Knowledge Spillovers and Strategic Network Formation

Kazuma Takakura*

April, 2024

Abstract

This study examines an efficient policy for accelerating research and development (R&D) in technology developing firms. To find an efficient R&D subsidy allocation, I estimate the degree of knowledge spillovers through collaborative research networks among firms. I address the interdependence of the R&D efforts and the choice of collaborators by constructing a two-stage model in which firms first decide their R&D investments and then choose collaborators in the second stage. Structural estimation provides the estimates of the magnitudes of spillover effects and determinants of its collaboration partners. I also find that when the network structure is treated exogenously, the spillover effects are estimated to be smaller between 8% and 18% than when endogenous network changes are considered. Counterfactual analysis reveals that subsidy allocation targeting firms currently involved in a lot of collaborations is more efficient, both in terms of promoting R&D investment and intense collaboration.

1 Introduction

Innovation is often the result of combining various types of knowledge. When inventors interact, knowledge spills over and plays a crucial role in developing new technologies (Kerr, 2008; Ductor et al., 2014; Akcigit et al., 2018; Zacchia, 2019; Anderson and Richards-Shubik, 2022). These spillovers are sources of externality, which implies that each firm's research and development (R&D) investment should be smaller than the social optimum. To promote innovation and achieve social optima, policies that provide additional incentives for R&D investment and encourage inventors to interact should be considered. Although both policies are certain to be effective, it is challenging to find a quantitatively optimal policy because of the interdependence

*Graduate School of Economics, University of Tokyo. kazuma.arukakat@gmail.com

of R&D investment decisions and interactions between inventors. The current literature usually assumes that R&D spending is exogenous and examines the mechanism of collaboration, or vice versa, but does not endogenize both. However, when R&D collaboration relationships affect firms' R&D investments, they also determine with whom they should collaborate. Without a model that addresses this interdependence, policy effects cannot be evaluated correctly.

This study constructs a model that considers endogenous collaboration network formation when firms determine their R&D investment in a competitive market and estimate it structurally. The model involves a two-stage game in which firms first choose their R&D effort based on their expected collaboration network and then propose partners with which to collaborate. Using data from U.S. manufacturing firms and their collaborative patent submission history, I estimate the degree of spillover, rivalry effects of product market competition, and the determinants of collaboration partners. These structural parameters enable counterfactual experiments to illustrate the impact of R&D acceleration policies on firms' investment and collaboration networks, showing that subsidies targeted at firms with high R&D expenditure are more effective than those that are not.

A crucial feature of this model is indirect spillover from a firm that does not collaborate directly. For example, assume that there are three firms: A, B, and C. There are two collaborative relationships between A and B and A and C, but not between B and C. Through research collaboration, firm A benefits from firm B's knowledge and vice versa. Simultaneously, firm C can also indirectly benefit from firm B's knowledge through collaboration with A. Although firms B and C do not directly collaborate, knowledge can still be transferred through the intermediation of A, which is connected to both firms. This indirect spillover is the reason for the joint endogeneity between R&D investment and collaboration. Firms' decisions on R&D investment should be based on the global structure of the collaborative network, considering the existence of indirect spillovers. If firms maintain a fixed R&D investment, the selection of their collaboration partners will be strategic because collaborators' partners can indirectly learn from their knowledge. Therefore, global collaboration networks and R&D investments depend on each other.

I use data from Bloom et al. (2013) and Zacchia (2019), which include the account information of manufacturing firms in the U.S., the list of scientists who were employed by those firms, and co-authored patents that scientists applied for. This dataset allowed me to connect firm-level R&D expenditures with the firm affiliations of the scientists involved in each research project. By analyzing the applicant information recorded in patents, patents filed by multiple companies or institutions can be identified and the collaboration network among these applicants can be observed. The panel data indicate a strong positive correlation between collaboration intensity and R&D expenditure.

I construct the model of a two-stage game in which firms first choose their R&D effort based on their expected collaboration network, and then choose partners to collaborate with. The first stage is based on a Cournot competition game, in which a firm determines its output and R&D investment. In this model, R&D investment lowers the marginal cost of its own production. Additionally, I incorporate a spillover structure in which collaborators' R&D investments contribute to a reduction in marginal cost. To account for the interaction of incentives to change network structure and R&D investment, I introduce the second stage in which firms can choose their collaboration partners. This model can be solved using backward induction. In the second stage, firms select their collaborators based on the decisions made in the first stage. In the first stage, firms make decisions regarding their R&D investments and output considering the expected spillovers and the anticipated network structure that will be realized in the second stage.

I first estimate the second stage and then use the estimated parameters to estimate the first stage. In the second stage, during which firms choose their collaboration partners, I employ the method developed by Leung (2015) to estimate the parameters that determine the network structure in the presence of externalities. Although the model involves externalities, I use the theoretical results of Comola and Dekel (2022) to restrict the characteristics of the equilibrium. Once the parameters of the second stage network formation game were estimated, the parameters that determine output and R&D investment in the first stage were estimated. I use a spatial autoregression model to address the reflection problem resulting from spillovers in the profit function.

The estimation results quantify the relative contributions of their own R&D investments and those of direct and indirect collaborators to the reduction in marginal cost. The degree of cost reduction from the collaborators' R&D is 13% relative to the reduction from a firm's R&D. Moreover, the effect of a collaborator's R&D is 4% of that of R&D. These direct and indirect spillover effects are larger than the estimates in existing literature. I examine whether these differences arise from an estimation bias resulting from ignoring the endogeneity of network formation.

I examine whether these differences come from the estimation bias that results from ignoring the endogeneity of network formation. As a result, the estimates that capture direct and indirect spillovers are 8% and 18% larger, respectively, than when I do not consider the second stage and the network structure is not endogenized. These magnitudes of bias are consistent with the differences in cost reduction estimates between our model and those in the literature.

Using the estimated model, I conduct a counterfactual analysis of policies that allocate R&D subsidies. I compare the efficiency of three allocations; (i) uniform allocation, (ii) allocation proportional to the extent of existing collaborations a firm originally had, and (iii) allocation inversely proportional to the number of a firm's original cooperative relationships. An examination of how R&D spending can increase depending on how subsidies of the same total amount are allocated reveals significant differences among the three allocation policies. The most cost-effective policy is the allocation method (ii). The subsidy policy has also been shown to significantly increase the number of collaborative research projects, which is a major factor in promoting R&D. The results suggest that targeting firms that have already a lot of collaborators makes subsidy allocation efficient.

