Knowledge Spillovers through Networks of Scientists

Paolo Zacchia*

August 2017

Abstract

In this paper I directly test the hypothesis that interactions between inventors of different firms drive knowledge spillovers. I construct a network of publicly traded companies in which each link is a function of the relative proportion of two firms' inventors who have former patent collaborators in both organizations. I use this measure to weigh the impact of R&D performed by each firm on the productivity and innovation outcomes of its network linkages. An empirical concern is that the resulting estimates may reflect unobserved, simultaneous determinants of firm performance, network connections and external R&D. I address this problem with an innovative IV strategy, motivated by a game-theoretic model of firm interaction. I instrument the R&D of one firm's connections with that of other firms that are sufficiently distant in network space. With the resulting spillover estimates, I calculate that among firms connected to the network the marginal social return of R&D amounts to approximately 112% of the marginal private return.

JEL Classification Codes: C31, D85, O31, O33

Keywords: R&D Spillovers, Co-authorship Networks, Indirect Friends

*IMT School for Advanced Studies, Lucca Italy

Theories of *knowledge spillovers* have occupied a central position in economic analysis since at least Marshall (1890) posited their role to explain the apparent productivity advantages for firms to cluster near one another in manufacturing districts of 19th century England. Since then, knowledge spillovers have entered economic theories of industrial innovation, geographic agglomeration, economic growth, international trade and more. However, the exact mechanisms through which knowledge spills from one agent or organization to another are still unclear. Conjectures about human interaction and spatial proximity as drivers of information exchange are typically associated with methods of measuring spillovers that are unable to test their hypotheses directly, as they are typically based on aggregate R&D metrics.

This paper contributes to the empirical literature on the quantitative assessment of R&D spillovers by directly measuring the role of individual relationships in the diffusion of industrially valuable knowledge. I estimate the effect of R&D performed by different firms, that are linked through their scientists, on their reciprocal performance and innovation rates. In particular, I exploit collaborations on past patents in order to identify inventors who are likely to maintain personal linkages across different organizations over time. For each pair of firms, I measure the degree of interaction between the two R&D teams and the potential for information exchange by the relative proportion of cross-connected inventors. This metric changes over time, as scientists move across firms or acquire new collaborators.

By combining firm-level data with patent data that identify individual inventors over the course of their patenting history, I am able to construct a dynamic network of knowledge exchange. This network includes the largest, most innovative and R&D intensive U.S. firms, and it becomes tighter over time thanks to the increase in the total number of connections. The R&D of connected firms, weighted by the intensity of the links, is significantly and positively correlated with firm performance and innovation rates as measured by patent counts. This contrasts with well-established measures of spillovers that are based, for instance, on the technological similarities between firms (Jaffe, 1986, 1989). Among companies that comprise the network, these measures are not significantly and robustly correlated with relevant firm-level outcomes.

It is arduous, however, to attribute a causal interpretation to these findings. As in other studies about spillovers and externalities of different kinds, these correlations may simply reflect the existence of common unobserved confounders simultaneously driving R&D, innovation, as well as firm performance (Griliches, 1998). For example, a sudden technological breakthrough in a technological niche where few connected firms operate could stimulate additional R&D efforts while, at the same time, facilitating productivity-enhancing follow-up discoveries. This corresponds to the problem of *correlated effects* in the classification by Manski (1993) of identification issues in the estimation of spillover effects. In addition, it is possible that network linkages are themselves endogenous. For instance, more reputed and better-connected scientists might be attracted to more productive and better paying firms. Under these circumstances, standard estimates of R&D spillovers may be biased in either directions.

Thanks to the characteristics of the network that I observe, I can formulate a novel empirical strategy that addresses these problems. The basic intuition is straightforward. Unobserved factors shared by a pair of connected firms – call them i and j – may bias standard estimates of spillovers as long as they are reflected in R&D expenditures. Suppose that a third firm k, which is not connected to i, shares some of these unobservables with j but not with i. It is not necessary, which is a crucial point, that j and k are themselves directly connected, but only that k is "closer" to j than it is to i in the network space; this may be the result, say, of a process of network formation in which firms are more likely to be connected if they are similar. If shared external circumstances affect R&D investment as hypothesized, *and* firms have private information on their own unobservables (so that these are not endogenously transmitted), R&D should be correlated within firm pairs (i, j) and (j, k), but not within the pair (i, k). Hence the R&D of firm k, while correlating with that of firm j, is orthogonal with respect to the unobservables of firm i. I argue that relationships of this kind are commonplace in networks, as evidenced by specific statistical regularities.

To formalize this idea, I describe a game of R&D investment played in a network of firms. R&D exerts reciprocal spillovers across linkages; in addition, firms are hit by shocks that are correlated with the characteristics of the network. Consequently, equilibrium R&D also co-varies across neighboring nodes, and the resulting correlation is endogenously amplified by the strategic anticipation of investment choices made by other firms. However, under reasonable assumptions that allow for both flexible patterns of dependence between the shocks and the network, and varying information structures of the game, the model predicts the existence of a degree of separation at which the R&D of different firms is independent. As the R&D choices of firms that lie at the bound also correlate with those of direct links, the former can serve as valid instruments. Empirically, I find no significant cross-correlation of firms' R&D at three degrees of separation. This evidence motivates the use of instrumental variables based upon the R&D of "indirect connections" located at distance three as my best choice.

Without applying this strategy, I find substantial effects of connected firms' R&D on firm performance – expressed in terms of productivity and market value – as well as on patent production. However, when instrumenting peers' knowledge with the R&D choices of indirect links located at distance two or three, I obtain larger point estimates of spillover effects, especially for the productivity and patent production outcomes. That such difference is only apparent when employing the distance three instrument in isolation is remarkably consistent with the proposed model. I interpret these findings as evidence that R&D is in fact driven by common unobservables across connected firms, and I advance several hypotheses to explain why this translates into a negative bias of OLS estimates. In light of the results, I estimate the social returns of R&D to be about 112% of the private returns among connected firms.

This paper builds on the traditional literature of industrial and innovation economics about the determinants of productivity at the firm level, especially the private and social returns of R&D.[1] The quest for R&D spillovers in particular, initiated with the original intuitions by Griliches (1964, 1979, 1992), has developed into its current empirical framework with the cited contributions by Jaffe. Successive research has experimented with metrics of spillovers, based for example on cross-industry transactions or flows of patent citations, that are alternative to Jaffe's concept of *technological proximity*; for a review of these studies see e.g. Hall, Mairesse, and Mohnen (2010). Other authors have assessed more specific mechanisms of knowledge diffusion. For example, Branstetter and Sakakibara (2002), as well as König, Liu, and Zenou (2014)[2] study the effect of R&D joint ventures. Griffith, Harrison, and Van Reenen (2006) instead examine the consequences of UK firms' technological outsourcing in the US.

In recent work Bloom, Schankerman, and Van Reenen (2013) have addressed a longstanding issue in the literature, disentangling R&D spillovers from the negative competition effect that is due to the R&D of product market rivals. In their article they also postulate a microfoundation of knowledge spillovers based on the frequency of personal or professional interactions between inventors, but they do not explicitly test this mechanism in their empirical analysis. In this contribution, I provide for the

[1] For a general survey see Syverson (2011).

[2] König et al. (2014) also take a network-based approach to their analysis of joint ventures, which bears similarities with the one in this paper. Their identification strategy is, however, different, and it does not address potential problems of correlated confounders.

first time a measure of cross-firms spillovers grounded on the observation of an actual social network of inventors: the patent collaboration network. This measure aims at capturing all kinds of individual interactions between inventors that eventually result in joint projects. While spillovers certainly also occur through less solid, harder to observe types of interactions, the proposed measure has the advantage of generality. In fact, collaborative cross-firm projects are common to other mechanisms examined in the literature, such as R&D joint ventures or technological outsourcing.

This work provides empirical evidence to support the hypothesis that spillovers are caused by the exchange of ideas between individuals. Therefore, it is related to the research about the micro-level determinants of performance in the workplace. Moretti (2004) argues that productivity is related to how well-educated the workforce is in the environment in which a plant is located, suggesting that knowledge spillovers have a local scope.[3] Mas and Moretti (2009) demonstrate how "peer effects" apply at work, as coworkers intensify their efforts when they observe others doing increasingly so. Serafinelli (2016) shows that firm productivity is related to positive flows of workers with experience from companies at the top of the productivity distribution. In the context of scientific production, which is especially relevant for this work, Azoulay, Graff Zivin, and Wang (2010) evidence the negative impact of superstars' deaths on the publication rate of scientific collaborators.[4]

The empirical strategy that I propose, centered on the idea of using the R&D of "sufficiently distant" firms to predict the R&D of direct neighbors, is itself a contribution to the literature of spatial and network econometrics. While instrumental variables of this kind are not novel as a concept (Bramoullé et al., 2009; De Giorgi et al., 2010), both my objective and conceptual framework[5] are different. In the cited

[3]Knowledge spillovers are theorized to be one of the determinants of agglomeration economies (Moretti, 2011). There are, in fact, complementarities between the empirical literature about spatial agglomeration effects, and studies on R&D spillovers. Jaffe et al. (1993, 2000) and Thompson and Fox-Kean (2005) discuss whether the spatial concentration of patent citations can be considered as evidence of localized knowledge spillovers. Bloom, Schankerman, and Van Reenen (2013) as well as Lychagin et al. (2016) attempt to identify a geographic scope of R&D spillovers, by attributing a differential effect to the R&D of "spilling" firms that are located closer in space.

