comparable to the reduced-form regression relating performance to centrality measures. Exploiting potential sources of exogenous variation, I substitute the coinvesment networks with the professional or the alumni networks and achieve similar results, which suggests that there are substantial similarities between these networks in predicting VC performances.

Next, in the Heckman-corrected model, I first note that VCs display strong homophilly in pursuing coinvestment relations and prefer partners with similar characteristics in terms of fund size, industry of specialization, and demographic factors. Past connections including both professional and alumni networks are also important in coinvestment decisions, suggesting that the VC industry is indeed a tight-knit community with significant barriers to entry in terms of past connections. After controlling for both reverse causality and unobserved variables, I find that network spillovers are still present and the magnitude is similar to the baseline model and that the estimates on the unobserved individual heterogeneity is insignificant. This could suggest that professional and alumni networks do not contain additional information that are not already captured by the coinvestment networks.

Lastly, estimation of the endogenous network formation model simultaneously recovers the elasticities of connection impact and of network formation. I find that a one percent increase in the social connectedness of a VC, either in terms of establishing new connections or improving existing connections, lead to a 0.2 percentage rise in its own exit rate. Conversely, a one percent increase in the performance of a coinvestor induces a 0.74 percentage point increase in connection intensity. The model also recovers the underlying social networks from VC performances, past connections, and characteristics without referencing to the coinvestment networks. I find that there are many similarities between these two networks, indicating the validity and robustness of the structural model. Importantly, there are also interesting differences between the recovered social networks and the observed investment networks, suggesting that personal and informal ties, if the interpretation stands, are relevant in determining VC performances.

The remainder of this article is organized as follows. Section 2 describes the data and establishes reduced-form evidence consistent with Hochberg, Ljungqvist, and Lu (2007). Section 3 describes the structural model, econometric specification, and estimation method following Battaglini, Leone Sciabolazza, and Patacchini (2020) and Battaglini, Patacchini, and Rainone (2021). Section 4 presents the estimation results of all three network models. Section 5 concludes.

1.1 Related literature

The paper is the first attempt to structurally estimate VC networks and makes three distinct contributions to the VC literature. First, the paper provides a micro-foundation to the connection between VC networks and performance. Past work has generally relied on reduced-form approaches (Hochberg, Ljungqvist, and Lu 2007; Tian 2011). The conclusion is consistent with information theory in the finance literature, and generalizes both the value-adding and screening explanations in the VC literature (Sørensen 2007; Hochberg, Ljungqvist, and Lu 2007; Sorenson and Stuart 2001; Sorenson and Stuart 2008; Das, Jo, and Kim 2011). Second, the structural approach provides an alternative solution to the persistent endogeneity problem in the literature. Past work generally relies on empirical designs (Hochberg, Ljungqvist, and Lu 2007; Hochberg, Ljungqvist, and Lu 2010) to circumvent this problem. In addition, the structural approach also enables the simultaneous recovery of the impact of connections and performances on each other. Previous work generally focuses on the former (Sorenson and Stuart 2001; Sorenson and Stuart 2008) or the latter (Lerner 1994; Du 2016; Bubna, Das, and Prabhala 2020) but not both. Third, I provide a first attempt to characterize the social networks in the VC industry in addition to the coinvestment networks. Social networks are widely assumed to play a key role in many economically relevant environments, and notably, the financial intermediary market. The literature is scarce in this aspect.

2 Data and Reduced-Form Evidence

2.1 VC data

The VC data for our analysis come from VentureSource, a comprehensive portfolio of venture capital markets acquired from Dow Jones by CB Insights. It contains all the deal flows of US-based VCs from 1990 to 2009. I choose 2009 as a cutoff date because a typical VC fund has a life cycle of around a decade, so I do not yet have a full picture of the performances of recently established VCs. Each round of funding consists of a target company and a handful of VCs known as a syndicate. The target company could raise a few rounds of funding before exiting, and the VC investors do not have to remain the same through the funding rounds. For companies that have already exited, I observe its exit date and the way it exited, through either an initial public offering (IPO), an acquisition by some other company, or a failure (write-off from the VC's portfolio). Otherwise, the company is deemed active with the following exceptions. If an active company has not received another round of funding in the past five years, I assume that it has failed (but not yet officially declared a write-off), observing that the typical life cycle of a startup is not more than a decade. Furthermore, I only consider VC firms that are traditional in the sense that they are small partnerships devoted exclusively to funding startups. Investment banks, business conglomerates, healthcare companies are generally excluded from the dataset because their large size precludes meaningful identification of connections. Lastly, I only include VCs with at least one partner having a LinkedIn profile because professional networks and alumni networks are important sources of exogenous variation in this paper. The data consists of 15,777 rounds of funding involving a total of 670 VCs during the period. Table 1 presents summary statistics of the VCs used in our estimation.

VC performance

Following the literature (Das, Jo, and Kim 2011; Du 2016; Lindsey 2008; Hochberg, Ljungqvist, and Lu 2007), I define VC performance as the percentage of portfolio companies that have successfully exited the market either through an IPO or an acquisition by some other company. I use exit rate and VC performance interchangeably in this paper. Ideally, I would use direct data on the returns to investment of these VCs, but systematic data are not available because VC firms are not required by regulations to disclose their investment outcomes. The lack of data notwithstanding, VCs with higher exit rates would enjoy large returns on investment. Additionally, the exit rate is a good measure for this study because it is a value between zero and one, making the estimation easier in terms of convergence. 1 also presents summary statistics on the exit rate of VCs in the dataset.

Coinvestment networks

I construct the coinvestment networks based on VC deals. Recall that for each round of funding, I observe all VCs involved. The adjacency matrix $\mathbf{G}$ is defined as such: For a given pair of VCs i and j, g_{ij} is the number of coinvestments of the pair throughout the period observed in the data. Note that the networks are in a sense weighted, whereby g_{ij} represents the intensity or strength of the connection between i and j. Following convention, I let $g_{ii} = 0$, i.e. a VC is not considered connected with oneself. For the estimation, I also define two more metrics of connection intensity, a binary variable indicating if two VCs have ever coinvested, and a log transformed number of coinvestments. Table 2 presents summary statistics on VC networks. The unit of observation is an ordered pair of VCs. For the n VCs in the dataset, there are a total of $n^2 - n$ pairs of VCs.

Given the adjacency matrix, I define four centrality measures of the VCs following network and graph theory. Consider each VC a vertex and each connection an edge.

Table 1: Summary statistics of VC firms

	mean	sd	min	max
No. rounds	95.33	196.95	1.00	2373.00
No. startups	46.79	85.84	1.00	989.00
Years of experience	12.27	9.65	0.00	39.19
No. coinvestments	296.68	640.81	1.00	8075.00
No. coinvestors	101.95	145.65	1.00	1327.00
Performance				
No. IPOs	5.45	15.82	0.00	186.00
No. acquisitions	17.99	36.97	0.00	375.00
No. write-offs	14.70	24.20	0.00	285.00
No. private companies	8.65	16.16	0.00	148.00
IPO rate	0.08	0.14	0.00	1.00
Exit rate	0.47	0.27	0.00	1.00
Attributes				
Percent business and financial services	0.21	0.21	0.00	1.00
Percent consumer goods and services	0.15	0.19	0.00	1.00
Percent healthcare	0.21	0.31	0.00	1.00
Percent information technology	0.37	0.26	0.00	1.00
Percent female	0.15	0.27	0.00	1.00
Percent Asian	0.18	0.31	0.00	1.00
Centrality				
Degree	101.95	145.65	1.00	1327.00
Betweenness	0.00	0.00	0.00	0.06
Harmonic	0.39	0.05	0.22	0.57
Eigenvector	0.11	0.15	0.00	1.00

Notes: Summary statistics of VC characteristics based on VC deals data.

I define the following centrality of each vertex. The *degree centrality* of a vertex is the number of edges it is connected to. I do not make the distinction between inward and outward degrees in this study.[2] Given that the networks are weighted, I could adjust the degree centrality by the weights of the edges. The *betweenness centrality* of each vertex is the number of shortest paths that pass through the vertex. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that the number of edges that the path passes through (or the sum of the weights of the edges for weighted graphs) is minimized. Intuitively, betweenness measures the extent

[2]See Hochberg, Ljungqvist, and Lu (2007) for a detailed discussion of the distinction in the context of the VC industry.