This article stands on the literature on knowledge spillovers resulting from individual R&D efforts. Bloom et al. (2013) have separately estimated R&D spillovers and the negative competition effect that is due to the R&D of product market rivals. While this study is groundbreaking in that it identifies two effects, it does not analyze the cooperation between individual firms. Zacchia (2019) studied the interactions between individual inventors from different companies that drive knowledge spillover among firms. In his study, in which he estimated the spillover effects given an exogenous network structure using the micro-data of inventors' collaboration, the endogeneity of network formation was not involved in the model. Arora et al. (2021) examined the relationship between spillover and R&D incentives. They found that the more sensitive a company is to the use of its research results by other companies, the more it reduces its share of R&D. My study adds to this literature, which had estimated spillover effects with these networks as exogenous, by showing that the results are significantly different when the networks are treated as endogenous. In addition, this study relates to a developing literature on collaborations and coauthorship among economics researchers(Kerr, 2008; Akcigit et al., 2018; Anderson and Richards-Shubik, 2022). However, no studies have simultaneously analyzed the effectiveness of collaborative research and the determinants of research partners. A two-step estimation shows how to deal with this problem.

I also apply findings from the literature on empirical network formation. The fundamental difficulty in identifying strategic network formation is the existence of multiple equilibria owing to externalities. Recent studies have developed structural models to address this problem (Miyauchi, 2016; Mele, 2017; Paula et al., 2018; Sheng, 2020). However, in these studies, the conditions for identification were often restrictive, or only partial identification was possible. For example, while Mele (2017) illustrates the network formation game as a potential game and identifies the parameters' stationary distribution, his procedure requires at least one of the externalities to be negative and sufficiently large. This method is not appropriate in this case because the presence of negative spillover effects cannot be confirmed a priori. Partial identification is also difficult because of the limited number of networks in my dataset. However, Leung (2015) shows that, when data are rationalized by a symmetric equilibrium, the two-step estimation provides point-identified estimators in a directed network formation game. Comola and Dekel (2022) extend the methods of Leung (2015) to allow undirected link formation. I incorporated their method into my two-stage game.

Recent studies have encompassed two strands of the literature: R&D spillovers and strategic network formation. Dasaratha (2022) describes a theoretical model of innovation in a strategic network formation game with many firms. His model captures the positive and negative competitive effects of network formation. They found an equilibrium and showed that several interventions failed to improve welfare. My empirical analysis is consistent with the theoretical predictions and quantitatively reveals effective interventions. König et al. (2019) and Hsieh et al. (2022) developed a structural model for the coevolution of networks and behavior of agents. König et al. (2019) estimate both R&D investment decisions and collaboration network formation. However, their estimation of network structure depends on the history of collaboration, and the model is independent of R&D investment decisions. As counterfactual simulations, they rank firms according to the size of spillovers when subsidizing them. Hsieh et al. (2022) applied the Bayesian double Metropolis–Hastings algorithm to estimate a structural model for the coevolution of networks and behavior; they estimated the spillover effects of R&D investment and collaboration decisions in the chemical and pharmaceutical industries. Their counterfactual analysis revealed key players in the network. In this study, I construct a two-stage model in which both R&D investment and collaboration decisions are consistent with profit maximization. The model was estimated using the method described by Leung (2015). My counterfactual experiment examined the efficiency of subsidy allocation and changes in the network structure under several allocation scenarios.

The remainder of this paper is organized as follows: Section 2 describes the proposed two-stage game model. Section 3 describes the data used and summary statistics. Section 4 outlines the identification and estimation procedures. Section 5 presents the estimation results and discusses economic interpretations. Finally, in Section 6, I use the estimated model for a counterfactual experiment to analyze the effectiveness of R&D subsidy allocations. Section 7 concludes the paper.

2 Theoretical Framework

This section provides a comprehensive description of the theoretical model that combines the Cournot competition and R&D collaboration considerations. In Section2.1, the R&D collaboration relationship is assumed exogenous. Subsequently, this relationship is endogenized in Section2.2 to accommodate the strategic interactions arising from network formation. The entire game unfolds into two stages: firms first determine their R&D investments and outputs based on anticipated network structures, and then form the R&D collaboration network.

2.1 Cournot competition

Consider a market comprising a set of firms $N = 1, ..., n$. Adopting the approach of König et al. (2019), a Cournot oligopoly framework is employed. Denote the price and quantity of good i are p_i and q_i, respectively. Given the potential for the goods to be imperfect substitutes, the inverse demand function is expressed as:

$$p_i = \bar{p}_i - q_i - \rho \sum_{j=1}^{n} b_{ij} q_j \tag{1}$$

Here, $\bar{p}_i$ embodies market size variations, while ρ signifies the substitutability extent between products. The coefficient b_{ij} describes the product market closeness of goods i and j.

Firms can reduce their marginal production costs by investing in research and development (R&D), representing the R&D effort of firm i as e_i. The corresponding marginal cost c_i can be expressed as:

$$c_i = \bar{c}_i - \underbrace{\varphi_0 e_i}_{\text{own effort}} - \underbrace{\varphi_1 \sum_{j \neq i} a_{ij} t_{ij} e_j}_{\text{direct spillover}} - \underbrace{\varphi_2 \sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{jk} t_{ik} e_k}_{\text{indirect spillover}} \tag{2}$$

Ensure that $\bar{c}_i$ is sufficiently large such that c_i remains positive for all entities $i \in N$. The R&D collaboration network is depicted by a symmetric adjacency matrix $\mathbf{A}$, with elements a_{ij} assigned a value of 1 if firms i and j are collaborating, and 0 otherwise. t_{ij} is the technological closeness between firm i and j. The spillover effect from the R&D collaboration network on cost reduction is captured by the parameters φ_1 and φ_2.

In this model, we explicitly assume that R&D efforts are used for process innovation. However, we can also capture product innovation with some modifications. In general, product innovation is characterized the differentiation, which is reflected in $\bar{p}_i$. If we define $\bar{p}_i = \tilde{p}_i - \varphi_0 e_i - \varphi_1 \sum_{j \neq i} a_{ij} t_{ij} e_j - \varphi_2 \sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{jk} t_{ik} e_k$, we can capture the effect of product innovation. On the other hand, since my data does not let us observe whether each R&D effort aims at process or product innovation, we cannot identify the degree of each effect. Then, my model only accounts for the impact of process innovation, but due to the linear representation of price and cost, we also can interpret it as a product innovation effect.