[4]However, in a related study Waldinger (2011) does not find similarly convincing evidence following the expulsion of scientists from Nazi Germany.

[5]The model developed in this paper is inspired by those in Calvó-Armengol et al. (2009), Conley and Udry (2010), Kranton et al. (2014), Blume et al. (2015), while differing from all of them. The empirical reduced form of the model that I estimate is a production function with R&D spillovers; it corresponds to a Spatial Durbin Error Model (SDEM) from the classification of spatial econometric models by Elhorst (2014), because it includes both an analogue of Manski's "exogenous effect" and

articles, in fact, IVs are meant to solve Manski's "reflection" problem, by extending methods originally devised for spatially autoregressive models (Kelejian and Prucha, 1998) to the case of networks. Here instead, the aim of the proposed methodology is to disentangle spillover effects from spatially distributed unobservables or "correlated effects."[6] The latter are, according to some authors (Angrist, 2014), the main cause for concern regarding studies on peer effects. Given that, as I illustrate, correlated effects and network formation are closely related, under the maintained hypotheses the proposed strategy is also robust to endogeneity of firms' connections. To the best of my knowledge this approach is new in the spatial econometrics literature, and it can be viewed as a spatial extension of familiar GMM methods for dynamic panels.

This paper is organized as follows. Section 1 illustrates the game-theoretic framework that models R&D investment in a network in the presence of spillovers. Section 2 describes the collaboration-based measures of connections, and provides a description of the resulting dynamic network. Section 3 outlines the econometric framework and discusses the empirical strategy of the paper. Section 4 presents the empirical results of the analysis and their economic implications. Finally, Section 5 provides some concluding remarks. A set of appendices accompanies this article, complementing both the theoretical and the empirical analyses.

1 Analytical Framework

In this section I outline the theoretical framework of this paper. The model I describe explores the equilibrium relationship of firms' choice of R&D investment when they exert network externalities on each other and are also subject to simultaneous correlated shocks. The objective of the model is to formalize the intuition motivating the identification strategy of the paper. In the appendices I provide formal proofs of the results, as well as additional comments on the model and possible extensions of it.

spatially correlated errors. The main feature of my model is that both are interdependent, inducing endogeneity. Note that a complete information version of this model can be manipulated so to return a spatially autoregressive (SAR) model of either R&D or final firm outcomes. In these cases, Manski's "endogenous effect" would be a function of the spillover parameter of the production function.

[6]The procedure described in Bramoullé et al. (2009) and implemented in De Giorgi et al. (2010) has been conceived for cross-sectional data made of multiple, separate networks. If correlated effects are identical within networks, taking fixed effects at the network level is sufficient to partial them out – as it is claimed in some of the empirical applications based on their framework. If common shocks are, instead, characterized by a more complex spatial dependence as a function of network topology, their approach would result in inconsistent estimates even with a sample of networks.

1.1 Model Setup

An economy consists of a set $\mathcal{I}$ of N firms, whose output depends on conventional inputs (e.g. capital, labor) as well as on *knowledge capital* (Griliches, 1979). Knowledge is the result of R&D activity performed by teams of researchers – be they professional scientists, occasional inventors, academic collaborators of firms or other individuals – who are linked together in a network of professional relationships. These networks transcend the borders of the individual firms: I represent the intensity of connections between any two firms as the N^2-dimensional set $\mathcal{G} = \{g_{ij} : i, j = 1, \dots, N\}$, where $(i, j) \in \mathcal{I} \times \mathcal{I}$ denotes any pair of firms. By adopting standard normalizations, I set $g_{ij} \in [0, 1]$ for all pairs such that $i \neq j$, as well as $g_{ii} = 0$ for every firm i in $\mathcal{I}$. The *firm-level network* of knowledge flows is thus given by the pair of sets $(\mathcal{I}, \mathcal{G})$. In this paper I assume that the network is undirected, that is $g_{ij} = g_{ji}$ for each pair of firms (i, j); the results of the model are however easily extended to directed networks.

Thanks to the formal and informal exchange of information that happens through the firm-level network, one firm's knowledge depends not only on R&D performed in-house, but also on the R&D of other, connected firms. Specifically, I assume that the knowledge capital $\tilde{S}_i$ of firm i is a Cobb-Douglas function of its own R&D investment, denoted as S_i, and the R&D investment S_j of any other j-th firm, as follows.

$$\tilde{S}_i = S_i^{\gamma} \left(\prod_{j=1}^{N} S_j^{g_{ij}} \right)^{\delta} \tag{1}$$

In the expression above, parameters $\gamma \in (0, 1)$ and $\delta \in (0, 1)$ represent, respectively, the relative contribution of in-house R&D and knowledge spillovers to the knowledge capital of some firm i. The actual intensity of knowledge flows directed from firm j to firm i, however, depends on the strength of their link in the network, expressed by the spillover weight g_{ij}. Note that this functional form implies that R&D is a strategic complement. A model featuring R&D as a strategic substitute would yield different empirical predictions about the sign of R&D cross-correlation in the network, but would not invalidate the main results that support identification.[7]

[7]Whether R&D is truly more a strategic complement or a strategic substitute is a controversial matter: it is a notoriously hard dichotomy to test. R&D is usually thought to be both – a complement and a substitute – to some degree. As Jaffe (1986) put it, this is a ultimately a question about the assumed functional form, and standard econometric techniques are not the best means to assess curvature parameters beyond first derivatives.

Knowledge capital $\tilde{S}_i$ enters as an additional input into the general production function of firm i, which is also Cobb-Douglas:

$$Y_i\left(X_{i1},\ldots,X_{iQ};S_1,\ldots,S_N\right)=\left(\prod_{q=1}^{Q}X_{iq}^{\beta_q}\right)S_i^{\gamma}\left(\prod_{j=1}^{N}S_j^{g_{ij}}\right)^{\delta}e^{\omega_i} \tag{2}$$

where X_{iq} for $q=1,\ldots,Q$ is any conventional input (like capital or labor), $\beta_q\in(0,1)$ being its associated elasticity parameter. In addition, output depends on a stochastic shock $\omega_i\in\mathbb{R}$, representing other technological and environmental factors that affect firms' productivity or profitability in either direction. For example, ω_i may enclose technological knowledge in the common domain that is specific to the industry in which firm i operates, or efficiency-enhancing managerial and organizational practices whose acquisition is independent of firms' R&D effort.

It is likely that two firms i and j that are connected in the network of knowledge flows ($g_{ij}\neq 0$) share some related technological factors that affect their performance. A concrete, famous case is the semiconductor industry. For decades, firms operating in that sector have been enjoying parallel trends in the development of integrated circuits with an increasingly higher count of transistors: the so-called "Moore's Law." Another pertinent example is the pharmaceutical sector, where firms developing new drugs typically enjoy common advantages based on the results of basic research. This cross-correlation of firms' technologies in the network can be related to the process of network formation: firms happen to learn from some other firms and their inventors (and not others) because they operate in similar technological niches.

In addition, the R&D cost function $z\left(S_i,\varpi_i\right)$ of firms also depends on a random variable $\varpi_i\in\mathbb{R}$ which can be spatially correlated in the network. For every firm i:

$$z\left(S_i,\varpi_i\right)=e^{\varpi_i}S_i$$

that is, the cost borne for an additional unit of R&D S_i increases with larger values of ϖ_i. Cross-correlation of R&D costs in the network can be due to the fact that firms with similar technological characteristics ω_i, which as argued may relate to the process of network formation, also face similar costs ϖ_i. In practice, the spatial dependence in R&D costs may reflect, for example, common developments in the supply of labor endowed with specific technological skills, or in financing opportunities. Note that I make no restriction on the sign of the covariance $\mathbb{C}\text{ov}\left(\omega_i,\varpi_i\right)$ between the productivity and cost shocks of a firm, but I expect it to be positive in high-tech industries.