Table 2: Summary Statistics of Pairwise Connection Intensities

	mean	sd	min	max
Coinvestment, G				
Having coinvested (Yes/No)	13.80	37.39	1.00	989.00
No. coinvestments	6.35	7.66	0.00	39.23
$\log$(1+ no. coinvestments)	6.35	7.66	0.00	39.23
Professional connections, $\mathbf{H}_p$				
Having professional connections (Yes/No)	1.69	7.24	0.00	186.00
No. professional connections	4.29	10.91	0.00	285.00
$\log$(1+ No. professional connections)	4.29	10.91	0.00	285.00
Alumni connections, $\mathbf{H}_a$				
Having alumni connections (Yes/No)	1.69	7.24	0.00	186.00
No. alumni connections	4.29	10.91	0.00	285.00
$\log$(1+ No. alumni connections)	4.29	10.91	0.00	285.00

Notes: The unit of observation is a VC-VC pair. The number of coinvestments is calculated based on the common funding round that both VCs participated in. Professional connections and alumni connections are calculated at the individual level and aggregated at the VC level. For example, if partner A from VC 1 and partner B from VC 2 have both worked at the same company prior to joining their respective VCs, this is one professional connection.

to which a vertex acts as a bridge in the networks. The *harmonic centrality* of a vertex is the average length of the shortest path between the node and all other nodes in the graph. (In a connected graph, the more common notion is called the closeness centrality.) Intuitively, harmonic measures how easy it is for a VC to reach other VCs in the entire industry. The *eigenvector centrality* is a measure of the influence of a node in a network based on the idea that connection to high-influence nodes increases the impact of oneself. The centralities can be characterized by a linear system $\lambda \mathbf{x} = \mathbf{G}\mathbf{x}$, where λ is the principal eigenvalue. Table 1 presents summary statistics on these measures.

Additionally, it is convenient now to introduce the concept of *alpha centrality* (sometimes also named after Katz 1953; Bonacich and Lloyd 2001) which will be useful later. The alpha centrality is an extension of the eigenvector centrality by incorporating external influences. Mathematically, it is solution to the system

$$\mathbf{x} = \delta \mathbf{G}\mathbf{x} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{\epsilon}$ is the amount of external influence that the nodes receive, and δ measures the relative importance between external influence against the importance of connectivity. We will encounter variations of equation (1) with different interpretations later. An alternative formulation of the alpha centrality is through a generalization of the degree centrality. It measures the number of all nodes that can be connected through a path, while the contributions of distant nodes are penalized.

Covariates

I also include relevant VC characteristics in our analysis as covariates. First, the performance of a VC is affected by its fund size and the industry the VC specializes in. I do not observe VC fund sizes directly but can observe the number of startups supported by the VC and the number of funding rounds participated. These variables are proxies for VC fund size, under the reasonable assumption that larger VCs generally invest in more startups. Second, VCs often specialize in one or a few industries of startups to avoid spreading resources too thin. In the data, there are four large industries: business and financial services, consumer goods and services, healthcare, and information technology. Third, I include two important demographic factor of VCs, the percentage of female partners and the percentage of Asian partners. The VC industry is dominated by white males, making it curious to understand how females and non-whites add to the dynamics. Gender and racial information is imputed based on partners' first and last names extracted from their LinkedIn profiles (to explained later). The algorithm is conservative in the sense that uncertainty resolves in favor of the male and the non-Asian. I focus on the Asian/non-Asian divide for two reasons. First, Asian first and last names are more identifiable compared to other minorities (black and Latino) in the US. Second, there is a sizable Asian community within the VC industry, giving rise to more data variation. The dataset suggests that as many as 10% of VC partners are Asian. Table 1 presents summary statistics on these covariates.

2.2 LinkedIn Data

I supplement the VC data with firm-level past professional and alumni networks based on information extracted from VC partners' LinkedIn profiles. LinkedIn is an online platform for professional networking and market. Individuals can voluntarily post their professional profiles on the site, including but not limited to work experience, education, skills, references, etc. The dataset is gather by a private company in 2017 by web-scraping all profiles on LinkedIn, including work experience and education background. I select individual profiles who have been partners and directors of the VCs in our dataset. One limitation of the LinkedIn dataset is that the experiences are self-reported and could potentially contain many errors or inaccuracies. I believe this is not a big issue on the large scale because partners have an incentive to accurately report their profiles. I have also taken care to remove spurious experiences from the dataset. Another issue is that some people do not have a LinkedIn profile, which tend to affect small VCs with fewer partners, so our analysis could understate the effect of alumni and professional networks. Note that the LinkedIn data is at the individual partner level, but for our purpose all data are aggregated at the VC firm level.

Professional networks

I construct the professional networks of VCs based on their partners' work experience. I define an adjacency matrix $\mathbf{H}_p$ where h_{ij} is the amount of connections that VCs i and j have based on how many work experiences their partners have shared in the past. If two partners from two VCs have ever worked in a same company, I consider it a shared experience. h_{ij} is a tally of all such shared experiences for all partners in VCs i and j. One difficulty is that two partners with experience at the same company may have never interacted or known each other. Hence the measure h_{ij} is noisy and potentially conflated. It is better to think $\mathbf{H}_p$ as a measure of the basis for networking and therefore a proxy of networks, rather than actual social links themselves. It is easier for partners with shared experience to network with each other. Table 2 presents summary statistics of pairwise work experience links based on LinkedIn data. Again, I present three metrics of the connection intensity: a raw count, a log transformed count, and a binary variable indicating if there's a connection at all.

Alumni networks

I similarly construct the alumni networks based on education background. I define the adjacency matrix $\mathbf{H}_a$ where h_{ij} is the amount of alumni connections that VCs i and j have based on how many pair of partners attended the same school. If two partners from two VCs have attended the same school, I consider it an alumni connection. h_{ij} is a tally of all such connections for all partners in VCs i and j. Table 2 presents summary statistics of alumni networks based on LinkedIn data.

2.3 Reduced-form evidence

I first demonstrate a correlation between VC performance and network position, following the empirical exercise by Hochberg, Ljungqvist, and Lu (2007). The econometric model is given in equation (2). The outcome variable $ExitRate_i$ is the exit rate of VCs measured by the percentage of successful exits through either an IPO and an acquisition by another company. The explanatory variable of interest is the centrality measure of various forms defined above. Note that Hochberg, Ljungqvist, and Lu (2007) adopts a time-lagged approach to address the endogeneity concern: For a fund of a given vintage year (the first year in which capital is delivered), measures of network centrality are constructed from coinvestment data for the 5 preceding years to eliminate reverse causality. The model here is a simplified version that does not address the endogeneity concern by incorporating the time dimension. I only intend to demonstrate the correlation and will deal with the endogeneity problem via a structural approach.

$$ExitRate_i = \gamma Centrality_i + X_i\beta + \epsilon_i \tag{2}$$

Table 3 presents the estimation results of equation (2) and will be the benchmark of other results in this study. As expected, all coefficients on the centrality measures are positive and significant. To be more concrete, having an additional connection (an increase in degree by 1) is associated with a 0.2 percentage point increases in exit rates, all else equal. Similarly, a one-standard-deviation increase (0.005) in the betweenness centrality is associated with a 5% increases in exit rates, all else equal. The magnitudes of these estimates are reasonable and generally consistent with those in Hochberg, Ljungqvist, and Lu (2007).

Before moving on, I note several issues with the reduced-form approach above. First, the above analysis suffers from critical endogeneity concerns due to both omitted variables and reverse causality. There could be unobserved characteristics like abilities that affect both network position and performances. Highly capable VCs could be better at both networking and bringing their portfolio companies to successful exits. In addition, superior performance could enable VCs to improve their network positions if VCs actively seek out better-performing peers to syndicate. More renowned and experienced VCs will find it easier to find coinvestors.

Second, the centrality measures (with the exception of degree) are highly abstract concepts that summarize social connections in a single index. They do not, however, tell us the marginal effect of having an additional connection because these measures are nonlinear in connections. In terms of interpretation, these measures are clumsy: I could only say something about the effect of a one-standard deviation increase in centrality, but not that of an addition of a connection or the improvement of my connections. The degree centrality does reflect the marginal impact, but only do so for the pair of connected individuals, not accounting for the impact it has on individuals who are further down the networks. In addition, it makes no distinction between the quantity and quality of connections. To un-