It's posited that the R&D effort's associated cost (or the R&D investment) increases with effort, displaying diminishing returns, specifically $\frac{1}{2}e_i^2$.[1] A network formation cost, ω_{ij}, is also introduced, which detracts from a firm's profit when $a_{ij} = 1$. Consequently, the profit for firm i is

$$\pi_i = (p_i - c_i)q_i - \frac{1}{2}e_i^2 - \sum_j a_{ij}\omega_{ij} \tag{3}$$

Optimizing with respect to e_i and q_i yields:

$$\mathbf{e} = \varphi_0 \mathbf{q} \tag{4}$$

$$2\mathbf{q} = \lambda - \rho \mathbf{B}\mathbf{q} + \varphi_0 \mathbf{e} + \varphi_1 \mathbf{A}\mathbf{e} + \varphi_2 \mathbf{A}_2 \mathbf{e} \tag{5}$$

Here, λ represents the vector $\bar{p}_i - \bar{c}_i$, and matrices $\mathbf{A}$, $\mathbf{A}_2$, and $\mathbf{B}$ symbolize $a_{ij} \times t_{ij}$, $a_{ik} \times a_{kj} \times t_{ij}$,

[1] Since I define the efficiency of R&D effort using parameters φ_0, φ_1 and φ_2, the relative size of R&D effort and R&D expenditure does not need specific restriction. Then, R&D expenditure can be standardized in any way. For simplicity, I specify the level of R&D expenditure by $\frac{1}{2}$.

and b_{ij}, respectively.

There exists a unique Nash equilibrium with the equilibrium R&D effort vector $\mathbf{e}^*$ given by

$$\mathbf{e}^* = \mu - \gamma \mathbf{B}\mathbf{e}^* + \phi_1 \mathbf{A}\mathbf{e}^* + \phi_2 \mathbf{A}_2 \mathbf{e}^* \tag{6}$$

This equilibrium is viable if the matrix $\mathbf{I} - \phi_1 \mathbf{A} - \phi_2 \mathbf{A}_2 + \gamma \mathbf{B}$ exhibits positive definiteness.[2] Since $\mathbf{e}$ and $\mathbf{q}$ corresponds uniquely in equation (4), I focus on $\mathbf{e}$ later.

2.2 Network formation

In this section, I examine the endogeneity of the matrices $\mathbf{A}$ and $\mathbf{A}_2$, which are treated as exogenous in Section 2.1. Given the structure of our game, the network matrices $\mathbf{A}$ and $\mathbf{A}_2$ are influenced by $\mathbf{e}$ and are hence denoted as $\hat{\mathbf{A}}_{(\mathbf{e})}$ and $\hat{\mathbf{A}_{\mathbf{2}}}_{(\mathbf{e})}$. The subsequent chapters provide a detailed exploration of the network formation game, emphasizing the characterization of its equilibrium.

2.2.1 Setting

Considering that the network formation game occurs after firms have made decisions regarding their output and R&D investment, these firms will choose collaboration partners based on others' R&D investments of others. While firms can observe $\mathbf{e}$ and $\mathbf{t}$, I introduce a layer of imperfect information: firm i remains unaware of the network costs associated with other firms, symbolized as $(\omega_{jk})_{j,k \neq i}$. The existence of an unobservable component implies that I deal with a game with imperfect information.

The decisions of firm i can be represented by a vector of binary variables, $s_{ij} \in 0, 1$. A value of $s_{ij} = 1$ indicates that firm i is proposing collaboration with firm j, and $s_{ij} = 0$ indicates the absence of such a proposal. The total number of proposals, represented by $\sum_j s_{ij}$, is not constrained.

The benefits of such collaborations for the firms were established in the initial stage. Collaboration occurs when the expected benefits surpass the collaboration costs, as dictated by the

[2] $\mu = \frac{\varphi_0 \lambda}{2-\varphi_0^2} \gamma = \frac{\rho}{2-\varphi_0^2}, \phi_1 = \frac{\varphi_0 \varphi_1}{2-\varphi_0^2}, \phi_2 = \frac{\varphi_0 \varphi_2}{2-\varphi_0^2}$. The existence proof for this unique Nash equilibrium mirrors the methodology employed in König et al. (2019).

following condition[3]:

$$s_{ij} = \mathbb{1}\left\{\mathbb{E}\left(\psi_1 t_{ij} e_i e_j + \psi_2 \sum_{k \neq i,j} a_{jk} t_{ik} e_i e_k - \omega_{ij} \mid \boldsymbol{e}, \boldsymbol{t}\right) \geq 0\right\} \tag{7}$$

Lastly, the network formation cost function is specified as:

$$\omega_{ij} = \beta_0 + \beta_1 \text{PreLink}_{ij} + \beta_2 \text{Geo}_{ij} - \epsilon_{ij} \tag{8}$$

Here, PreLink_{ij} denotes a prior collaboration, where β_1 captures the costs associated with initiating a new contract. Similarly, Geo_{ij} signals geographical proximity, and β_2 quantifies communication costs. The unobservable factor, ϵ_{ij}, represents private information and is derived from a Type-I EV distribution. Although idiosyncratic shocks exist and are universally distributed, their precise values remain concealed from other firms.

2.2.2 Equilibrium

I examine the stable matching where a_{ij} takes the value of 1 if and only if both s_{ij} and s_{ji} are 1; otherwise, a_{ij} equals 0. This relationship can be expressed as $a_{ij} = s_{ij} \times s_{ji}$. I introduce X as a vector capturing the observable characteristics that influence a firm's actions. This includes elements such as e_i, t_i, PreLink_i, and Geo_i.

Consider $\sigma^{A_{-i}}$ as the matrix representing firm i's beliefs about other firms' collaboration. Here, A_{-i} represents the matrix A excluding the ith row and column. Consequently, i's expected marginal benefit from collaboration with j is contingent on both X and $\sigma^{A_{-i}}$. Given that the marginal payoff function satisfies linearity, separability, and anonymity, firm i's action with respect to firm j becomes independent. This can be represented as:

$$s_{ij} = \mathbb{1}\{\mathbb{E}[v_{ij}(X, A_{-i}; \theta_0)|X, \sigma^{A_{-i}}] + \epsilon_{ij} \geq 0\} \tag{9}$$

Here, θ_0 is the vector of true parameters and includes $\psi_1, \psi_2, \beta_0, \beta_1$, and β_2. Additionally, $v_{ij}(X, A_{-i}; \theta_0)$ can be derived from (7) and (8).