To allow for dependence between the network, the technological characteristics of different firms and their cost factors, I do not explicitly model network formation. Instead, I keep the model general by treating $\mathcal{G}$, the vector of technological shocks $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)$ and the vector of cost shocks $\boldsymbol{\varpi} = (\varpi_1, \ldots, \varpi_N)$ as random draws from some joint distribution $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{\varpi}, \mathcal{G})$. I impose no restriction on this distribution, except that the process of network formation or other determinants enclosed in $\mathcal{F}$ are unlikely to set firms with similar characteristics too far apart in network space. To formalize this idea, it is useful to introduce a notion of *distance* between any two firms in the network. Specifically, let $d_{ij} \in \mathbb{N}$ be the *minimum path length* between firms i and j: the lowest number of firms linked together as a sequence (*path*) indirectly connecting i to j.[8] Note that d_{ij} is a function of $\mathcal{G}$, and is itself a random variable. Armed with this concept, it is easier to express the following assumption.

Assumption 1. *Consider the set $\mathbb{G}$ of all realizations of $\mathcal{G}$ with positive probability in $\mathcal{F}(\cdot)$. There exists some positive integer C such that, for all $\mathcal{G} \in \mathbb{G}$, the conditional distribution $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{\varpi} | \mathcal{G})$ has the following property for all appropriate pairs (i, j):*

$$\mathbb{C}\text{ov}\left(\omega_i, \omega_j \middle| d_{ij} > C\right) = 0$$
$$\mathbb{C}\text{ov}\left(\varpi_i, \varpi_j \middle| d_{ij} > C\right) = 0$$

that is, if the minimum path length between i and j is higher than C, their productivity shocks ω_i and ω_j are mutually independent, and so are their cost shocks ϖ_i and ϖ_j.

In the context of this article Assumption 1 is interpreted as follows: if two firms have similar technologies, say one mainly operates in semiconductors and the other in ICT applications, they are not to be found very far away from one another in the network. This implies that if any two firms are sufficiently distant in a given observed network, their technological and cost characteristics are expected to be unrelated. In Appendix B I explore some models of network formation, inspired by those discussed by Graham (2015, 2017) and related ideas, that involve multi-technology firms. I illustrate how Assumption 1 is implied by these models either exactly or approximately, with low associated values of C.

[8]This count includes the end of the path: thus $d_{ij} = 1$ if $g_{ij} \neq 0$. For every node i, by definition $d_{ii} = 0$. In undirected networks, $d_{ij} = d_{ji}$ for every pair (i, j). Minimum path length is popularly referred to as the "degree of separation" between any two nodes in a network.

By specifying a vector of linear cost parameters $(\xi_1, \dots, \xi_Q) \in \mathbb{R}^Q_{++}$ for the Q conventional inputs, the firm profit function (revenues minus costs) can be written as

$$\pi_i \left(X_{i1}, \dots, X_{iQ}; S_1, \dots, S_N\right) = \left(\prod_{q=1}^{Q} X_{iq}^{\beta_q}\right) S_i^{\gamma} \left(\prod_{j=1}^{N} S_j^{g_{ij}}\right)^{\delta} e^{\omega_i} - \sum_{q=1}^{Q} \xi_q X_{iq} - e^{\varpi_i} S_i \quad (3)$$

for any firm $i = 1, \dots, N$. Note that individual profits depend both on firm-specific shocks ω_i and ϖ_i, as well as on the R&D choices of firms that are connected in the spillovers network. This in turn makes firm R&D, in equilibrium, dependent on the shocks of other firms. Thus, any notion of equilibrium should specify an information structure of the game. Denote as Ω_i the set of shocks ω and ϖ observed by firm i. I make a fairly general assumption about the structure of this set.

Assumption 2. *Every firm always observes its own individual shocks: $\omega_i, \varpi_i \in \Omega_i$. Moreover, there exists some integer L such that individual information sets do not include the shocks of firms located at distances higher than L: $(\omega_j, \varpi_j) \notin \Omega_i$ if $d_{ij} > L$.*

Assumption 2 states the obvious consideration that firms are aware of their own circumstances (shocks ω_i and ϖ_i). In addition, it specifies that for sufficiently high distances in network space, any two firms i and j that are that far away are ignorant of their respective shocks. In other words, this assumption rules out the case of complete information for networks of moderate diameter, which is arguably unrealistic. More concretely this means that the management of, say, a biotech firm is unlikely to know – or to take into account when making business decisions – the specific circumstances affecting a firm specialized in mechanical engineering, and vice versa.

I characterize the problem of firms' optimal input choice as a simultaneous game of incomplete information with the following timing.

1. Nature draws $(\boldsymbol{\omega}, \boldsymbol{\varpi}, \mathcal{G})$ from the common knowledge distribution $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{\varpi}, \mathcal{G})$. Every firm i observes the network $\mathcal{G}$ as well as its own information set Ω_i.
2. Firms simultaneously make their R&D and conventional input choices.
3. Payoffs (profits) are paid out.

For simplicity, network connections are not treated as strategic choices. In Appendix B I discuss a variation of this game in which, after nature has drawn $(\boldsymbol{\omega}, \boldsymbol{\varpi})$, each pair of firms cooperatively establishes connections g_{ij}, and then individual firms choose their inputs. As the input choice subgame remains unchanged, the model's empirical predictions still hold if the network formation stage conforms to Assumption 1. This is the case if, for dissimilar firms, the cost of establishing a link exceeds the benefit.

1.2 Equilibrium Predictions

The solution of the game is identified as a Bayes-Nash equilibrium. Define an individual strategy as a mapping from individual information sets onto valid choices of R&D investment and conventional inputs: $(\mathrm{S}_i, \mathrm{X}_i) : \Omega_i \mapsto (S_i; X_{i1}, \ldots, X_{iQ}) \in \mathbb{R}_{++}^{Q+1}$ for every firm $i = 1, \ldots, N$. Denote the vector of all other firms' R&D strategies as $\mathrm{S}_{-i} = \{(S_1, \ldots, S_N) \setminus S_i\} \in \mathbb{R}_{++}^{N-1}$. A Bayes-Nash equilibrium is a profile of strategies $(\mathrm{S}^*, \mathrm{X}^*) = [(\mathrm{S}_1^*, \mathrm{X}_1^*), \ldots, (\mathrm{S}_N^*, \mathrm{X}_N^*)]$ of all firms, such that

$$\mathbb{E}\left[\pi_i\left(\mathrm{S}_i^*, \mathrm{X}_i^*; \mathrm{S}_{-i}^*\right) \middle| \Omega_i\right] \geq \mathbb{E}\left[\pi_i\left(\mathrm{S}_i, \mathrm{X}_i; \mathrm{S}_{-i}^*\right) \middle| \Omega_i\right] \quad \forall\left(\mathrm{S}_i, \mathrm{X}_i\right) \neq \left(\mathrm{S}_i^*, \mathrm{X}_i^*\right)$$

for every firm $i = 1, \ldots, N$. The following result characterizes the equilibrium.

Proposition 1. *If*

$$\vartheta \equiv \frac{\delta}{1 - \gamma - \sum_{q=1}^{Q} \beta_q} < \min\left\{1, \left[\max_i \left(\sum_{j=1}^{N} g_{ij}\right)\right]^{-1}\right\}$$

there exists a unique Bayes-Nash equilibrium strategy profile which can be expressed, for $i = 1, \ldots, N$, as

$$\log S_i^* = \frac{\log \gamma + \sum_{q=1}^{Q} \beta_q \left(\log \beta_q - \log \xi_q - \log \gamma\right)}{1 - \gamma - \sum_{q=1}^{Q} \beta_q} b_i^*\left(\mathcal{G}; \vartheta\right) + s_i^*\left(\Omega_i, \mathcal{G}\right) \tag{4}$$

$$\log X_{iq}^* = \log S_i^* + \log \beta_q - \log \xi_q - \log \gamma + \varpi_i \quad \text{for } q = 1, \ldots, Q \tag{5}$$

where $b_i^*\left(\mathcal{G}; \vartheta\right)$ is the Katz-Bonacich network measure of centrality with attenuation factor ϑ for $i = 1, \ldots, N$, while $s_i^*\left(\Omega_i, \mathcal{G}\right)$ is a non-linear, spatially recursive function of firm i's information set Ω_i and network topology:

$$s_i^*\left(\Omega_i, \mathcal{G}\right) = \frac{1}{1 - \gamma - \sum_{q=1}^{Q} \beta_q} \left\{\tilde{\omega}_i + \log \mathbb{E}\left[\prod_{j=1}^{N} \exp\left(g_{ij} \delta \cdot s_j^*\left(\Omega_j, \mathcal{G}\right)\right) \middle| \Omega_i\right]\right\}$$

where $\tilde{\omega}_i \equiv \omega_i - \left(1 - \sum_{q=1}^{Q} \beta_q\right) \varpi_i$ for $i = 1, \ldots, N$.