Table 3: Reduced-form evidence of network effect on VC performance

	Dependent variable:								
	Exit rate								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Degree	0.002^{***}					0.002^{***}			
	(0.0002)					(0.0004)			
Betweenness		9.857^{***}					15.972^{***}		
		(2.068)					(4.092)		
Harmonic centrality			8.517^{***}					6.517^{***}	
			(0.850)					(0.952)	
Eigenvector centrality				0.637^{***}					0.929^{***}
				(0.057)					(0.121)
No. startups					0.001^{***}	-0.001^{***}	0.002^{***}	0.001^{***}	-0.001^{***}
					(0.0001)	(0.0003)	(0.0003)	(0.0001)	(0.0003)
Percent business & finance					0.306^{***}	0.277^{***}	0.289^{***}	0.335^{***}	0.261^{***}
					(0.093)	(0.090)	(0.092)	(0.090)	(0.089)
Percent consumer G&S					0.190^{*}	0.164^{*}	0.177^{*}	0.184^{*}	0.150
					(0.099)	(0.097)	(0.098)	(0.096)	(0.095)
Percent healthcare					0.425^{***}	0.394^{***}	0.392^{***}	0.419^{***}	0.385^{***}
					(0.090)	(0.088)	(0.089)	(0.087)	(0.086)
Percent info tech					0.405^{***}	0.364^{***}	0.370^{***}	0.379^{***}	0.330^{***}
					(0.092)	(0.090)	(0.091)	(0.089)	(0.089)
Percent female					0.014	0.019	0.016	0.010	0.021
					(0.039)	(0.038)	(0.038)	(0.038)	(0.037)
Percent Asian					−0.016	−0.017	−0.016	−0.026	−0.019
					(0.034)	(0.033)	(0.033)	(0.032)	(0.032)
Constant	0.339^{***}	0.398^{***}	0.100^{***}	0.332^{***}	0.047	0.035	0.064	-0.168^{*}	0.049
	(0.013)	(0.012)	(0.033)	(0.013)	(0.083)	(0.080)	(0.082)	(0.086)	(0.079)
Observations	670	670	670	670	670	670	670	670	670
R^2	0.133	0.033	0.131	0.156	0.143	0.190	0.163	0.200	0.214
Adjusted R^2	0.131	0.031	0.129	0.155	0.134	0.181	0.153	0.190	0.204

Notes: Estimates of equation (2) of various specifications are presented. Columns (1)-(4) only use centrality measure as the explanatory variable. Columns (5)-(8) include additional covariates. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels.

derstand the marginal effect, I must first establish a model with a micro-foundation that relates performances to social connections themselves, rather than through a summary statistic. This will be the focus of the next section.

3 Structural Model, Specifications, and Estimation Method

Following Battaglini, Leone Sciabolazza, and Patacchini (2020) and Battaglini, Patacchini, and Rainone (2021), I present a condensed version of the structural model and the econometric specification. Readers are refereed to the appendix as well as the two papers for details. The key idea of the model is the equilibrium outcome in which the performances of all VCs are jointly determined. I present three network models in progression, each intended to address an empirical challenge outlined in the introduction. I start with laying out the micro-foundation of network models in the VC setting by introducing a production function that leverages on the insights from information economics.

3.1 Baseline Network Model

Production function

Financial intermediary markets are characterized by uncertainty, risk, and information asymmetry. To alleviate these problems, VCs perform two important roles in entrepreneurship finance: screening and value-adding. The startup industry is saturated with business ideas and new ventures, but only a tiny fraction of unicorns will eventually succeed. VCs must sift these projects diligently before committing both capital and effort. After a decision to invest, VCs continue to nurture and add value to their portfolio companies by providing guidance and resources, until such time arrives when they are viable enough to either make an IPO or be acquired by another company. Both of these roles require significant informational input: Screening is by definition an information game, whereas value-adding relies heavily on VCs having the right contacts and references. In a network setting, the content or strength of one's information depends on his connections, where both the quantity (what might be called the extensive margin) and the quality (the intensive margin) matter. This motivates a simple production function of the following form, in which a VC's performance depends on the performance of its connected peers, as well as the effort exerted and individual characteristics.

Consider a market that consists of n VCs with $\mathcal{N} = \{1, \cdots, n\}$. Each VC wants to maximize its performance, measured by the probability of bringing its portfolio companies into successful exits, which I call "effectiveness" and denote E_i. The effectiveness of each VC i, E_i, is a Cobb-Douglas production function with two inputs: its social connectedness s_i and its effort level l_i.

$$E_i = \rho s_i^{\alpha} l_i^{1-\alpha} + \epsilon_i \tag{3}$$

where $s_i = \sum_{j \in \mathcal{N}} g_{ij} E_j$ is an average of the effectiveness of the VCs that i has a connection with, weighted by g_{ij}, a measurement of the social link between i and j. g_{ij} could be a binary variable or a continuous variable (which can be interpreted as the connection intensity.) ϵ_i is the individual heterogeneity. Because of the Cobb-Douglas functional form, α is the elasticity of E_i with respect to s_i, that is, the responsiveness of a VC's effectiveness with respect to his social connectedness s_i. Notice that g_{ij} captures the quantity (extensive margin) of one's connections whereas E_j captures the quality (intensive margin).

Before moving on, I provide a quick interpretation of the mechanism through equation (3). The effectiveness E_i is the probability of VC's ability to bring a portfolio to a successful exit. This can be empirically observed and estimated by the exit rate of VC i. But how is one's effectiveness affected by his peers' effectiveness in a network? Several interpretations are consistent with the model. First, we can interpret E_i as the "signal"–to be slightly technical–that i receives in the screening game when VCs decide if they want to invest in a startup. Having more connections boost the accuracy of the signal so that the VC could make a more informed decision. Second, E_i could be interpreted as the social and human capital that the VC possesses during value-adding that largely boils down to knowledge and information, where having more connections allows the VC to offer more targeted help to their portfolio companies. Since VCs are not directly involved in the day-to-day running of the startups, what they do is providing contacts, references, and expertise. Third, E_i could be conversely interpreted as the reputation of VCs in a matching market where they also face reverse selection from startups. Reputation is key in any financial intermediary market, especially in the VC industry that is characterized by small partnerships with elusive outward appearance. Being well-connected to other reputable VCs gives VCs more exposure and signals to startups its strength and competence. I believe that all of these factor are relevant to the performance of VCs.

Equilibrium in the baseline model

Assume that the cost of effort is simply l_i. Under some regularity conditions and assumptions on the parameter space, we can solve for the optimal level of efforts and then derive a unique equilibrium. (Details are in the appendix.)

Proposition 1. *There is a unique equilibrium $\boldsymbol{E}$ characterized by the following autoregressive system*

$$\boldsymbol{E} = \delta \mathbf{G} \boldsymbol{E} + \boldsymbol{\epsilon} \tag{4}$$

where $\delta = \rho^{\frac{1}{\alpha}} (1-\alpha)^{\frac{1-\alpha}{\alpha}}$.

While the derivation is relegated to the appendix, equation (4) makes intuitive sense and says that being connected to effective individuals also make oneself effective. In comparing equation (4) with the definition of alpha centrality in (1), we see that $\boldsymbol{E}$ is exactly the alpha centrality with weight δ and exogenous influence $\boldsymbol{\epsilon}$. The difference is that whereas I exogenously calculate centrality measures as explanatory variables from the networks, this centrality measure $\boldsymbol{E}$ endogenously emerges out of the equilibrium model. This lends a nice interpretation to the model: performance or effectiveness is a direct measure of centrality itself, rather than a consequence of centrality. Being central is equivalent to being effective. In comparing model (4) with the reduced-form model (2), we see that the structural model explains the performance through connections directly, rather than indirectly through a centrality measure that is a summary for connections.

In terms of estimation, I make the parametric assumption that heterogeneity is a linear function of observable characteristics $\mathbf{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n]'$. It follows that

$$\boldsymbol{E} = \delta \mathbf{G} \boldsymbol{E} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{5}$$

System (5) is our baseline estimation model, and it corresponds to the linear structural model common in the networks literature. It is a spatial autoregressive (SAR) system that can be estimated easily using the maximum likelihood. Note that in this case, we make inference directly on the parameter δ that is a composite of ρ and α in the Cobb-Douglas production function. A direct regression does not allow me to recover both parameters. This issue will be dealt with shortly when I introduce Bayesian methods later. If it were the case that $\delta = 0$, then social spillover is absent and (5) is reduced to a simple linear regression on individual characteristics with no regard to networks.

To implement the regression, I use both the coinvestment networks $\mathbf{G}$ or the alumni or professional networks $\mathbf{H}_a$ and $\mathbf{H}_p$ as the adjacency matrix. The former is our benchmark model that relates coinvestment networks directly to VC performances. This approach suffers from critical endogeneity issue due to both simultaneity and omitted variables. The second method treats the past connections as exogenous and alleviate some endogeneity concerns. The trouble is that it does not shine light on the effect of current networks by regressing on the past networks. To resolve this issue, I connect the past and current networks by introducing a two-step estimation procedure much like the standard Heckman selection model.

3.2 Heckman-Corrected Network Model

As mentioned in the introduction, the baseline network model suffers from endogeneity problem due to simultaneity and omitted variables. A VC partner's intrinsic ability and his socioeconomic background, for example, could both affect performance and shape his networks. To address the endogeneity concern and account for the selection bias in network formation, I introduce a two-step Heckman correction procedure to equation (5). I use the alumni and professional networks as controls for the unobserved individual factors in the model. To preview my approach, the coinvestment connection between VCs i and j is first explained in terms of past connections between partners in VCs i and j through educational background and professional experiences. This first step accounts for the selection bias due to intrinsic ability and socioeconomic background, and the residual of this regression captures all unobservable characteristics in network formation not explained by the past.