A belief matrix, σ^A, can be a Bayesian Nash equilibrium if, for all $i, j \in N$, it satisfies the

[3] $\psi_1 = \frac{\varphi_1}{\varphi_0}$, $\psi_2 = \frac{\varphi_2}{\varphi_0}$

subsequent condition :

$$\sigma^A_{ij} = \Pr(a_{ij} = 1 | X, \sigma^A) \tag{10}$$

Furthermore, where σ^A is an equilibrium belief, the equilibrium network A conforms to:

$$a_{ij} = \underbrace{\mathbb{1}\{\mathbb{E}[v_{ij}(X, A_{-i}; \theta_0)|X, \sigma^{A_{-i}}] + \epsilon_{ij} \geq 0\}}_{s_{ij}} \underbrace{\mathbb{1}\{\mathbb{E}[v_{ji}(X, A_{-j}; \theta_0)|X, \sigma^{A_{-j}}] + \epsilon_{ji} \geq 0\}}_{s_{ji}} \tag{11}$$

Comola and Dekel (2022) incorporate the symmetric equilibrium condition, which requires that pairs of agents with matching characteristics exhibit identical ex-ante linking probabilities. This condition is expressed as follows:

$$(X_i = X_k \text{ and } X_j = X_l) \text{ or } (X_i = X_l \text{ and } X_j = X_k) \Rightarrow \sigma^A_{ij} = \sigma^A_{kl} \tag{12}$$

The presence of a symmetric equilibrium is verified in Comola and Dekel (2022). While this condition may appear constraining when all attributes are discrete, it becomes less stringent when continuous attributes are incorporated. I thus move to an enhanced condition: similar pairs of agents should have analogous ex-ante linking probabilities. Formally, an equilibrium σ^A should satisfy that for all $\varepsilon > 0$, there exists $\delta > 0$ such that for all pairs $ij \neq kl \in N$

$$||X_{ij} - X_{kl}|| < \delta \Rightarrow |\sigma^A_{ij} - \sigma^A_{kl}| < \varepsilon \tag{13}$$

I then assume that condition (13) is satisfied within the equilibrium. By positing that only symmetric equilibria are chosen, I can identify within a single network as suggested by (Leung, 2015).

3 Data

In the empirical assessment, I employ data drawn from Zacchia (2019), originally sourced from Bloom et al. (2013) and Li et al. (2014). This dataset encompasses 736 R&D-intensive firms from the U.S. stock market, curated from COMPSTAT and categorized under the SIC four-digit sector. These firms have corresponding entries in the NBER patent dataset, as described in Hall et al. (2001). The patents are segregated into 426 distinct classes as per the USPTO's categorization.

Bloom et al. (2013) computed the technological proximity of two firms in terms of their patent allocation using a measure adapted from Jaffe (1986):

$$t_{ij} = \frac{(T_i T_j')}{(T_i T_i')^{\frac{1}{2}}(T_j T_j')^{\frac{1}{2}}} \tag{14}$$

Here, $T_i = (T_{i1}, T_{i2}, ..., T_{i426})$ is the vector that gathers $T_{i\tau}$ where $T_{i\tau}$ signifies the patent proportion of firm i in the technology class τ amidst the 426 classes.

Similarly, product market closeness was measured using SIC and four-digit industry codes. This measure was calculated as follows:

$$b_{ij} = \frac{(S_i S_j')}{(S_i S_i')^{\frac{1}{2}}(S_j S_j')^{\frac{1}{2}}} \tag{15}$$

where $S_i = (S_{i1}, S_{i2}, ..., S_{i597})$ denotes the vector of S_{ik}. S_{ik} is the share of sales of firm i in four digit industry k across the 597 classes. In their dataset, t_{ij} and b_{ij} are discrete variables that take integer values between zero and 100. In my estimation, I standardize these variables to the values between zero and one.

In addition, the patent dataset of Li et al. (2014) was combined. This dataset consists of 1315060 patents granted to 565019 scientists with information on which firms they belong to. I use the co-authorship of jointly filed patents as the connection between scientists and the firms with which they are associated. Specifically, for two inventors m and n, the link element a_{mnt} takes the value 1 if at time $t+1$ the USPTO has received at least one patent application filed at any time in the past by both m and n. I associate inventors m and n with firms i and j that belong to and construct a_{ijt}, the link information between firms. [4]

Because our model is not dynamic and cannot use panel data, I employed a sample from 1996, when collaboration between firms was most observed in the panel. This year, there were 1146 links between the 461 firms that have positive R&D expenditures.

Table 1 reports the firm-level summary statistics for 1996. On average, one firm had collaborative research relationships with 3.2 firms. The measures of technological and product market closeness are very small, on average, at 0.049 and 0.025, respectively. If I limit the sample to those who are involved in a collaboration, the overage values become 0.304 and 0.166, respectively.

Figure 1 illustrates the collaborative networks of the 175 firms in 1996. This figure only

[4]The details of this association are provided by Zacchia (2019).

Table 1. Summary Statistics (1996)

Variable	Unit	Observations	Mean	SD	Minimum	Maximum
Number of Collaborations	firm	461	3.8	7.774	0.00	62.00
R&D Expenditure	firm	461	0.172	0.638	0.001	8.900
t_{ij} (Tech Closeness)	pair	106030	0.049	0.036	0.000	0.990
b_{ij} (Market Closeness)	pair	106030	0.024	0.091	0.000	1.000
$t_{ij} \mid a_{ij} = 1$	pair	876	0.304	0.255	0.000	0.990
$b_{ij} \mid a_{ij} = 1$	pair	876	0.166	0.249	0.000	1.000

The values are measured in 1996 prices in $billion. Standardized between 0 and 1 and discretized in 0.01 increments. The sample is limited to firms whose R&D expenditure is positive.

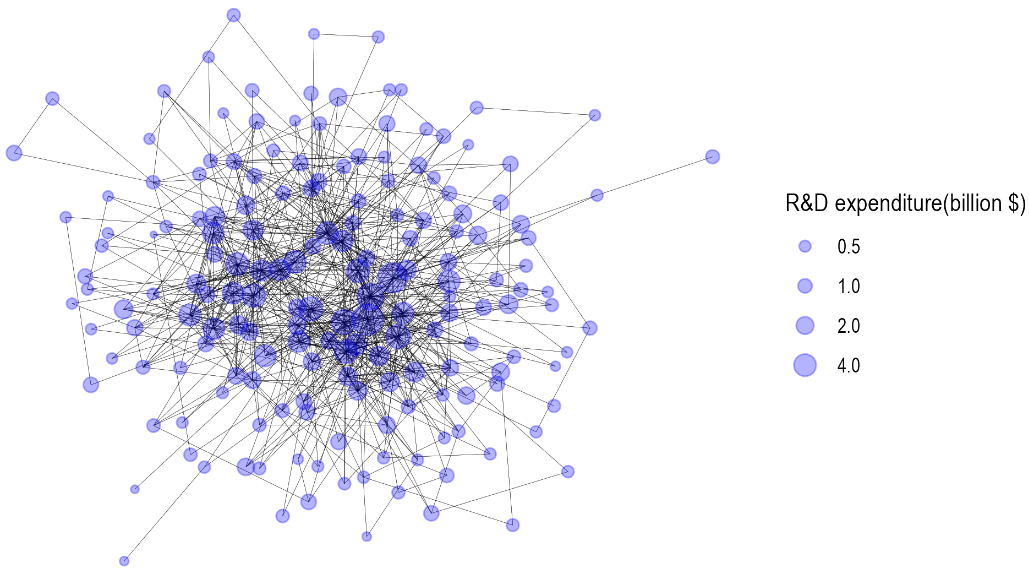

Figure 1. Collaborations between firms (1996)

includes firms with more than two collaborations. Companies with small-scale R&D tend to be located on the periphery, whereas those with large-scale R&D tend to be located in the center of the network. In addition, companies located at the center of the network are involved in a number of joint research projects, three or more.