This result is easily interpreted. First, consider that (5) is simply a set of constant relative input shares conditions, which is typical of the maximization of Cobb-Douglas functions. By contrast, equilibrium R&D given in (4) can be decomposed in two parts. The first one represents the deterministic component, for firm i, of the marginal return of R&D. It accounts for the complementarity of private R&D with both conventional inputs and the "certain" component of peers' R&D, itself a function $b_i^*(\mathcal{G};\vartheta)$ of firm i's position in the network. The second part represents the best equilibrium prediction that firm i can make, on the basis of private information, on how random shocks to both productivity and R&D costs of all firms in the network (including itself) would alter its own net marginal return of R&D. In equilibrium, in fact, all the shocks may affect the R&D investment of connections, which is complementary to private R&D.

The Bayes-Nash equilibrium expressed in Proposition 1 is unique for values of the spillover parameter δ that are sufficiently small relative to the overall spillover weights of all other firms. This condition is necessary to rule out the existence of "explosive" equilibria in which some firms invest infinite amounts in R&D. In theory, an explosive equilibrium could be "catalyzed" by the presence of a single, highly connected firm in the network: due to strategic complementarities from the many connections, that one firm would invest infinite amounts in R&D; in equilibrium however, its connections would also do so (again because of complementarities) and so on. In practice, explosive equilibria are not encountered in the real world, and the empirical results of this paper are consistent with the necessary condition for uniqueness.

The next result motivates the empirical strategy of this paper.

Proposition 2. *Suppose that Assumptions 1 and 2 hold. It follows that*

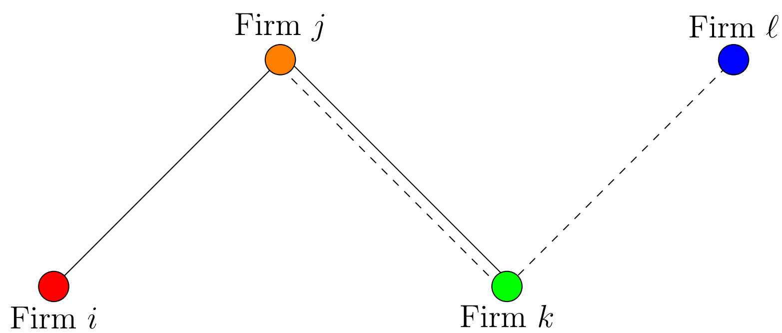
$$\mathbb{C}\text{ov}\left(\omega_i, \log S_j^* \middle| \, d_{ij} > C + L\right) = 0 \tag{6}$$

$$\mathbb{C}\text{ov}\left(\log S_i^*, \log S_j^* \middle| \, d_{ij} > C + 2L\right) = 0 \tag{7}$$

$$\mathbb{C}\text{ov}\left(\log S_i^*, \log S_j^* \middle| \, d_{ij} \leq C + 2L\right) \neq 0 \tag{8}$$

that is, the unobserved shock of one firm and the equilibrium R&D strategy of another one are independent as long as the two are distanced by a minimal path length higher than $C+L$; similarly, the equilibrium strategies of any two firms at distance higher than $C+2L$ are also independent, but may be correlated at distance $C+2L$ or lower.

Proposition 2 places a bound, in terms of "degrees of separation", on the equilibrium correlation across R&D choices and unobserved shocks in the network. The intuition is the following: even if in equilibrium firms endogenously internalize the shocks of other organizations that are "sufficiently close" (up to distance L), and this in turn amplifies the exogenous cross-correlation (up to distance C), if both mechanisms are bounded their combined effect also is. In other words the shocks of other firms that are "very distant" in the network, whose R&D investment is of little relevance, are never internalized by individual firms. An implication of this result is that, for any firm i, the R&D choices of firms that are "sufficiently distant" in the network can be used as exogenous predictors of the R&D investment of its own direct links, which are located at distance 1. Intuitively, the R&D of such "predicting" firms depends on some technological and cost factors also affecting the R&D of firm i's connections, but not the R&D of firm i itself. Given (6), (7) and (8), the appropriate "predicting" firms are located at any distance D between $C + L + 1$ and $C + 2L + 1$.



Graph 1: A Tetrad, or Two Semi-Overlapping Open Triads

An example of this is provided in Graph 1, which displays a network of four firms (i, j, k, ℓ): a *tetrad*. In fact, this graph is composed by two open triads[9] that partially overlap, as they share two nodes and an edge (the link between j and k). Consider first the simplest situation in which $C = 1$ and $L = 0$. In this case, firms only observe their private shocks, featuring spatial cross-correlation extending up to immediate connections, but not beyond. Therefore, R&D is correlated in equilibrium only across firms that are reciprocally connected, as it solely reflects private shocks. Consequently, from the point of view of firm i, the R&D of firm k (S_k^*) can serve as an "exogenous predictor" of firm j's R&D (S_j^*), because the two are correlated but the former is independent from firm i's R&D (S_i^*). However, the R&D of firm

[9]An *open triad* is a network (or a subset of a network) composed by three nodes; two of these three nodes are not connected to one another, but are both connected to the third one. In Graph 1 the two semi-overlapping open triads are represented by a solid and a dashed line, respectively.

ℓ, (S^*_ℓ) is not a valid predictor, as it is uncorrelated with that of firm j. Similarly, S^*_i exogenously predicts S^*_j from the point of view of firm k. The same properties symmetrically apply to the "dashed" triad made of nodes (j, k, ℓ).

Consider now some slightly more complex cases. If $C = 2$ and $L = 0$ firms are still unable to observe the shocks of others, but in this case the cross-correlation of R&D extends up to two degrees of distance as it reflects the primitive cross-correlation of the shocks. Hence, from the point of view of firm i, S^*_ℓ can act as a valid predictor of S^*_j; symmetrically S^*_i would predict S^*_k for firm ℓ. In the case where $C = 0$ and $L = 1$ the only mechanism driving R&D cross-correlation is the endogenous reflection of shocks, which can be observed between connected firms. Note that, in this case, the cross-correlation of R&D extends up to two degrees of distance. In fact, observe that both S^*_i and S^*_k depend on (ω_j, ϖ_j). Yet, S^*_k is still a valid predictor of S^*_j for firm i, as it is uncorrelated with ω_i – and vice versa. Observe how S^*_ℓ also correlates with S^*_j: they are both a function of (ω_k, ϖ_k). Thus, firm ℓ's R&D is a valid predictor of firm i's spillovers. The same logic applies when inverting the order of the nodes. Finally consider the case in which $C = 1$ and $L = 1$. Observe how R&D is correlated across the entire tetrad, but the R&D of firms at a distance of at least three degrees of separation are still valid predictors as per (6).

The result that follows is an immediate implication of Proposition 2.

Proposition 3. *Under Assumptions 1 and 2, the equilibrium conventional input choices of one firm are also uncorrelated with the equilibrium R&D of firms located at a distance higher than $C + 2L$.*

$$\mathbb{C}\text{ov}\left(\log X^*_{iq}, \log S^*_j \middle| \, d_{ij} > C + 2L\right) = 0 \text{ for } q = 1, \dots, Q \tag{9}$$

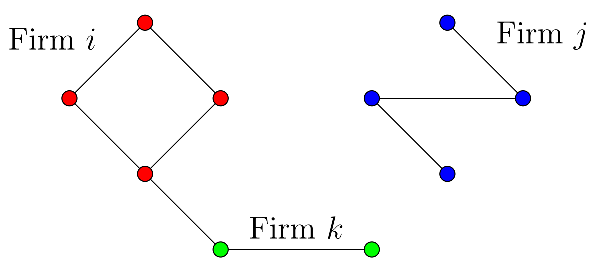
This result further supports the use of the R&D of firms that are distant enough as an instrument for the R&D of direct connections. Specifically, it motivates their exogeneity relative to other potentially endogenous control variables employed in the empirical analysis, as long as such instruments are taken at the furthest valid distance $C + 2L + 1$. The intuition is very simple: according the equilibrium conditions in (5), R&D and conventional inputs reflect the same information a firm knows about the state of the network. As the stochastic properties of both R&D and conventional inputs are a function of the same information set, the same bound applies to both relationships (7) and (9).

2 Networks and Data

This section is divided in three parts. In the first one, I formally introduce cross-firm *measures of connection* based on the underlying professional connections between the inventor teams of any two firms. In the second part, I describe the dynamic network of R&D spillovers calculated by measuring connections over time for each pair of firms from a panel of US companies. In the third part, I discuss some descriptive statistics that are relevant with respect to the empirical analysis, with a special focus on the spatial cross-correlation of R&D in the network.