For the causal inference to be valid, both the alumni and professional networks must be relevant and exogenous. The relevance condition is easy to see. Alumni and professional networks offer a well-established leverage for people to pursue professional opportunities in general. Shared experiences have long-lasting effects on the propensity to socialize later in life. I will demonstrate this pattern in the VC setting in the first-stage regression. The exogeneity condition relies on the assumption that past connections affect VC activities only through coinvestment. This assumption is a strong one and precludes past connections from acting in any informal engagement. As I have argued in the introduction, this is not the case since VCs also share personal and informal ties beyond coinvestment. I will deal with this issue in the third model, but for the moment we can concede that these instruments alleviates *some* endogeneity concerns due to intrinsic ability and socioeconomic background, but not necessarily personal ties. They are plausible albeit imperfect.

The first step is a standard dyadic model of link formation that explains coinvestment connections in terms of past connections as well as the distances between two VCs in terms of their characteristics.

$$g_{ij} = \gamma_0 + \gamma_1 h_{ij} + \sum_l \gamma_{l+1} d(X_i^l, X_j^l) + \eta_{ij} \tag{6}$$

where h_{ij} is the connection between i and j through work or college, and $d(X_i^l, X_j^l)$ is a distance metric measuring the difference between i and j in terms of characteristic l. Intuitively, the social connection between two VCs is increasing in their past connections, and decreasing in the distance between their characteristics.

In the second step, either a standard instrumental variable estimation with estimated $\hat{g}_{ij}$ or a Heckman correction procedure exploiting the residuals $\hat{\eta}_{ij}$ are possible. The first approach is difficult because the inference on standard errors are complicated, so I will adopt the second approach here. It requires an assumption on the covariance of the residuals ϵ and η_{ij}, which are outlined in the appendix. Roughly speaking, the correlation between unobserved characteristics determining link formation η_{ij} and unobserved characteristics driving the outcome ϵ_i must be the same across all VCs. The equilibrium is similarly characterized by the following modified system with an additional term.

$$\boldsymbol{E} = \delta \mathbf{G} \boldsymbol{E} + \mathbf{X} \boldsymbol{\beta} + \psi \boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{7}$$

where $\xi_i = \sum_{j \neq i} \eta_{ij}$. The term $\psi \boldsymbol{\xi}$ includes all unobserved characteristics of VC i and captures the selection bias in the link formation.

3.3 Endogenous Network Formation Model

One shortcoming of the exogenous variable approach is that it takes a naive approach to network formation based on homophilly but fails to account for the structural aspects. Suppose that a VC's performance improves due to some exogenous reasons, it is conceivable that other VCs would want to strengthen the tie because a closer connection with strong VCs would improve their own performances. This reverse causality is not modelled in the baseline network model (4) and treated away as an endogeneity problem. Further, the analysis above regards the observed coinvestment connections as all that is and all that matters in this empirical setting, but personal and informal links could also be present and play important roles. This consideration introduces a further source of endogeneity that cannot be accounted for in the Heckman-corrected model, and motivates a model of endogenous social network formation.

The setup of the model is similar except that it is a two-period game now. In the first period, VCs establish connections in rational expectations of the impact on their effectiveness. In the second period, agents choose their effort levels. The equilibrium levels of effectiveness are then realized. By backward induction, VCs would optimally choose their connections in the first period taking into consideration their impacts on their equilibrium effectiveness. The equilibrium is defined by the pure strategy $(\boldsymbol{g}_i, l_i)$, where $\boldsymbol{g}_i = (g_{i1}, \cdots, g_{in})$ maps the type of the VC to a vector of connection intensities, and l_i maps the type and the networks to the effort level.

The final piece of the setup is the cost of establishing social link $C(g_{ij}, \theta_{ij})$ in the first period on VC i's part. It is increasing in the strength of connection g_{ij} and decreasing in the compatibility θ_{ij} between i and j (to be specified below). The following parametric form is assumed and the parameter λ would have a convenient interpretation to be made clear shortly. For now, λ is a parameter that captures the curvature of the cost function and therefore measures the responsiveness of the i's choice of connection intensity to its peers' performance.

$$C(g_{ij}, \theta_{ij}) = \frac{\lambda}{1+\lambda}\left(\frac{g_{ij}}{\theta_{ij}}\right)^{1+\frac{1}{\lambda}} \tag{8}$$

Given the setup, Battaglini, Patacchini, and Rainone (2021) define the following equilibrium concept.

Definition 1. *Given the game described above, $(l, \boldsymbol{E}, \mathbf{G})$ is a network competitive equilibrium if*
1. $\boldsymbol{g}_i = (g_{i1}, \cdots, g_{in})$ is optimal at $t = 1$ given $\boldsymbol{E}$ (agents are price-taking),
2. l_i is optimal at $t = 2$ given $\boldsymbol{E}$ and $\mathbf{G}$, and
3. $\boldsymbol{E}$ satisfies the production function given l and $\mathbf{G}$ (price must clear the market).

Now, under some assumptions on the parameter space, the game can be solved by backward induction with a pure-strategy equilibrium. Further, under regularity conditions, the solution is interior and unique.

Proposition 2. *A network competitive equilibrium exists and is characterized by $\boldsymbol{E}$ with*

$$E_i = \varphi \sum_j (\theta_{ij} E_j)^{1+\lambda} + \epsilon_i \tag{9}$$

for all i, where φ is a function of ρ, α, and λ.[3] *The equilibrium level of intensities $\mathbf{G}$ are given by*

$$g_{ij} = \theta_{ij}^{1+\lambda}(\alpha\delta E_j)^{\lambda} \tag{10}$$

for all $i \neq j$. Under regularity conditions, the equilibrium is unique.

That is, the equilibrium effectivenesses are characterized by a system of nonlinear equations. Note φ measures the network spillover. Comparing system (9) with the baseline exogenous system (4), we see that the effectiveness are a modified version of the alpha centrality with the added nonlinear component λ on the index. The reason is that since the networks g_{ij} are endogenous, agents also choose their equilibrium connection intensity g_{ij} proportional to E_j^{λ} as in equation (10). Hence, E_i is a function of $E_j^{1+\lambda}$ in equilibrium.

[3]Detailed proofs are in the appendix. The parameter $\varphi = \alpha^{\lambda}\delta^{1+\lambda}$, where $\delta = \rho^{\frac{1}{\alpha}}(1-\alpha)^{\frac{1-\alpha}{\alpha}}$ is a intermediate parameter in the model identical to that in equation (4).

Note further that the elasticity of a link g_{ij} with respect to the E_j is precisely equal to λ given the parametric assumption above.[4] This lends to a convenient interpretation as we can simultaneously estimate the effect of connections on effectiveness and how effectiveness induces connection formation. Observe that if $\lambda = 0$, then equations (9) and (10) collapse into equation (4), the baseline network model. The parameter λ, therefore, measures the extent to which the endogenous network formation model fits the data better than the baseline network model and how much active networking happens in the VC industry.

For econometric specification, I assume that the individual heterogeneity is a linear function of characteristics. Then the equilibrium is characterized by the following system which is the main estimation equation of the third model.

$$E_i = \varphi \sum_j (\theta_{ij} E_j)^{1+\lambda} + X_i \beta + \epsilon_i \tag{11}$$

Lastly, θ_{ij}, the compatibility between VCs i and j, is modelled as Bernoulli random realization from a logistic function of χ_{ij}, a measure of connectivity between i and j, which in turn depends on past connections and the distance between i and j in terms of their characteristics. That is,

$$\begin{gathered} \mathbb{P}(\theta_{ij} = 1 | \chi_{ij}) = \frac{e^{\chi_{ij}}}{1 + e^{\chi_{ij}}} \\ \text{with } \chi_{ij} = \gamma_0 + \gamma_1 h_{ij} + \sum_l \gamma_{l+1} d(X_i^l, X_j^l) \end{gathered} \tag{12}$$

I make a final note on the difference between equation (12) and the link formation model in the first-step of Heckman-corrected network model in equation (6). In equation (12), the outcome variable is θ_{ij}, a measure of compatibility, whereas in the Heckman estimation, the outcome variable is g_{ij}, the connection itself. The similarity in form is deceiving: It should be clear that social networks g_{ij} are endogenously determined and recovered from the data in this model in equation (10), rather than the observed coinvestment networks. As we will see, there are both similarities and differences between the observed coinvestment networks and the recovered social networks.