In Table 2, I checked the correlation between collaboration probability and pairwise characteristics which are indicated in the second stage of my model. I regressed the collaboration dummy on $t_{ij}e_ie_j$, $\sum_{k\neq i,j} a_{jk}t_{ik}e_ie_k$, dummy of collaboration in 1995, and the same state dummy. All four variables show a positive significant correlation with collaboration probability. In specification (2), I controlled b_{ij}. The result does almost not change, and the coefficient on b_{ij} also shows a positive and significant estimate. This positive correlation between market

Table 2. Correlation between Collaboration and Firm Characteristics

Variable	(1)	(2)	(3)
Intercept	-0.0002	-0.0007	-0.0001
	(0.0002)	(0.0002)	(0.0002)
R&D of partner (t weighted)	0.0122	0.0113	0.0118
	(0.0070)	(0.0068)	(0.0069)
R&D of partner's partners (t weighted)	0.0015	0.0015	0.0015
	(0.0003)	(0.0003)	(0.0003)
Collaboration in last year	0.7390	0.7374	0.7311
	(0.0167)	(0.0167)	(0.0171)
Same state	0.0033	0.0029	0.0026
	(0.0009)	(0.0009)	(0.0009)
b_{ij}		0.0222	0.0206
		(0.0054)	(0.0054)
Firm characteristics			×
Adjusted R^2	0.5877	0.5882	0.5882

The sample is 461 firms that hold positive R&D expenditure in the 1996 dataset. In specification (3), I controlled each firm's stock of R&D, total assets, total sales, employment, and market value. The robust standard errors are in parentheses.

competitiveness and collaboration probability is intuitive because firms producing more differentiated goods have less incentive to collaborate. I also included the other firm characteristics, such as stock of R&D, total assets, total sales, employment, and market value. The result is almost the same as (1) and (2). In addition, adjusted R^2 in the baseline model is 0.5877, which can partially support my assumption of symmetric equilibrium.

4 Estimation

The estimation process initiates with the network formation stage, followed by the estimation of parameters governing $\hat{\mathbf{A}}$ and $\hat{\mathbf{A}}_2$. Subsequently, the parameters affecting the first stage were estimated.

4.1 Beliefs

Given a symmetric equilibrium, the belief estimation is expressed as

$$\hat{\sigma}^A_{ij} = \frac{\sum_{l,k>l} a_{kl} \cdot \mathbb{1}\{X^d_{ij} = X^d_{kl}\} K\left(\frac{d(X^c_{ij}, X^c_{kl})}{h}\right)}{\sum_{l,k>l} \mathbb{1}\{X^d_{ij} = X^d_{kl}\} K\left(\frac{d(X^c_{ij}, X^c_{kl})}{h}\right)} \tag{16}$$

where X^d_i and X^c_i represent the discrete and continuous attributes of i, respectively. The function $d(X^c_{ij}, X^c_{kl})$ signifies the vector of attribute distances between two pairs, with h as the bandwidth and $K(\cdot)$ as a kernel function. A proof showcasing that $\hat{\sigma}^G_{ij}$ is consistent with σ^G_{ij}, a belief in symmetric equilibrium, for all $i, j \in N$ can be found in Comola and Dekel (2022).

4.2 Network Formation

From the equilibrium condition (11), the likelihood of observing network A is:

$$L(\theta, \sigma^A) = \prod_{i,j>i}^{N} \left(\frac{1}{1+\exp(-\mathbb{E}[v_{ij} \mid \theta, \sigma^A])} \frac{1}{1+\exp(-\mathbb{E}[v_{ji} \mid \theta, \sigma^A])} \right)^{a_{ij}} \\ \times \left(1 - \frac{1}{1+\exp(-\mathbb{E}[v_{ij} \mid \theta, \sigma^A])} \frac{1}{1+\exp(-\mathbb{E}[v_{ji} \mid \theta, \sigma^A])} \right)^{(1-a_{ij})} \tag{17}$$

Assuming that unobserved utility shocks, ϵ_{ij}, are independently sourced from the Gumbel distribution, parameters are estimated by maximizing the log-likelihood function. The parameter θ is then estimated using the deduced symmetric equilibrium beliefs $\hat{\sigma}^A$. Notably, Comola and Dekel (2022) verifies that the estimator complies with consistency and asymptotic normality.

4.3 First stage

Given the derivations of $\hat{\mathbf{A}}_{(\mathbf{e})}$ and $\hat{\mathbf{A}}_{2(\mathbf{e})}$, the equation for $\mathbf{e}$ is:

$$\mathbf{e} = \mu - \gamma \mathbf{B}\mathbf{e} + \phi_1 \hat{\mathbf{A}}_{(\mathbf{e})}\mathbf{e} + \phi_2 \hat{\mathbf{A}}_{2(\mathbf{e})}\mathbf{e} + \varepsilon \tag{18}$$

Here, ε represents the idiosyncratic shock influencing R&D investments, and is i.i.d., following the normal distribution $N(0, \sigma^2)$. When the matrix $\mathbf{\Omega} = \mathbf{I} - \phi_1 \hat{\mathbf{A}}_{(\mathbf{e})} - \phi_2 \hat{\mathbf{A}}_{2(\mathbf{e})} + \gamma \mathbf{B}$ is positive definite, it can be restructured as:

$$\mathbf{e} = \mathbf{\Omega}^{-1}(\mu + \varepsilon) \sim N(\mathbf{\Omega}^{-1}\mu, \sigma^2 \mathbf{\Omega}'\mathbf{\Omega}) \tag{19}$$

Then, employing the spatial lag regression model as per (Arbia, 2014), I derive the log-likelihood function:

$$\ell(\mu, \varphi, \rho, \sigma^2) = C - \frac{n}{2}\log\sigma^2 + \log|\mathbf{\Omega}| - \frac{1}{2\sigma^2}\{\mathbf{\Omega}e - \mu\}^T\{\mathbf{\Omega}e - \mu\} \tag{20}$$

To conclude, by maximizing this likelihood, I estimate three parameters in the model of the first stage (μ, φ, ρ) and the variance of the distribution of the error term (σ^2).

5 Results

Table 3. Parameter Estimates: Network Formation

Parameter	Symbol	Estimates
Direct Synergy	ψ_1	0.2776
		(0.0458)
Indirect Synergy	ψ_2	0.0514
		(0.0036)
Cost: Constant	β_0	3.2451
		(0.0420)
Cost: Previous Collaboration	β_1	-4.5001
		(0.1051)
Cost: Same State	β_2	-0.4623
		(0.0969)

Standard errors are in parentheses.