2.1 The Measures of Connection

Assume that there are three R&D intensive firms whose scientists are related to each other even beyond the borders of their respective organizations. Denote as $\mathcal{M}_i$, $\mathcal{M}_j$ and $\mathcal{M}_k$ the sets of inventors belonging to each firm, with $\mathcal{M} = \mathcal{M}_i \cup \mathcal{M}_j \cup \mathcal{M}_k$. I define an existing *co-patenting relationship* between any two elements of $\mathcal{M}$, be they m and n, with the notation $p^t_{mn} = 1$. This indicates that two individuals, at time t, share some professional collaboration on *any past* research project that has resulted in a patent application featuring both their names. If such a relationship is absent, it is $p^t_{mn} = 0$. One could visualize the resulting network as a graph where each element of $\mathcal{M}$ is a node, and nodes are linked by edges if $p = 1$.



Graph 2: Inventors Network Example, $t = 0$

Graph 2 displays the first part of a stylized example on such a co-patenting network (hypothetically observed at some point in time $t = 0$). The inventors of each firm are nodes of the network displayed with different colors: red for i, blue for j, green for k. The co-patenting relationships p^0_{mn} are visualized as edges connecting two inventors. The only existing cross-firm co-patenting relation is that between an inventor of firm i and an inventor of firm k.

The central hypothesis of this paper is that firms learn about other firms' R&D activities thanks to the inventors who are connected to scientists in other firms, because of continuing professional relationships or more informal channels. A natural implication of such an assumption is that the tighter the connection is between two R&D teams, the stronger the spillovers are that occur between two organizations. For this reason I define measures that quantitatively capture such a differential effect. A *measure of connection* $c^f_{(ij)t}$ between, say, firm i and firm j at time t, is a monotonic function f of the fraction of inventors of either firm who are connected to inventors of the other firm, relative to the total size of both R&D teams:

$$c^f_{(ij)t} = f\left(\frac{\#\text{ inv.s of } i \text{ connected to } j \text{ at } t + \#\text{ inv.s of } j \text{ connected to } i \text{ at } t}{\#\text{ inv.s of } i \text{ at } t + \#\text{ inv.s of } j \text{ at } t}\right) \tag{10}$$

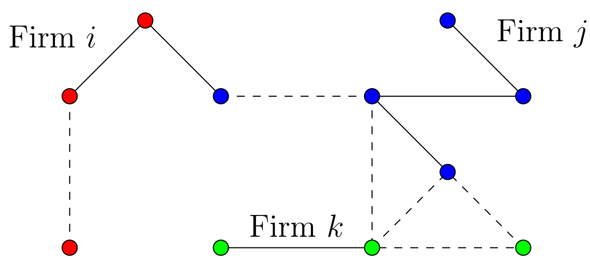
where $f : [0,1] \to [0,1]$, $f(0) = 0$ and $f(1) = 1$. These restrictions on f ensure that measures of connection take values between 0 (no connection) and 1 (full connection) as per the standard normalization of the strength of edges in a weighted network. For the three firms in the example of Figure 2, $c^f_{(ij)0} = c^f_{(jk)0} = 0$, while $c^f_{(ik)0} = f(1/3)$.

The facts that $c^f_{(ij)t} \in [0,1]$, and that any measure of connection is symmetric ($c^f_{(ij)t} = c^f_{(ji)t}$) bear important implications. The former means that an extra unit of external R&D cannot be more valuable for a firm than internally performed R&D, which is a reasonable hypothesis because in-house R&D is under direct control of the firm's management. The latter implicitly assumes that the spillover relationship is symmetric between the two firms i and j, regardless of the relative size of their R&D departments.[10] In addition, it must be stressed that such a connection measure essentially captures the relative number of personal relationships established in the past; it is silent about the relative importance of single linkages between inventors.[11] In Appendix E I characterize alternative definitions of connections which depart from these assumptions, and I explore the implications of their use in the empirical analysis.

[10]This is apparent from the example in Figure 2 where the two connected firms have different size. This assumption can have advantages: for example, it conveniently handles measurement errors in the assignment of individual inventors to firms. It may not be the most appropriate description of reality, however. For example, it might be that few "insiders" are enough to grasp much of another firm's knowledge. In such a case, the symmetric ratio in (10) would downplay spillovers received by the smaller firm in a pair, and an "asymmetric" measure would be better suited.

[11]Alternatively, one can assume that relationships between inventors that are prolonged over the years, or collaborations resulting in many joint patents, are more relevant than others. Similarly, connections involving superstar inventors who issue many patents, of which some have been extremely well cited, can be more valuable for a firm than linkages to "ordinary" inventors.

Connection measures between two firms can change over time. Their dynamics are the result of conceptually different types of events that are in principle observable, although I am not able to do so with the available data. Said events are: *i.* cross-firm R&D collaborations, such as joint ventures, resulting for instance in joint patents; *ii.* the movement of inventors between firms. Both situations are usually thought of as drivers of knowledge transfer between firms, and they positively impact measures of connection. In addition, *iii.* entry and exit of inventors from the network also affect the calculated metrics. However, their net effect is ambiguous and depends on the specific circumstances of the inventors in the process in question.[12]



Graph 3: Inventors Network Example, $t = 1$

Graph 3 extends the previous example by advancing one time period to $t = 1$, and examining the consequences of various changes in the underlying network of past patent collaborations. New linkages between inventors, due to newly appearing joint patents, are represented by dashed lines. In period $t = 1$ some incumbent inventors of firm j have been observed to patent jointly with researchers from firm k, including an entrant inventor from that company. A new entrant in firm i, not connected to anyone elsewhere, has also appeared. Instead, among firm i's incumbents one inventor has now moved to j, while the one who used to maintain the connection with firm k has exited the network. As a result, $c^f_{(ij)1} = f\,(1/4)$, $c^f_{(jk)1} = f\,(1/2)$ and $c^f_{(ik)1} = 0$.

In the applied analysis I employ connection measures based on the square root function.

$$g_{(ij)t} = \sqrt{\frac{\#\text{ inv.s of } i \text{ connected to } j \text{ at } t + \#\text{ inv.s of } j \text{ connected to } i \text{ at } t}{\#\text{ inv.s of } i \text{ at } t + \#\text{ inv.s of } j \text{ at } t}} \tag{11}$$

This choice responds to a precise economic assumption. The typical anecdotal narra-

[12]New entrants increase the denominator of (10), but can also generate new cross-firm linkages, tightening connections. Similarly, the exit of scientists can decrease the denominator of (10), as well as the numerator if the leaving inventors were playing the role of connecting firms to each other.

tive on technological spillovers usually involves some solitary individual who transfers, perhaps by mistake, much of the knowledge internally developed by one firm to some of its partners or competitors. The very expression "spillovers" is verbally associated in such anecdotes to the "leakage" of few accumulating "drops" of knowledge. By applying the square root function to the ratio of connected inventors, I attribute more importance to the pairs of firms with relatively fewer connections. In the remainder of this paper I use the expression "connection" to indicate the squared root metric. In Appendix E I present the empirical results from applying alternative definitions of connections based on different choices of f.

2.2 Firm-level Network

In the empirical analysis I combine different data sources. The firm-level network is constructed from the data assembled by Bloom et al. (henceforth BSV) for their cited study. This is an unbalanced panel[13] consisting of 736 mostly manufacturing, R&D-intensive firms listed on the US stock market, observed over the years 1976-2001. The BSV dataset combines accounting data from COMPUSTAT, firm-level patent counts, as well as Jaffe-type measures used by BSV to disentangle different types of spillover effects. Via firm and patent identifiers, I match the BSV data to the "disambiguated" patent dataset by Li et al. (2014). The latter provides unique identifiers for individual inventors across different patents, thanks to a disambiguation algorithm that exploits information available in the USPTO database. Ultimately, this results in the selection of 1,315,060 patents granted to 565,019 inventors.

To calculate the connection measures, I need to associate inventors to each other as well as to firms. The first task is accomplished by looking at jointly filed patents. Specifically, for two inventors m and n, I assign $p_{mn}^{t} = 1$ if at time $t+1$ the USPTO has received at least one patent *application* (to be eventually granted) filed *at any time in the past* by both inventors. The implicit assumption is that the two inventors are involved in a professional relationship at least one year prior to the application.[14] Similarly, in order to assign inventors to firms one has to extrapolate facts on the basis

[13] The panel is unbalanced because of the entry and exit of firms, as they go public or are subject to mergers and acquisitions (see the description in BSV's appendix). The average length of the panel is 17.8 years.