[4] The elasticity of a link g_{ij} with respect to the effectiveness of j is

$$\epsilon_{g_{ij}, E_j} = \frac{\partial g_{ij}}{\partial E_j} \frac{E_j}{g_{ij}} = \theta_{ij}^{1+\lambda} (\alpha \delta E_j)^{\lambda - 1} \alpha \delta \frac{E_j}{g_{ij}} = \lambda$$

3.4 Estimation Methods

I now briefly describe the estimation strategy. The main estimation equations are the baseline network model (5), the Heckman-corrected model (7), and the endogenous network formation model (11). Although equations (5) and (7) are linear in E, ordinary least squares (OLS) estimation is not feasible because the variable E appears on both sides of the equation. The simultaneity of the autoregressive model would render an OLS estimation inconsistent. Instead, I will estimate equations (5) and (7) using nonlinear least squares (NLLS). Standard errors are bootstrapped with 500 replications.[5]

The endogenous network formation model (11) is highly nonlinear and is estimated using Bayesian methods. Specifically, I use the Approximate Bayesian Computation (ABC) (see Marjoram et al. 2003), a technique modified from the Metropolis-Hastings algorithm (see Metropolis et al. 1953; Hastings 1970). See section 4.2 in Battaglini, Patacchini, and Rainone (2021) as well as the appendix for more details on the procedure. I provide a loose description of the algorithm here. Starting from an initial value of the parameters ω, the algorithm proposes a candidate new parameter ω' according to some pre-specified transition kernel. If the proposed parameter ω' explains the observed data E well according to the equilibrium condition (11) and better than the current parameter ω does, then the algorithm moves to the proposed parameter ω'. The algorithm generates a Markov chain with a limiting stationary distribution, which, under the assumption that the model is correctly specified, coincides with the true conditional distribution of the parameter $P(\omega|E)$, the object of our interest.

4 Results[6]

4.1 Baseline Model Results

I begin the analysis by discussing the estimation results from the baseline network model. Recall that the baseline model relates VC performance on networks directly without explicitly accounting for endogeneity issues or the underlying social networks. Table 4 reports these results. As a benchmark, Column (1) reports the OLS regression of exit rate on VC characteristics directly, a naive model that disregards the effect of networks. Unsurprisingly, the exit rates is strongly predicted by the number of the startups that the VC

[5]The standard errors reported in this draft are asymptotic standard errors.
[6]Note that post-estimation statistics such as the R^2 are not reported in this draft.

supports. This is consistent with the literature that bigger VCs with larger funds are generally associated with better performances. Additionally, specialization in the healthcare and information technology industries are also associated with stronger performances, possibly due to the burgeoning of these industries in the past few decades.

Columns (2) through (4) present results from the baseline estimation model in equation (5). Column (2) uses the observed coinvestment connections $\mathbf{G}$ as the networks. Columns (3) and (4) use the professional and alumni connections $\mathbf{H}_p$ and $\mathbf{H}_a$ as the networks. Observed that the coefficient on the key parameter δ is positive and significant in all three models, indicating the presence of network spillover effects. There are several ways to interpret this result. At the intensive margin, if the exit rate of a VC's coinvestor increases by 10 percentage point, then the VC's own exit rate would increase by 0.1 percentage point, (assuming the equilibrium effects of these changes are negligible.) At the extensive margin, if the VC makes a new connection with an exit rate of, say 10%, then the VC's own exit rate would also increase by 0.1 percentage point. Compare this with the reduced-form result in table 3 relating exit rates to the degree centrality, where an additional coinvestor raises the exit rate by 0.2 percentage point. The magnitudes of the structural estimates are reasonable. Additionally, they allow a richer interpretation in terms of both the quantity and quality of social connections compared to the reduced-form estimates.

4.2 Heckman-Corrected Model Results

First step: link formation

For the Heckman-corrected network model, I begin the analysis by showing that the past networks are relevant to VC coinvestments. Table 5 presents the OLS estimation results of the first step coinvestment network formation in equation (6). Columns (1) and (2) use binary variables in the networks, whereas columns (3) and (4) use the raw number of connections. Perhaps unsurprisingly, there is strong evidence of homophilly as all coefficients on the distance between VC characteristics are negative and statistically significant. VCs are more likely to be coinvestors if they share similar characteristics including their fund size, industry of specialization, and demographics. In particular, VCs with a higher representation of female or Asian partners are more likely to syndicate with their alike, suggesting either that underrepresented groups prefer and trust their own group more in coinvestment decision, or that the VC industry is indeed an old boys' club difficult for newcomers to navigate.

Table 4: Estimation results of the baseline network model

	Dependent variable:			
	Exit rate			
	(1) No networks	(2) Coinvestment networks	(3) Professional networks	(4) Alumni networks
δ		0.00934***	0.00127**	0.0107***
		(0.00126)	(0.000548)	(0.000605)
No. startups	0.000997***	-0.00098***	0.000799***	-0.000469*
	(0.000131)	(0.000323)	(0.000159)	(0.000269)
Percent business & finance	0.306***	0.265***	0.295***	0.0694
	(0.0927)	(0.0882)	(0.0925)	(0.166)
Percent consumer G&S	0.19*	0.151	0.179*	-0.0839
	(0.0993)	(0.0945)	(0.0991)	(0.179)
Percent healthcare	0.425***	0.388***	0.415***	0.14
	(0.0898)	(0.0845)	(0.0896)	(0.162)
Percent information tech	0.405***	0.354***	0.392***	0.0845
	(0.0918)	(0.0873)	(0.0917)	(0.165)
Percent female	0.0143	0.0198	0.0097	-0.00277
	(0.0388)	(0.0374)	(0.0387)	(0.0695)
Percent Asian	-0.0163	-0.0212	-0.0195	0.049
	(0.0335)	(0.032)	(0.0334)	(0.06)
Constant	0.047	0.037	0.0444	-0.0701
	(0.0826)	(0.0782)	(0.0823)	(0.148)
Observations	670	670	670	670

Notes: Column (1) reports the OLS of exit rate on VC characteristics. Columns (2) through (4) report the estimates of equation (5), with the coinvestment networks, professional networks, and alumni networks, respectively. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels.

For the purpose of this paper, the more relevant result is the coefficient on professional and alumni networks. Note that both coefficients are positive and statistically significant. In columns (1) and (2), two VCs having alumni connections increases the probability of coinvestment by 5 percentage point, and having professional connections increases the probability of coinvestment by 13 percentage point. Past professional connections are more than twice as impactful as alumni connections are in inducing coinvestments. This is true in the specifications using raw number of connection counts. In columns (3) and (4), have one additional professional connection between two VCs leads to an increase of 0.03 coinvestments, whereas an additional alumni connection increase the number of

Table 5: First step in the Heckman-corrected model: coinvestment network formation

	Dependent variable:			
	If coinvest		No. coinvestments	
	(1)	(2)	(3)	(4)
Professional connections	0.131***		0.0278***	
	(0.00125)		(0.000582)	
Alumni connections		0.0461***		0.0154***
		(0.00065)		(0.000214)
No. startups (absolute distance)	-0.00367**	-0.00349*	-0.013	-0.011
	(0.00178)	(0.00179)	(0.0134)	(0.0134)
Percent business & finance (absolute distance)	-0.0124***	-0.0128***	-0.0595***	-0.0556***
	(0.00106)	(0.00106)	(0.00799)	(0.00796)
Percent consumer G&S (absolute distance)	-0.0189***	-0.019***	-0.083***	-0.0864***
	(0.000877)	(0.000883)	(0.00662)	(0.0066)
Percent healthcare (absolute distance)	-0.0251***	-0.0242***	-0.107***	-0.0994***
	(0.000782)	(0.000788)	(0.0059)	(0.00588)
Percent information tech (absolute distance)	-0.0102***	-0.011***	-0.0416***	-0.041***
	(0.0013)	(0.00131)	(0.00981)	(0.00978)
Percent female (absolute distance)	-0.0182***	-0.0169***	-0.129***	-0.113***
	(0.000565)	(0.000575)	(0.00425)	(0.00424)
Percent Asian (absolute distance)	-0.015***	-0.0156***	-0.108***	-0.0946***
	(0.000572)	(0.000578)	(0.00429)	(0.00429)
Constant	0.052***	0.0461***	0.268***	0.241***
	(0.000456)	(0.000517)	(0.00332)	(0.00334)
Observations	448900	448900	448900	448900

Notes: Results from the estimation of equation (6). Columns (1) and (2) uses the binary outcome of coinvestment as the outcome variable. Columns (3) and (4) uses the number of coinvestments as the outcome variable. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels.

coinvestments by 0.015.

Second step: networks controlling for endogeneity

Table 6 presents estimation results of the second step of the Heckman-corrected network model. As two benchmarks, column (1) reports the OLS estimates of exit rate on VC characteristics directly, while column (2) reports results from the baseline network model using the coinvestment connections in table 4. Columns (3) to (6) report the estimates of equation (7) after controlling for endogeneity using the additional bias term gathered from the first-step estimation. The estimate of the network spillover effect δ is positive and statistically significant across specifications, indicating the presence of network externalities. The estimates are quantitatively similar to the results in the baseline model.