The estimated parameters ψ_1 and ψ_2, as delineated in Table 3, suggest a propensity for firms to engage in collaborative research with others. This likelihood increases under three primary conditions: (i) large R&D expenditure by the firm, (ii) significant R&D expenditure by a prospective collaborating firm, and (iii) close technological field alignment between the firm and its potential collaborator. Notably, the parameter ψ_2 is 80% smaller than ψ_1, indicating that while firms do consider the existing collaborators of potential partners, this consideration is less pronounced, accounting for approximately 20%. The positive estimates of ψ_2 signify that the benefits derived from indirect learning through a collaborator's network generally outweigh the risks associated with knowledge leakage and potential business stealing. Additionally, the statistically significant positive coefficients of β_1 and β_2 imply that a history of prior collaboration and a shorter physical distance between firms contribute to diminishing the costs associated with collaboration.

Table 4. Parameter Estimates: First Stage

Parameter	Symbol	Endogeneous Network	Fixed Network
Direct Reflection	ϕ_1	0.1093	0.1008
		(0.0079)	(0.0076)
Indirect Reflection	ϕ_2	0.0159	0.0131
		(0.0041)	(0.0042)
Competition	γ	0.0059	0.0070
		(0.0035)	(0.0037)
Variance of ε	σ^2	0.3045	0.4575
		(0.0200)	(0.0310)

Standard errors are in parentheses.

In the estimation of the first stage, as presented in the first column of Table 3, the coefficients ϕ_1 and $\phi_2$are estimated to ascertain the extent of reflexivity in the autoregression model, specified in equation (18). These coefficients indicate the impact of direct and indirect collaborations, respectively, on a firm's R&D expenditure decisions. The results revealed that a firm's R&D spending decisions are significantly influenced by the R&D expenditures of its direct collaborators. Echoing findings from the network formation stage, the influence of indirect collaborative connections is quantitatively less than that of direct collaborations, constituting 15% of the effect size. Furthermore, the coefficient γ is estimated to be positive, suggesting that R&D expenditures function as strategic substitutes among firms that compete within closely related product markets. This positive estimation of γ indicates a tendency for firms to adjust their R&D spending in response to the R&D activities of their competitors.

In addition, I examine the estimation bias of utilizing the observable collaboration network directly instead of relying on the collaboration structure formulated during the network formation stage. This methodology evaluates the extent of estimation the bias incurred when the network structure is assumed to be exogenous, thus overlooking the dynamic nature of firms' incentives to modify their collaborative networks. As illustrated in the second column of Table 4, the findings reveal an underestimation of the reflective impacts of both direct and indirect collaborations when the endogenous nature of network formation is not accounted for. Coefficients ϕ_1 and ϕ_2 are found to be lower by 8% and 18%, respectively, under this exogenous assumption. The influence emanating from competitive interactions is found to be overestimated by 19%. These comparative analyses underscore the importance of recognizing the endogenous formation of collaborative networks. Neglecting this aspect leads to a disproportionate emphasis on the effects of competition, thereby failing to fully capture the underlying incentives driving collaborative partnerships.

Table 5. Parameter Estimates: Cournot Competition

Parameter	Symbol	Estimates	Calculation
R&D Own Effort	φ_0	0.7424 (0.1159)	$\sqrt{2\frac{\phi_1+\phi_2}{\psi_1+\psi_2}/1+\frac{\phi_1+\phi_2}{\psi_1+\psi_2}}$
R&D Direct Spillover	φ_1	0.2061 (0.0380)	$\varphi_0\psi_1$
R&D Indirect Spillover	φ_2	0.0382 (0.0048)	$\varphi_0\psi_2$
Market Rivalry	ρ	0.0085 (0.0000)	$(2-\varphi_0^2)\gamma$

Standard errors are in parentheses. Standard errors are calculated by 100 times bootstraps.

Finally, the parameters estimated in both the first and second stages enable the reconstruction of the original parameters defining Cournot competition. These estimated parameters are detailed in Table 5. A critical observation from this analysis is the relative magnitudes of φ_0 and φ_1, which suggest that R&D expenditure by collaborators has a 27% efficient impact in reducing marginal costs compared to a firm's own R&D expenditure. This value does not consider the technology difference because it is not weighted by t_{ij}. If we weight this effect by the average value of t_{ij} among collaborations and standardize the own effort efficiency to one, the average spillover effect is $\bar{t}_{ij} \times \varphi_1/\varphi_0 = 0.084$. This degree is larger than existing literature. For example, IV estimation in König et al. (2019) showed that the degree of direct spillover effect is 0.058. This difference may depend on whether or not the structural model is used to remove bias when dealing with the endogenous nature of network structure. Additionally, the coefficient φ_2 indicates that indirect spillovers from a firm's collaborators' network can also lead to reductions in marginal costs, albeit at a lower efficiency rate of approximately 18% relative to direct spillovers. The positive coefficient ρ quantifies the degree of competitive intensity within the Cournot game framework. This positive value of ρ highlights the importance of strategic interactions among firms in shaping their R&D investment decisions and their subsequent impact on market structure and firm behavior.

6 Counterfactual Simulations

Using the estimated parameters, counterfactual simulations can be executed to examine the variations in R&D investment patterns that may occur under several subsidy regimes. This subsidy, designed to incentivize R&D activities, is proposed as a per-unit grant for R&D expenditure as follows:

$$\pi_i = (p_i - c_i)q_i - \frac{1}{2}e_i^2 + \underbrace{s_i\frac{1}{2}e_i^2}_{\text{subsidy}} - \sum_j a_{ij}\omega_{ij} \tag{21}$$

where s_i is the proportion of the subsidies.

Given the estimated parameters and artificially determined s_i, both the R&D expenditures and the network structure that will be realized can be estimated. This estimation result can be obtained as a fixed point by alternatively updating R&D investment and network structure. This updating process was confirmed to be contraction mapping, at least under the size of the subsidy set.

In this analysis, I evaluate three distinct subsidy allocation policies; (i) uniform allocation, (ii) allocation proportional to the extent of existing collaborations a firm originally had, and (iii) allocation inversely proportional to the number of a firm's original cooperative relationships. The first approach is a uniform allocation, where the subsidy s_i is consistent across all firms $i \in N$. Second, I explore a policy framework wherein the subsidy percentage increases in direct proportion to the extent of existing collaborations a firm has in the absence of subsidies. This approach is based on the hypothesis that subsidizing firms with extensive collaborative networks can effectively catalyze R&D activities in other firms through large spillover effects. On the other hand, the third policy proposes an inverse relationship between subsidy allocation and the number of original cooperative relationships. Under this scheme, firms with fewer preexisting cooperative relationships receive a higher percentage of subsidies. This approach aims to stimulate R&D by encouraging new collaborations, especially among firms with limited partnerships. The rationale behind this is the potential for significant R&D promotion through the spillover effects arising from new collaborations.