[14] Given the lag structure of R&D outcomes (patents) it is likely that this is an overly restrictive assumption. On the other hand, it is desirable to avoid assigning relationships that did not exist in reality. The results are very robust to perturbations of this assignment rule.

of limited available information. I use the sequence of patents co-filed by inventor m and assigned to firm i in order to define a time interval in which one can reasonably presume that the individual was crucial for the R&D activity of that organization. The details of the assignment rule are provided in Appendix C.

I calculate measure (11) for each pair of firms and for every year from 1981 to 2001, using patents granted since 1976. As calculating connections requires the observation of enough antecedent patents, I abstain from doing it for 1976-1980; since the BSV panel is thin in those years this is a small loss. In total 460 firms display at least one positive connection with another firm in any year of the time interval under analysis. The number of firms that are actually connected in any year varies over time: some of the initially unconnected firms would eventually develop bonds. Conversely, firms that are already connected in 1981 may experience variations in the number of their connections (possibly resulting in the loss of all of them), or leave the sample. Thus, one never observes all the 460 firms of the dynamic network in each cross section.

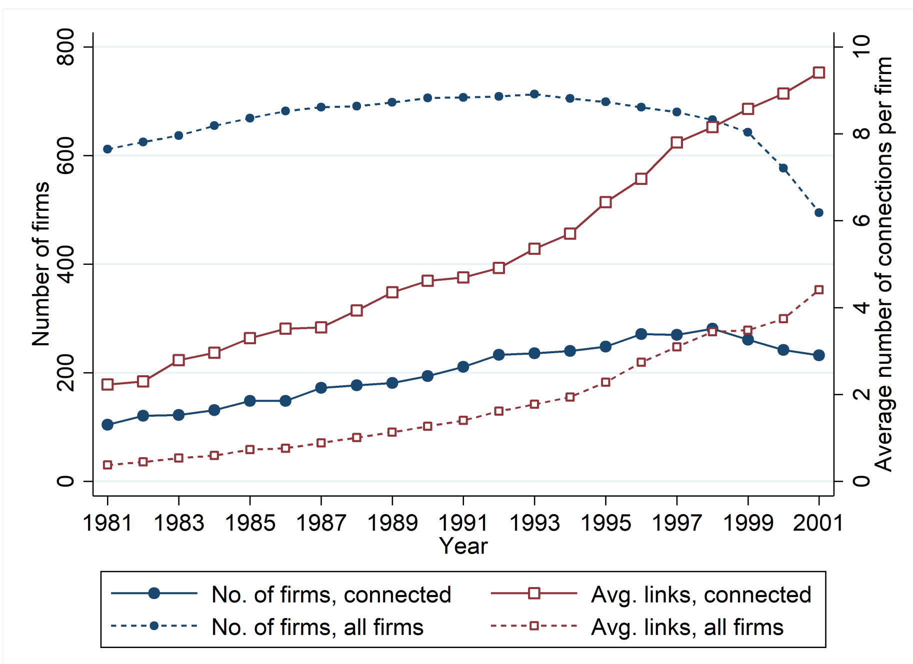


Figure 1: Firms, Connected Firms and Average Connections over time

Figure 1 portrays, for every year from 1981 to 2001, the overall number of firms in the original panel (blue dashed line), the number of those displaying any nonzero connection (blue continuous line), as well as the average number of connections per firm, both in the whole sample and for the subset of "connected" firms (red dashed and continuous lines, respectively). Figure 1 displays a steady increase in the number of connected firms between 1981 and 1998, followed by a drop from 1998 to 2001 – partly because of attrition in the original panel, which is particularly severe in later years. Among connected firms, the average number of connections increases steadily over the entire time frame.[15] By construction, this is also reflected by the "unconditional" average represented by the dashed red line. Another way to appreciate the temporal evolution of the network is to visualize it in the form of graphs; selected graphs for the years 1985, 1990, 1995 and 2000 are reported in Appendix D.

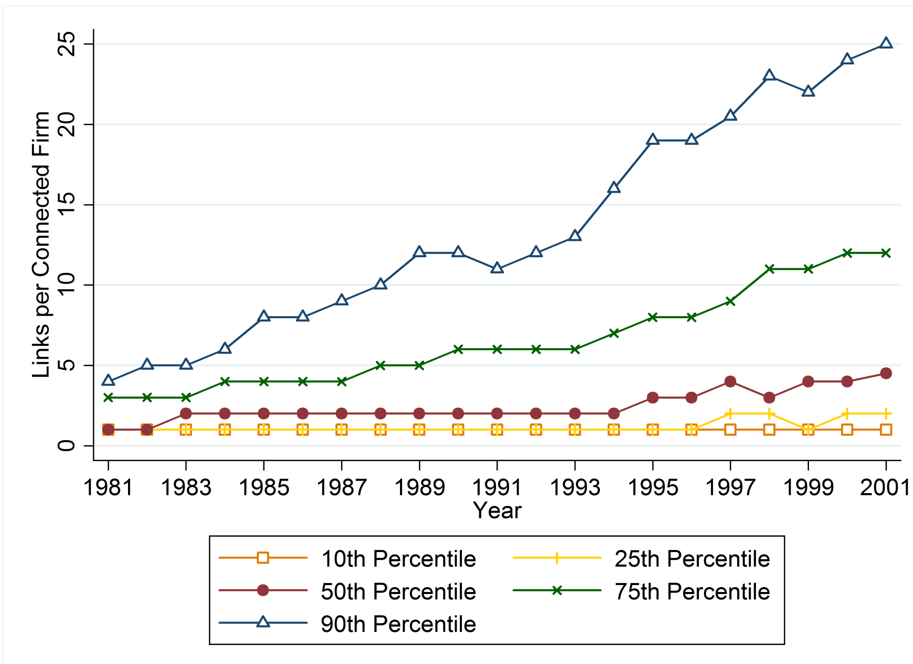


Figure 2: Degree distribution over time (connected firms)

Figure 2 summarizes the yearly distributions in the number of connections per firm (called "degree") of connected firms. Like in most networks, the degree distribution is very asymmetric; moreover it tends to widen over time. The most connected firms go from less than 10 links in the early 80s to but several dozens of them by the year 2000, while the median number of connections only increases from 1 to 5. Connections

[15]This increase can be attributed to several factors, for example: the mechanical effect due to the overall increase in the number of patents over the two decades, the diffusion of R&D joint ventures, and the emergence of collaborations between universities and firms, linking inventors from various firms together. More informative patent data are necessary to distinguish between these mechanisms.

themselves are also asymmetrically distributed, as shown by Figure 3 (however, their distribution looks quite stable over time). The average of $g_{(ij)t}$ is 0.083, with 0.066 standard deviation.[16] A measure useful for interpreting the empirical results is the *row sum* of connections, that is the sum of all of one firm's connections in one year: $\bar{g}_{it} = \sum_{j \neq i} g_{(ij)t}$. Among connected firms, the mean and standard deviation of $\bar{g}_{it}$ are respectively 0.50 and 0.57. The yearly empirical distribution of $\bar{g}_{it}$, not shown for brevity, spreads out over time mirroring the dynamics of the degree distribution.

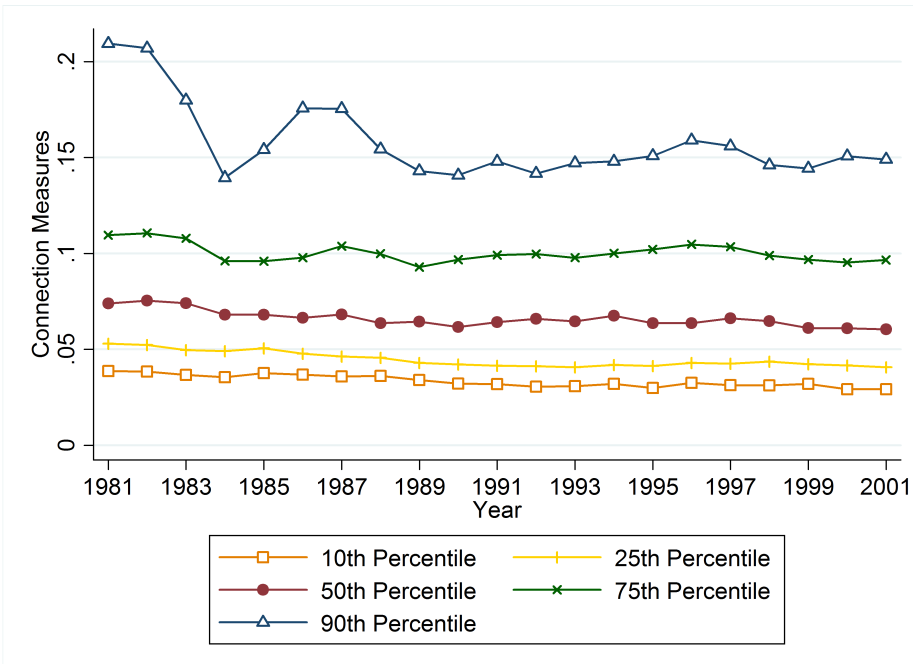


Figure 3: Distribution of the connections ($g_{(ij)t}$) over time

Figure 4 displays the temporal evolution of the *Triad Census*, namely the total count of open and closed triads.[17] Between years 1981 and 2001, one counts in total 160,365 open triads and 15,623 closed triads; the latter are about 10% of the former. The number of both types of triads grows over time in analogy with average degree. The preponderance of open triads indicates that the network does not feature an excessive degree of clustering for the purpose of this article's empirical strategy. If most triads were closed, in fact, it would not be possible to identify "indirect connections"

[16] Recall that this refers to the squared-root connection measure as defined in (11). The average for baseline linear measure is 0.012, with 0.028 standard deviation.