Table 6: Estimation results of the Heckman-correct network model

	Dependent variable:					
	Exit rate					
	(1) No networks	(2) Baseline model	(3) Heckman (professional, binary)	(4) Heckman (alumni, binary)	(5) Heckman (professional, count)	(6) Heckman (alumni, count)
δ		0.00934***	0.00934***	0.00934***	0.00742***	0.00744***
		(0.00126)	(0.00126)	(0.00126)	(0.000394)	(0.000394)
No. startups	0.000997***	-0.00098***	-0.000979***	-0.00098***	-0.00466***	-0.00465***
	(0.000131)	(0.000323)	(0.000323)	(0.000323)	(0.000263)	(0.000262)
Percent business & finance	0.306***	0.265***	0.265***	0.266***	0.328***	0.332***
	(0.0927)	(0.0882)	(0.0884)	(0.0885)	(0.115)	(0.115)
Percent consumer G&S	0.19*	0.151	0.15	0.152	0.307**	0.309**
	(0.0993)	(0.0945)	(0.0947)	(0.0948)	(0.123)	(0.123)
Percent healthcare	0.425***	0.388***	0.388***	0.389***	0.155	0.159
	(0.0898)	(0.0845)	(0.0847)	(0.0847)	(0.108)	(0.108)
Percent information tech	0.405***	0.354***	0.353***	0.355***	0.361***	0.365***
	(0.0918)	(0.0873)	(0.0877)	(0.0877)	(0.113)	(0.114)
Percent female	0.0143	0.0198	0.0198	0.0198	0.0754	0.0748
	(0.0388)	(0.0374)	(0.0374)	(0.0374)	(0.0491)	(0.0491)
Percent Asian	-0.0163	-0.0212	-0.0211	-0.0212	0.0691*	0.0689*
	(0.0335)	(0.032)	(0.032)	(0.032)	(0.0405)	(0.0406)
Unobservables (ξ)			-3.55e-05	5.63e-05	-9.85e-05	-8.77e-05
			(0.000367)	(0.000357)	(6.61e-05)	(6.74e-05)
Constant	0.047	0.037	0.0376	0.0361	0.0934	0.09
	(0.0826)	(0.0782)	(0.0785)	(0.0785)	(0.103)	(0.103)
Observations	670	670	670	670	670	670

Notes: Column (1) reports the OLS of exit rate on VC characteristics. Column (2) reports the estimates of equation (5). Columns (3) to (6) report the estimates of equation (7), with the professional networks and alumni networks, either binary or raw count, respectively. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels.

Furthermore, the estimates on the unobserved individual-level factors ξ are not statistically significant. This result suggests that the professional and alumni networks do not contain additional information not already captured by the coinvestment networks.

4.3 Endogenous Network Formation Model Result

Lastly, I discuss the estimation results of the endogenous network formation model. Tables 7 and 8 present the median value of the posterior distributions of the estimation of the endogenous model described in section 3.2. Table 7 presents the the median values of the posterior distributions of parameters φ, ρ, α, λ, and β of the network competitive equilibrium characterized by equation (11). Table 8 shows the median values of the posterior distributions of parameters γ in the first-stage network formation model in equation (12). The tables also report in brackets the empirical p-value of zero on the estimated posterior distribution (as opposed to the standard error that is typically reported). Statistical significance correspond to values near 1 or 0: If the p-value is equal to 1, the support of the empirical posterior distribution is greater than zero, whereas if the p-value is equal to 0, the support of the empirical posterior distribution is less than zero.

Before discussing the results, it is important to remember that both VC performance and networks are equilibrium outcomes, so the following interpretation of the parameters assumes that any small changes have little cascading impact on the entire networks. I begin the analysis by discussing φ in the context of equation (11). It is positive and statistically significant, indicating the presence of social spillover. The magnitude of φ, however, is not easy to interpret because of the nonlinear term $E_j^{1+\lambda}$ involved. As an example, if the exit rate of a VC's coinvestor increases from 10% to 20%, after accounting for endogenous network formation, the VC's own performance would increases by 0.09 percentage point. If the VC makes a new connection with a 10% exit rate, its own exit rate would rise by 0.08 percentage point.

The parameters α and λ are easier to interpret given the structure of the model. Recall that α is the elasticity of effectiveness of social connectedness with respect to the effectiveness of the others. A one percent increase in the social connectedness, measured by the weighted sum of the effectiveness of his peers of i, $\sum g_{ij}E_j$, either at the intensive or the extensive margin, induces a 0.24 percentage point increase in the effectiveness E_i. On the other hand, λ is the elasticity of connection intensity g_{ij} with respect to effectiveness of its peers E_j. That is, a one percent increase in E_j leads to a marginal increase in g_{ij} by 0.74 percentage point, assuming that the impact on others are negligible in equilibrium. This latter indicates that VCs do strongly respond to the effectiveness of their peers and actively seek out better connections.

Table 7: Results from the endogenous network model

	Dependent variable:
	Exit rate
φ (Social spillover)†	0.0002***
	[1.0000]
ρ (Cobb-Douglas coefficient)	0.2700***
	[1.0000]
α (Elasticity of connection impact)†	0.2403***
	[1.0000]
λ (Elasticity of network formation)†	0.7411***
	[1.0000]
No. startups	0.0010***
	[1.0000]
Percent business & finance	0.3395***
	[1.0000]
Percent consumer G&S	0.2165***
	[1.0000]
Percent healthcare	0.4897***
	[1.0000]
Percent information tech	0.4705***
	[1.0000]
Percent female	0.0148***
	[1.0000]
Percent Asian	-0.0232***
	[0.0000]
Pseudo-R^2	0.8352
Penalized pseudo-R^2	0.8341
MSE	0.1648
MASD	0.4320
Observations	670

Notes: $^\dagger\alpha$ is the elasticity of performance E_i with respect to the social connections $\sum g_{ij}E_j$. λ is the elasticity of link g_{ij} with respect to the performance of j, E_j. φ is calculated based on the estimates of ρ, α, and λ. Estimates of parameters in equation (11) are reported in column (1). The median of the posterior distribution estimated with the ABC algorithm is reported for each parameter. The empirical p-value of zero on the estimated posterior is reported in the brackets. p-value is equal to 1 if the support of the empirical posterior distribution is greater than zero, whereas p-value is equal to 0 if the support of the empirical posterior distribution is less than zero. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels based on empirical p-values.

Table 8 shows characteristics that matter for social connections in the VC industry. Again, I find strong evidence of homophilly, that VCs display strong preferences for peers with similar characteristics in terms of demographics and industry of specialization. One unexpected result is that the coefficient on distance between the number of startups is positive, suggesting that VCs could in fact prefer to interact with those with different fund sizes. A possible explanation is that VCs might attempt to reduce their risk exposure by interacting with others of different characteristics, therefore receiving valuable information otherwise not available to them. Big and small VCs could be complementary in their expertise, skills, and contacts, so having personal ties with VCs on the other end of the spectrum expands one's own information set.

For our purpose, the most relevant result is the coefficient on past connections, and it is indeed positive and significant. (The magnitude is not easy to interpret because of the parametric form in equation (12).) Recall that this set of parameters are recovered from VC performances and embedded in the structural model of social networks rather than the observed coinvestment networks, so it is quite remarkable that they confirm our intuition. Relying on past connections and VC characteristics alone, the model identifies the effect of social networks on VC performances without referencing coinvestment networks at all. The social networks recovered from the model, in fact, are not the same as the coinvestment networks. This is a direction that I intend to further pursue later on.

Lastly, I compare the endogenous network formation model with two benchmarks. The first is the standard model without networks that relates exit rates to VC characteristics, a model similar to a direct OLS estimation. In terms of implementation, this is equivalence to imposing $\rho = 0$ in the production function (3) and consequently $\varphi = 0$ in equation (11). The second benchmark is the case where performance depends on networks but the networks are exogenous, i.e. VCs do not endogenously choose their connections. This is equivalent to imposing the elasticity of network formation $\lambda = 0$. Then g_{ij} is identical to θ_{ij} in equation (10) and the system (11) is reduced to the baseline exogenous network model in equation (5). Table 9 columns (1) and (2) report the results from the two benchmark estimations. Note that the estimates are different from the baseline estimation results because the underlying network is still recovered from the data rather than imposed as the coinvestment networks.