Table 6. Counterfactual: Subsidy for R&D Expenditure

Subsidy Expenditure		(i)	(ii)	(iii)
10	R&D increase	17.30	25.48	12.26
	New collaborations	10	32	2
50	R&D increase	98.49	152.82	59.14
	New collaborations	135	359	4
100	R&D increase	216.48	333.42	117.40
	New collaborations	415	916	11

Expenditure is measured in 1996 prices in billion dollars. I evaluate three distinct subsidy allocation policies; (i) uniform allocation, (ii) proportional to the extent of existing collaborations a firm originally has, and (iii) inversely proportional to the number of a firm's original cooperative relationships. There are three patterns of total subsidy expenditure: 10 billion, 50 billion, and 100 billion dollars. R&D increase is the difference in the total amount of R&D expenditure between original data and counterfactual. New collaborations counts the number of collaborations which exist in counterfactual but not in the original data.

Table 6 presents the outcomes of various counterfactual experiments wherein hypothetical subsidy policies are implemented with total allocations of \$10 billion, \$50 billion, and \$100 billion, respectively. This experiment assesses the impact of these subsidy policies on both the resultant R&D expenditure and the formation of collaborative research networks. Each policy is evaluated using two metrics, the increase in total R&D expenditures relative to a no-subsidy scenario, and the increase in the number of collaborations compared to a no-subsidy situation.

Initially, a uniform subsidy allocation is used as the benchmark. Under this approach, a subsidy of $10 billion is found to enhance overall corporate R&D spending by $17.3 billion, with the efficiency of the subsidy escalating with its size; a $50 billion subsidy boosts R&D by $98.5 billion, and a $100 billion subsidy results in $216.5 billion increase. Additionally, the likelihood of initiating new collaborations increases nonlinearly with subsidy size, indicating that subsidies not only directly increase R&D expenditure but also indirectly foster innovation on a larger scale through enhanced collaboration.

Among the tested policies, the policy focusing on subsidies to firms with a pre-existing extensive network of collaborations was the most effective in both of two metrics. In particular, the number of joint research projects has increased non-linearly with the increase in subsidies. This result suggests that the number of networks of firms doing large-scale R&D may not have been efficient in the baseline situation without subsidies.

Conversely, policies aimed at supporting firms without existing collaboration were found to be relatively ineffective. Although these policies lead to an increase in total R&D expenditures, they are less successful in fostering new collaborative projects. The number of new joint research initiatives remained significantly lower, between 1% and 6%, compared with the second policy, uncovering the critical role of grant distribution methods in the efficacy of promoting collaborative research.

Figures 2, 3, and 4 graphically describe the results when the total amount of subsidy is 100 billion dollars. Figure 2 shows that allocation (iii) does not indicate a dense network or high R&D expenditure. Figure 3, which shows the results for subsidy scenario (i), describes a more dense network structure. Figure 4 illustrates the results of the most efficient allocation, (ii). Both the network density and the size of R&D investment are the largest in the three scenarios in the analysis. Especially, some of the companies in the center of the network have substantial R&D investments and are collaborating with a great number of companies.

7 Conclusion

This study elucidates the factors influencing R&D investment in competitive markets by focusing on the endogenous decision-making processes governing interfirm collaboration. The findings

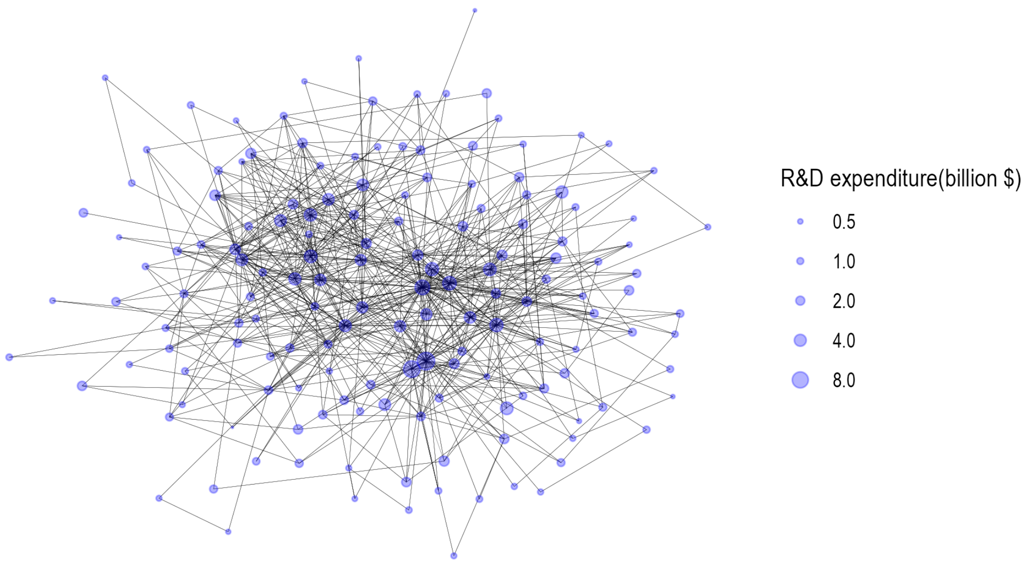

Figure 2. Subsidy Allocation (iii): More subsidy for firms with less collaborations

from counterfactual analyses suggest that an allocation of subsidies with targeting, primarily towards firms already engaged in extensive collaborative networks, can effectively stimulate additional R&D investment and foster further collaboration. This insight emerges from an analysis that incorporates endogenous alterations within a collaborative research network, a dimension previously unexplored in the extant literature. The methodologies employed in this study offer a framework for examining the decision-making processes in the presence of network spillover effects. This approach enables the assessment of various hypothetical policy scenarios, considering the interdependent nature of behavioral and network changes among agents.