[17] In analogy with the definition of open triad, a closed triad is a (sub)network whose three nodes are all connected to one another.

(other firms located at distance 2 or higher) for many firms in the network.

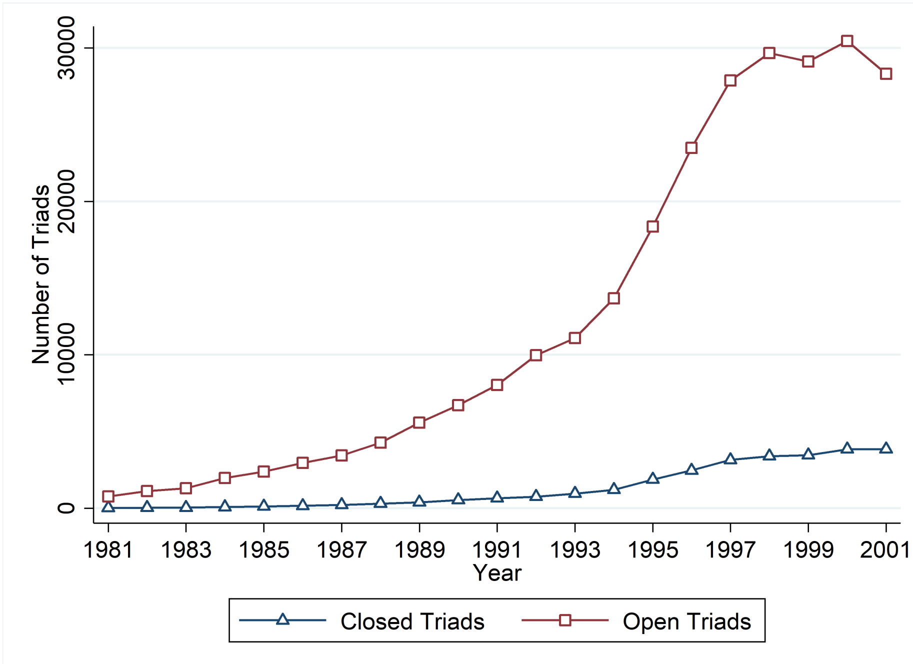


Figure 4: Triad Census

2.3 Summary Statistics and Spatial Correlation

In Table 1 I report some firm-level summary statistics. To this end I split the sample into five groups: one for the firms that never enter the network, and four groups for those that do. Specifically, I calculate the *overall* sum of connections for each firm as $\bar{\bar{g}}_i = \sum_t \bar{g}_{it}$, and I assign each firm to a group on the basis of its classification within quartiles of $\bar{\bar{g}}_i$. Quartile 1 contains the least connected firms in the network over the time interval; while quartile 4 contains the most connected ones. For each of these groups, I report the mean and standard deviation of specific variables by pooling all years in the sample. In addition to real sales Y_{it}, other outcome measures (Tobin's q, citation-weighted patents P_{it}) and number of employees E_{it}, I also report – for easier interpretation – the *ratio* of Y_{it} to several input or spillover measures. Finally, the last row reports the ratio of Y_{it} on the connections-based spillover component of knowledge capital, that is the main independent variable of the analysis. Table 1 highlights the fact that firms inside the network – in particular the most connected among them – are larger, more R&D intensive and more productive than those outside of of it.

Table 1: Summary Statistics, 1981-2001

	No Network	Quartile of $\sum_t \bar{g}_{it}$ 1	2	3	4
Y_{it}: Sales (Millions 1996$)	751	1066	1383	2172	10462
	(3792)	(2357)	(2504)	(4533)	(20058)
V_{it}/A_{it}: Tobin's q	1.886	1.885	2.573	2.734	3.410
	(2.031)	(1.839)	(3.080)	(3.306)	(4.118)
P_{it}: Patent Stock (cit. weighted)	7.453	16.09	24.65	74.03	652.0
	(48.17)	(44.75)	(50.91)	(143.8)	(1322.1)
E_{it}: Employees (Thousands)	4.068	6.940	9.328	12.40	57.09
	(12.52)	(15.80)	(16.63)	(22.43)	(96.80)
Y_{it}/E_{it}: Labor Productivity	135.6	134.5	157.1	156.5	192.4
	(80.06)	(106.6)	(95.43)	(117.7)	(153.3)
Y_{it}/K_{it}: Capital Productivity	6.932	5.308	5.142	4.941	4.184
	(6.083)	(3.167)	(3.992)	(3.292)	(2.883)
$Y_{it}/R\&D_{it}$	39.31	19.71	51.10	11.12	4.342
	(134.1)	(70.47)	(479.9)	(34.46)	(3.932)
$Y_{it}/$ Jaffe Measure (i,t)	80.28	107.7	140.0	211.6	962.5
	(407.7)	(238.5)	(264.9)	(435.4)	(1787.8)
$Y_{it}/\prod_j R\&D_{jt}^{g(ij)t}$		953.9	846.2	577.6	198.9
		(2224.0)	(1762.1)	(1858.7)	(1339.6)
No. of Observations	4363	1854	1819	1949	2028

Notes: The table is divided in five columns: one for firms in the BSV sample that are never part of the network, and four for each quartile of $\bar{\bar{g}}_i$. The four quartile groups contain different numbers of observations because of panel attrition. $R\&D_{it}$ denotes the R&D *stock* of firm i in year t. All descriptive statistics are calculated by pooling observations over years; means and standard deviations (in parentheses) are reported.

In light of the empirical strategy adopted in the paper, an important set of descriptive statistics that is worth examining is the empirical spatial cross-correlation of R&D between firms in the network. This is reported in Figure 5 in the form of the Moran's I statistic, which is calculated for both R&D flows and R&D stocks across different degrees of separation (distances) in the network. Moran's I statistic, a standard tool in spatial analysis, consistently estimates the spatial correlation of a given variable of interest at a given level of distance (Kelejian and Prucha, 2001). I perform the calculation by pooling together all pairs of firms at the same level of distance throughout all the years. Figure 5 illustrates a strong correlation for direct connections (distance 1), a correlation of half strength for indirect links (distance 2) and zero correlation for all further distances: this is a typical pattern encountered in many other real-world networks (Christakis and Fowler, 2013). The correlation of R&D stocks is mechanically weaker than the one of R&D flows, as it accounts for past time periods when two firms were not connected.

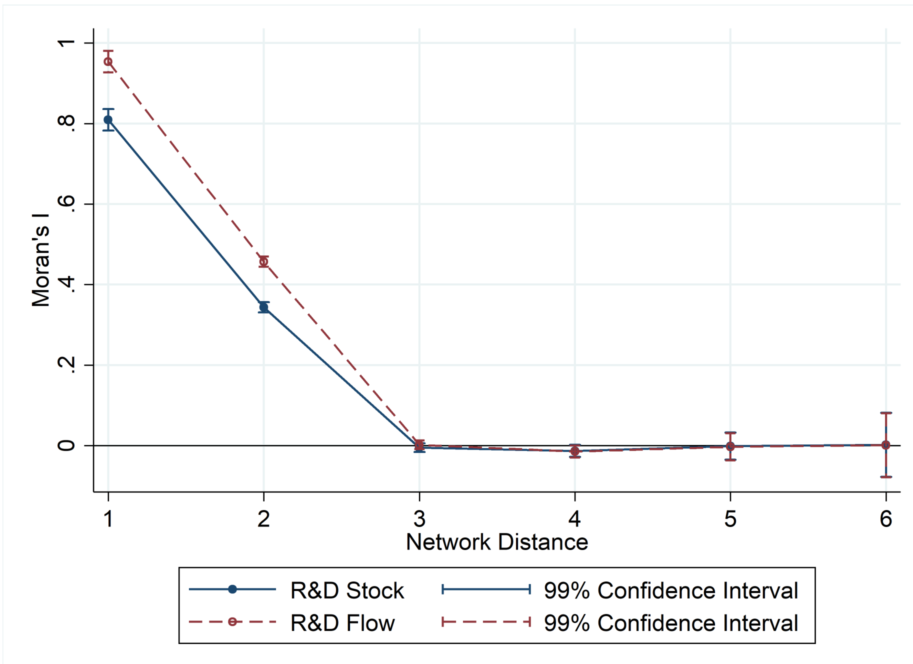


Figure 5: Spatial Correlogram of R&D Measures, 1981-2001

According to the analytical framework of the paper, the spatial cross-correlation of R&D reflects the exogenous cross-correlation of firm-specific characteristics (ω_i, ϖ_i), the endogenous strategic dependence between firms' R&D choices, or both. In light of Proposition 2, the evidence in Figure 5 is compatible either with a situation where $(C, L) = (0, 1)$ or one in which $(C, L) = (2, 0)$ (this is analogous to the analysis of time series correlograms generated by MA-types of processes). In the former case, the R&D of firms located at either distance 2 or at distance 3 are valid predictors of direct connections' R&D. In the latter, only indirect links at distance 3 can function as appropriate predictors. Consequently, in the empirical analysis I experiment with instruments constructed by aggregating the R&D stocks of indirect connections at both levels of distance. Instruments based on higher distances present no correlation with the R&D of direct connections, as evidenced by Figure 5.