For future work, I intend to compare if the endogenous network model produce better fit than the exogenous version. To complete the analysis, I would also estimate the model using the professional and alumni networks as a proxy for social connections following the baseline and Heckman-corrected network models. These two are not degenerate ver-

Table 8: Results of link formation in the endogenous network model

	Dependent variable:
	Compatibility
Professional connection	1.3400***
	[1.0000]
No. startups (absolute distance)	0.0039***
	[1.0000]
Percent business & finance (absolute distance)	-4.1258***
	[0.0000]
Percent consumer G&S (absolute distance)	-3.1104***
	[0.0000]
Percent healthcare (absolute distance)	-0.8625***
	[0.0000]
Percent information tech (absolute distance)	-1.9955***
	[0.0000]
Percent female (absolute distance)	-0.1731***
	[0.0000]
Percent Asian (absolute distance)	-0.3480***
	[0.0000]
Constant	-1.8462***
	[0.0000]
Observations	448,900

Notes: Estimates of parameters in equation (12) are reported in column (1). The median of the posterior distribution estimated with the ABC algorithm is reported for each parameter. The empirical p-value of zero on the estimated posterior is reported in the brackets. p-value is equal to 1 if the support of the empirical posterior distribution is greater than zero, whereas p-value is equal to 0 if the support of the empirical posterior distribution is less than zero. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels based on empirical p-values.

sions of the endogenous model, but can be seen as alternative ways to control for the endogeneity problems.

Finally, we compare the exogenous versus the endogenous approaches. Results are reported in table 10.

5 Conclusion

The VC industry is characterized by risk, uncertainty, and information asymmetry. Networks prove to be a solution to VC's information problems by facilitating the exchange

Table 9: Comparison between the main estimation and two benchmarks

	Dependent variable:		
	Exit rate		
	(1) No networks	(2) Exogenous networks	(3) Endogenous networks
φ (Social spillover)†	-	0.0012***	0.0002***
		[1.0000]	[1.0000]
ρ (Cobb-Douglas coefficient)	-	0.1153***	0.2700***
		[1.0000]	[1.0000]
α (Elasticity of networks)†	-	0.0555***	0.2403***
		[1.0000]	[1.0000]
λ (Elasticity of network formation)†	-	-	0.7411***
			[1.0000]
No. startups	0.0010***	0.0011***	0.0010***
	[1.0000]	[1.0000]	[1.0000]
Percent business & finance	0.3539***	0.3300***	0.3395***
	[1.0000]	[1.0000]	[1.0000]
Percent consumer G&S	0.2403***	0.2298***	0.2165***
	[1.0000]	[1.0000]	[1.0000]
Percent healthcare	0.4730***	0.4323***	0.4897***
	[1.0000]	[1.0000]	[1.0000]
Percent information tech	0.4546***	0.4581***	0.4705***
	[1.0000]	[1.0000]	[1.0000]
Percent female	0.0154	0.0108***	0.0148***
	[0.6571]	[1.0000]	[1.0000]
Percent Asian	-0.0158	-0.0172***	-0.0232***
	[0.3185]	[0.0000]	[0.0000]
Observations			

Notes: $^\dagger\alpha$ is the elasticity of performance E_i with respect to the social connections $\sum g_{ij}E_j$. λ is the elasticity of link g_{ij} with respect to the performance of j, E_j. φ is calculated based on the estimates of ρ, α, and λ. Estimates of parameters in equation (9) are reported in column (3). Column (1) reports the estimates with the constraint $\lambda = 0$. Column (2) reports the estimates with the constraint $\rho = 0$. The median of the posterior distribution estimated with the ABC algorithm is reported for each parameter. The empirical p-value of zero on the estimated posterior is reported in the brackets. p-value is equal to 1 if the support of the empirical posterior distribution is greater than zero, whereas p-value is equal to 0 if the support of the empirical posterior distribution is less than zero. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels based on empirical p-values.

and flow of information. The study of VC networks, however, is plagued by endogeneity issues. In this paper, I have presented a structural model to address these problems. To the best of my knowledge, this is the first study that examines VC networks from a structural perspective. I find that better-connected VCs indeed realize better performances, as measured by the proportion of portfolio investments that successfully exited through

Table 10: Comparison between the endogenous model and the exogenous model

	Dependent variable:			
	Exit rate			
	(1) SAR	(2) Two-step I	(3) Two-step II	(4) Endogenous
φ (Social spillover)†			0.5052***	0.0015***
			[1.0000]	[1.0000]
ρ (Cobb-Douglas coefficient)			0.5057***	0.4790***
			[1.0000]	[1.0000]
α (Elasticity of networks)†			0.9999***	0.1430***
			[1.0000]	[1.0000]
λ (Elasticity of network formation)†	-	-	-	0.0896***
				[1.0000]
Number of rounds		-3.5286***	-3.4840***	-2.0082***
		[0.0003]	[0.0003]	[0.0000]
Number of startups		1.9522**	1.9172**	1.3973***
		[0.9806]	[0.9780]	[1.0000]
Experience (year)		1.3841***	1.3828***	0.6822***
		[1.0000]	[1.0000]	[1.0000]
Pseudo-R^2		0.7918	0.7944	0.8352
Penalized pseudo-R^2		0.7883	0.7911	0.8341
MSE		0.2082	0.2056	0.1648
MASD		0.2564	0.2576	0.4320
Observations				

Notes: † α is the elasticity of performance E_i with respect to the social connections $\sum g_{ij} E_j$. λ is the elasticity of link g_{ij} with respect to the performance of j, E_j. φ is calculated based on the estimates of ρ, α, and λ. Estimates of parameters in equation (9) are reported in column (4). In columns (1) to (3), $\lambda = 0$. In column (1), $\Theta = H$ the alumni network which reduces the model to equation (5). In column (2), $\Theta = G$ and ξ is added as a regressor as in equation (7). In column (3), $\Theta = \hat{G}$ where $\hat{G}$ is derived based on the dyadic model in equation (6), which makes the model equivalent to model (**??**). The median of the posterior distribution estimated with the ABC algorithm is reported for each parameter. The empirical p-value of zero on the estimated posterior is reported in the brackets. p-value is equal to 1 if the support of the empirical posterior distribution is greater than zero, whereas p-value is equal to 0 if the support of the empirical posterior distribution is less than zero. *, **, and *** indicates statistical significance at the 10, 5, and 1% levels based on empirical p-values.

an IPO or an acquisition by another company. Furthermore, I find strong evidence for the presence of social networks that explain VC performances even without referencing coinvestment connections. My study suggests that both coinvestment and informal and personal ties play important roles in the industry.

Three methodological contributions are at the center of my study. First, I bridge the gap between information, network, and performance by introducing a parsimonious model with a simple production function. This micro-foundation grounds my analysis and subsumes many theories in explaining the role and goal of VCs as important financial intermediaries. Second, I exploit past professional and alumni networks as possible sources of exogenous variation in the model to partially address the endogeneity issues. The results demonstrate that the VC industry is in many ways an exclusive club with substantial barriers to entry. Third, I adopt an endogenous network formation model that recovers the underlying social networks from VC performances alone. I would highlight that this last method is the pièce de résistance of this study because social networks have not been closely examined in the literature. The intuition is performances, past connections, and characteristics are sufficient information to infer the structure of social networks. The results demonstrate that the recovered social networks share many similarities with the coinvestment networks, but also contain many interesting differences.

References

Battaglini, Marco, Valerio Leone Sciabolazza, and Eleonora Patacchini (2020). "Effectiveness of Connected Legislators". *American Journal of Political Science* 64.4, pp. 739–756.

Battaglini, Marco, Eleonora Patacchini, and Edoardo Rainone (2021). "Endogenous Social Interactions with Unobserved Networks". *The Review of Economic Studies*, rdab058.

Bonacich, Phillip and Paulette Lloyd (2001). "Eigenvector-like measures of centrality for asymmetric relations". *Social Networks* 23.3, pp. 191–201.

Bubna, Amit, Sanjiv R. Das, and Nagpurnanand Prabhala (2020). "Venture Capital Communities". *Journal of Financial and Quantitative Analysis* 55.2, pp. 621–651.

Da Rin, Marco, Thomas Hellmann, and Manju Puri (2013). "A Survey of Venture Capital Research". *Handbook of the Economics of Finance*. Ed. by George M. Constantinides, Milton Harris, and Rene M. Stulz. Vol. 2. Elsevier, pp. 573–648.

Das, Sanjiv R., Hoje Jo, and Yongtae Kim (2011). "Polishing diamonds in the rough: The sources of syndicated venture performance". *Journal of Financial Intermediation* 20.2, pp. 199–230.

Du, Qianqian (2016). "Birds of a feather or celebrating differences? The formation and impacts of venture capital syndication". *Journal of Empirical Finance* 39, pp. 1–14.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications". *Biometrika* 57.1, pp. 97–109.

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu (2007). "Whom You Know Matters: Venture Capital Networks and Investment Performance". *The Journal of Finance* 62.1, pp. 251–301.

Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu (2010). "Networking as a Barrier to Entry and the Competitive Supply of Venture Capital". *The Journal of Finance* 65.3, pp. 829–859.

Katz, Leo (1953). "A new status index derived from sociometric analysis". *Psychometrika* 18.1, pp. 39–43.