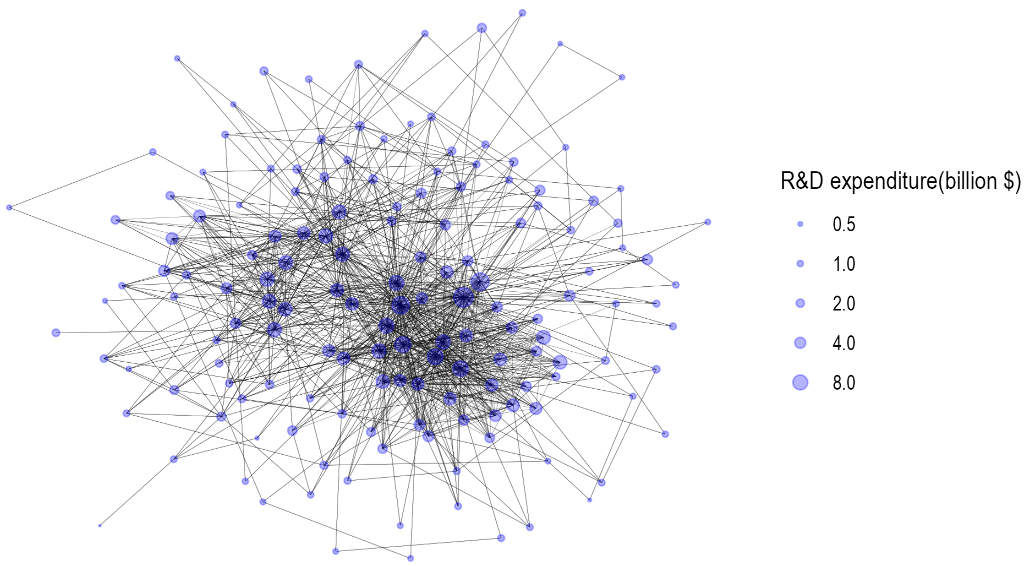

Figure 3. Subsidy Allocation (i): Unifrom subsidy

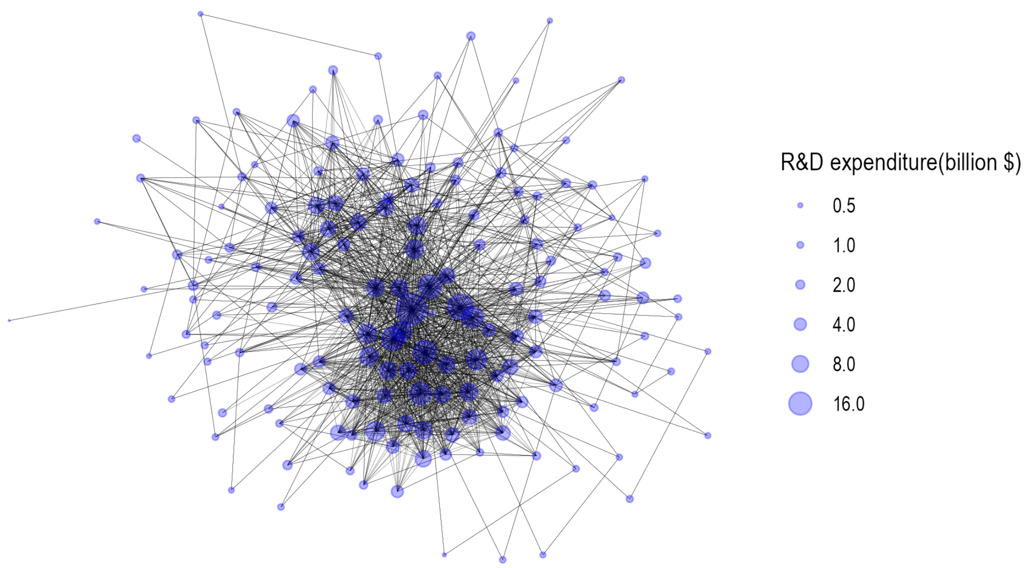

Figure 4. Subsidy Allocation (ii): More subsidy for firms with more collaborations

References

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018) "Dancing with the Stars: Innovation Through Interactions," March.

Anderson, Katharine A and Seth Richards-Shubik (2022) "Collaborative production in science: An empirical analysis of coauthorships in economics," *Rev. Econ. Stat.*, 104 (6), 1241–1255.

Arbia, Giuseppe (2014) *A Primer for Spatial Econometrics*: Palgrave Macmillan UK.

Arora, Ashish, Sharon Belenzon, and Lia Sheer (2021) "Knowledge Spillovers and Corporate Investment in Scientific Research," *Am. Econ. Rev.*, 111 (3), 871–898.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013) "Identifying technology spillovers and product market rivalry," *Econometrica*, 81 (4), 1347–1393.

Comola, Margherita and Amit Dekel (2022) "Estimating Network Externalities in Undirected Link Formation Games."

Dasaratha, Krishna (2022) "Innovation and Strategic Network Formation," *Rev. Econ. Stud.*, 90 (1), 229–260.

Ductor, Lorenzo, Marcel Fafchamps, Sanjeev Goyal, and Marco J van der Leij (2014) "Social networks and research output," *Rev. Econ. Stat.*, 96 (5), 936–948.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg (2001) "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," October.

Hsieh, Chih-Sheng, Michael D König, and Xiaodong Liu (2022) "A structural model for the coevolution of networks and behavior," *Rev. Econ. Stat.*, 1–13.

Jaffe, Adam B (1986) "Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value," *Am. Econ. Rev.*, 76 (5), 984–1001.

Kerr, William R (2008) "Ethnic Scientific Communities and International Technology Diffusion," *Rev. Econ. Stat.*, 90 (3), 518–537.

König, Michael D, Xiaodong Liu, and Yves Zenou (2019) "R&D networks: Theory, empirics, and policy implications," *Rev. Econ. Stat.*, 101 (3), 476–491.

Leung, Michael P (2015) "Two-step estimation of network-formation models with incomplete information," *J. Econom.*, 188 (1), 182–195.

Li, Guan-Cheng, Ronald Lai, Alexander D'Amour, David M Doolin, Ye Sun, Vetle I Torvik, Amy Z Yu, and Lee Fleming (2014) "Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010)," *Res. Policy*, 43 (6), 941–955.

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

Miyauchi, Yuhei (2016) "Structural estimation of pairwise stable networks with nonnegative externality," *J. Econom.*, 195 (2), 224–235.

Paula, Áureo, Seth Richards-Shubik, and Elie Tamer (2018) "Identifying preferences in networks with bounded degree," *Econometrica*, 86 (1), 263–288.

Sheng, Shuyang (2020) "A structural econometric analysis of network formation games through subnetworks," *Econometrica*, 88 (5), 1829–1858.

Zacchia, Paolo (2019) "Knowledge Spillovers through Networks of Scientists," *Rev. Econ. Stud.*, 87 (4), 1989–2018.

Appendix

A Details on Data

Table A1. Top 10 SIC Codes in Sample

SIC	Industry	Observations
367	Electronic components and accessories	34
382	Measuring and controlling devices	31
357	Computer and office equipment	29
366	Communications equipment	25
384	Medical instruments and supplies	24
283	Drugs	23
371	Motor vehicles and equipment	23
356	General industrial machinery and equipment	14
737	Computer and data processing services	13
353	Construction, mining, and materials handling	10