Lerner, Joshua (1994). "The Syndication of Venture Capital Investments". *Financial Management* 23.3, pp. 16–27.

Lindsey, Laura (2008). "Blurring Firm Boundaries: The Role of Venture Capital in Strategic Alliances". *The Journal of Finance* 63.3, pp. 1137–1168.

Marjoram, Paul, John Molitor, Vincent Plagnol, and Simon Tavaré (2003). "Markov chain Monte Carlo without likelihoods". *Proceedings of the National Academy of Sciences* 100.26, pp. 15324–15328.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller (1953). "Equation of State Calculations by Fast Computing Machines". *The Journal of Chemical Physics* 21.6, pp. 1087–1092.

Nahata, Rajarishi (2008). "Venture capital reputation and investment performance". *Journal of Financial Economics* 90.2, pp. 127–151.

Sørensen, Morten (2007). "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital". *The Journal of Finance* 62.6, pp. 2725–2762.

Sorenson, Olav and Toby E. Stuart (2001). "Syndication Networks and the Spatial Distribution of Venture Capital Investments". *American Journal of Sociology* 106.6, pp. 1546–1588.

Sorenson, Olav and Toby E. Stuart (2008). "Bringing the Context Back In: Settings and the Search for Syndicate Partners in Venture Capital Investment Networks". *Administrative Science Quarterly* 53.2, pp. 266–294.

Tian, Xuan (2011). "The causes and consequences of venture capital stage financing". *Journal of Financial Economics* 101.1, pp. 132–159.

Appendix A Details of the Structural Model

A.1 Baseline network model

See Battaglini, Leone Sciabolazza, and Patacchini (2020). Given the production function

$$E_i = \rho s_i^{\alpha} l_i^{1-\alpha} + \epsilon_i \tag{A.1}$$

where $s_i = \sum_{j \in \mathcal{N}} g_{ij} E_j$ measures the social connectedness of i. The following conditions are imposed on the parameter space. The effort $l \in [0, \bar{l}]$ with $\bar{l} > 0$. The cost of effort is simply normalized as l_i. The social link between i and j intensity $g_{ij} \in [0, \bar{g}]$ with $\bar{g} > 0$. Following convention in the literature, self-connection is assumed to be zero $g_{ii} = 0$. Individual heterogeneity $\epsilon_i \in [\underline{\epsilon}, \bar{\epsilon}]$ with $\underline{\epsilon} > 0$ and $\bar{\epsilon} \in (0, 1)$. Now assume that $\rho \bar{g}^{\alpha} \bar{l}^{1-\alpha} + \bar{\epsilon} < 1$, a sufficient condition that guarantees $E_i \in (0, 1)$.

The agent chooses l to maximize his effectiveness net of effort cost $\rho s_i^{\alpha} l_i^{1-\alpha} + \epsilon_i - l_i$. The first-order condition yields the optimal effort level $l_i^* = (\rho(1-\alpha))^{\frac{1}{\alpha}} s_i$. Plugging this back into the production function gives

$$E_i^* = \delta \sum_{j \in \mathcal{N}} g_{ij} E_j^* + \epsilon_i \tag{A.2}$$

where $\delta = \rho^{\frac{1}{\alpha}} (1-\alpha)^{\frac{1-\alpha}{\alpha}}$. Further, since the system is linear, it can be inverted and solved by a unique equilibrium.

$$\boldsymbol{E}(\mathbf{G}, \boldsymbol{\epsilon}; \delta) = [\mathbf{I} - \delta \mathbf{G}]^{-1} \boldsymbol{\epsilon} \tag{A.3}$$

A.2 Heckman-Corrected Network Model

The first step in the Heckman-corrected network model is a link formation model.

$$g_{ij} = \gamma_0 + \gamma_1 h_{ij} + \sum_{l} \delta_{l+1} d(X_i^l, X_j^l) + \eta_{ij} \tag{A.4}$$

where h denotes alumni or professional connection, and $X^1, \cdots, X^L$ is a vector of VC characteristics.

Further assume the covariance matrix of (ϵ, η) has the following properties. $\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$ and $\eta_i = (\eta_{i1}, \cdots, \eta_{in})'$ are jointly normal with $E(\epsilon_i^2) = \sigma_\epsilon^2$, $E(\epsilon_i \eta_{ij}) = \sigma_{\epsilon\eta}$ for all $i \neq j$, $E(\eta_{ij}^2) = \sigma_\eta^2$, and $E(\eta_{ij}\eta_{ik}) = 0$ for all $j \neq k$. Then it can be shown that the expected value of the error term conditional on the link formation is $E(\epsilon_i|\eta_{i1}, \cdots, \eta_{in}) = \psi \sum_{j \neq i} \eta_{ij}$, where $\psi = \sigma_{\epsilon\eta}/\sigma_\eta^2$. The model after correction is given by

$$\boldsymbol{E} = \delta \mathbf{G} \boldsymbol{E} + \mathbf{X} \boldsymbol{\beta} + \psi \boldsymbol{\xi} + \boldsymbol{\epsilon} \tag{A.5}$$

where $\xi_i = \sum_{j \neq i} \eta_{ij}$.

A.3 Exogenous Network Formation Model

See Battaglini, Patacchini, and Rainone (2021). The cost of establishing social link is given by the following. In this model, the cost C_{ij} is assumed to be born by i only. (Conversely, C_{ji} is born by j only.) This assumption is not important and can be generalized easily.

$$C(g_{ij}, \theta_{ij}) = \frac{\lambda}{1+\lambda} \left(\frac{g_{ij}}{\theta_{ij}} \right)^{1+\frac{1}{\lambda}} \tag{A.6}$$

Another important assumption is that $\bar{l} > \left(\frac{(1-\alpha)\rho}{c} \right)^{\frac{1}{\alpha}}$. This guarantees interior solutions of $l_i < \bar{l}$.

Consider a two-period game. In period 1, agents choose connections. In period 2, agents choose effort levels. Type $\omega_i = (\epsilon_i, (\theta_{ik})_k, M_i)$. Denote Ω the space of types. We solve for a pure strategy (g, l) where $g : \Omega \to [0, \bar{g}]^{n-1}$ maps the VC type to a vector of connection intensities and $l : \Omega \times G \to [0, \bar{l}]$ maps type and networks to the effort level.

We solve the game by backward induction. At $t = 2$, the agents choose l to maximize effectiveness net of effort cost $\rho s_i^\alpha l_i^{1-\alpha} + \epsilon_i - l_i$. This is the same as the baseline network model above. At $t = 1$, agents form networks. First, observe that the continuation value is (ignoring discounting)

$$E_i(G, \epsilon) - l_i(G, \epsilon) = \alpha \delta \sum g_{ij} E_j(G, \epsilon) + \epsilon_i \tag{A.7}$$

The agent maximizes the continuation value net of the costs of connection formation

$$\sum_{j=1}^{n} \left(\alpha \delta g_{ij} E_j(G, \epsilon) - \frac{\lambda}{1+\lambda} \left(\frac{g_{ij}}{\theta_{ij}} \right)^{1+\frac{1}{\lambda}} \right) \tag{A.8}$$

Equations (A.2) and (A.8) characterize the network competitive equilibrium. The equilibrium exists and is characterized by E^* and G^* where

$$E_i^* = \delta \sum_j g_{ij}^* E_j^* + \epsilon_i \tag{A.9}$$

$$g_i^* \leq \theta_{ij}^{1+\lambda} (\alpha \delta E_j^*)^\lambda \tag{A.10}$$

If an interior solution exist for all i, j, then the second equation, which is based on the first-order condition of equation (A.8), is binding and enables further simplification.

$$E_i^* = \varphi \sum_j (\theta_{ij} E_j^*)^{1+\lambda} + \epsilon_i \tag{A.11}$$

where $\varphi = \alpha^\lambda \delta^{1+\lambda}$. That is, the equilibrium effectiveness are characterized by a system of nonlinear equations.

Note further that the elasticity of a link g_{ij} with respect to the effectiveness of j, E_j is precisely λ

$$\epsilon_{g_{ij}, E_j} = \frac{\partial g_{ij}}{\partial E_j} \frac{E_j}{g_{ij}} = \theta_{ij}^{1+\lambda} (\alpha \delta E_j)^{\lambda - 1} \alpha \delta \frac{E_j}{g_{ij}} = \lambda \tag{A.12}$$

Lastly, assume that $\bar{g} > (\alpha\delta)^\lambda \bar{\theta}^{1+\lambda}$, where $\bar{\theta} = \max \theta_{ij}$. A sufficient condition for unique equilibrium is that δ is sufficiently small. If $\delta \leq \frac{1}{\bar{\theta}} \left(\frac{1}{(1+\lambda)\alpha^\lambda \bar{m}} \right)^{\frac{1}{1+\lambda}}$, then the equilibrium is unique.