3 Econometric Model

This section concerns the empirical methodology of this article, and it is divided in four parts. In the first one, I discuss the workhorse model for the evaluation of R&D spillovers on productivity effects: an augmented production function. In the second part I describe the empirical strategy for addressing endogeneity. In the third part I introduce models for the estimation of spillovers on firms' market value and innovation rate. Finally, in the fourth and last part I illustrate the clustering approach for the calculation of standard errors, which is common to all empirical models.

3.1 Production Function

The workhorse empirical model of the empirical analysis is an augmented production function. It is the empirical counterpart of equation (2) adapted to panel data:

$$\log Y_{it} = \alpha_i + \sum_{q=1}^{Q} \beta_q \log X_{itq} + \gamma \log S_{it} + \delta \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tau_t + \upsilon_{it} \tag{12}$$

where the unobserved shock ω_{it}, which is now allowed to vary across firms and over time, is decomposed as $\omega_{it} = \alpha_i + \tau_t + \upsilon_{it}$, that is by a firm-invariant effect (α_i), a year effect (τ_t), and finally a residual error term (υ_{it}). Here S_{jt} denotes the R&D *stock* of firm j at time t, and $g_{(ij)t}$ is the connection measure between firms i and j at time t, with $g_{(ii)t} = 0$ for all i and for all t. The R&D stock S_{it} is constructed, following a customary approach in the literature, as the depreciated sum of past expenditures on R&D up to year $t-1$ (Griliches, 1998).[18] To account for the known fact that the innovation and productivity effects of R&D materialize with a temporal lag, current expenditures in R&D are excluded from the calculation of the yearly stock.

Parameter δ represents the overall strength of the R&D spillovers in the network. It is interpreted as the elasticity of a connection-weighted neighbor's R&D on firm productivity. It is useful for different kinds of thought experiments: for example, a firm i connected to a neighbor j with connection $g_{ij} = 0.4$ receives a 0.4δ percentage increase in productivity following a 1% increase in the R&D stock of firm j. Similarly, a firm with row sum of connections $\bar{g}_{it} = 4$ benefits from a 4δ percentage increase in

[18] In the BSV dataset the R&D stock is constructed, like in other studies, by applying the perpetual inventory method with a 15% depreciation rate. The R&D stock measure employed in this analysis reads as $S_{it} = R_{i(t-1)} + .85 S_{i(t-1)}$, where R_{it} is the *flow* of R&D investment at time t.

productivity following a 1% rise in the research effort of *all* its neighbors. By contrast, parameter γ measures the elasticity of firm productivity with respect to changes in the private (in-house) R&D stock.

Since actual physical quantity Y_{it} is not observed in the BSV dataset, I proxy it by the deflated sales of firm i in year t, as it is customary in studies dealing with the estimation of production functions. Deflated sales, however, conflate both supply and demand factors. To control for the latter, like BSV I include in the specification of (12) some industry-level market outcomes, namely current and lagged industry sales, and current industry prices. In addition to demand side controls and conventional inputs (capital and labor in year t), I also include other R&D spillover variables on the right-hand side of (12). Their purpose is to more convincingly restrict the interpretation of δ to the sole effect of other firms' R&D induced through the collaboration network. For simplicity, in expression (12) I do not explicitly distinguish between conventional inputs and other demand or spillover controls; with some abuse of notation I treat all of these as different elements of the set of covariates $\{X_{itq}\}_{q=1}^{Q}$.

The additional R&D spillover variables deserve further elaboration. Two of them are the key regressors of the BSV study: Jaffe's classical measure of (beneficial) R&D spillovers, and BSV's original measure of "business stealing," which aims at capturing the negative externality in terms of residual demand that is due to market rivals' R&D. Both are constructed weighing external R&D S_j by some metric of similarity between firm j and a reference firm i: the correlation of patent technological classes in the case of Jaffe's measure, and that of sales' allocation across sectors for the BSV business stealing measure. In addition, I include one spillover variable which accounts for the geographic co-location of firms' R&D, so to control for the possibility that connections $g_{(ij)t}$ simply capture firms' spatial proximity and other spatially correlated factors. To construct this measure, I weigh external R&D by a *measure of proximity* analogous to (11): the square root of the relative proportion of two firms' inventors who reside in the same statistical metropolitan area (CBSA). Appendix C provides additional details and descriptive statistics about this "Geographic Spillovers" measure.

3.2 Endogeneity and Instrumental Variables

The estimation of δ in equation (12) suffers from two potential endogeneity problems. The first is the possible presence of common confounders driving both R&D choices and productivity for connected firms, corresponding to the *correlated effects* as per the analysis by Manski (1993) of spillovers in the classroom. Formally, if for any two connected firms i and j both $\mathbb{E}\left[\upsilon_{it}\upsilon_{jt}\right] \neq 0$ and $\mathbb{E}\left[\log S_{it} \cdot \upsilon_{it}\right] \neq 0$ hold simultaneously, it follows that $\mathbb{E}\left[\log S_{jt} \cdot \upsilon_{it}\right] \neq 0$ and OLS estimates of δ are inconsistent. In the production function context, this is a spatial generalization of the classical "transmission bias" due to unobserved shocks. The second problem is endogeneity of connections: the possibility that firms differ in their propensity to establish connections as a function of their unobserved shocks ($\mathbb{E}\left[g_{(ij)t}\upsilon_{it}\right] \neq 0$ for at least some connection j). For example, it is likely that highly prolific, well connected inventors are more inclined to move towards more productive firms.[19]

The analysis conducted in Section 1 suggests a strategy to address both problems at the same time: to predict the R&D of a firm's direct connections with the R&D of other firms that are "sufficiently" distant in the network. For two appropriate firms i and k, equation (6) from Proposition 2 can be recast in a panel data setting as:[20]

$$\mathbb{E}\left[\log S_{kt} \cdot \upsilon_{it} \middle| h^{D}_{(ik)t} = 1\right] = 0 \tag{13}$$

where $h^{D}_{(ik)t} = \mathbb{1}\left[d_{(ik)t} = D\right]$ is a dummy variable indicating that firms i and k are located, in year t, at a given distance $C + L + 1 \leq D \leq C + 2L + 1$. Because of (13), by iterating expectations one gets:

$$\mathbb{E}\left[h^{D}_{(ik)t} \log S_{kt} \cdot \upsilon_{it}\right] = 0 \tag{14}$$

which is a valid *and* relevant moment condition for the identification of δ, as Equation (8) from Proposition 2 also guarantees that the instrument correlates with the R&D of some of firm i's direct connections.

The econometrics of moment condition (14) deserves more discussion. Intuitively, the reason why (14) addresses the common shocks problem is that for appropriate D, $\log S_{kt}$ is generated by some variation that is independent of υ_{it}. Moreover (14) is

[19] In such a situation it is unclear whether the resulting bias would be positive, because moving inventors link together both high- and low-productivity firms.

[20] Firms accumulate R&D stock (knowledge capital) over time. In Appendix B I discuss conditions under which Proposition 2 can be extended to a dynamic model. In particular, it is necessary that the network tends to become denser instead of sparser (intuitively, if distances are lengthened (13) might not hold, because both $\log S_{kt}$ and υ_{it} would preserve the memory of past linkages). Moreover, if errors are serially correlated and firms do not pre-commit to long term R&D plans, firms must be unable to *observe* past R&D choices of sufficiently distant firms ("finite spatial memory").

also robust to network endogeneity because direct connections $g_{(ij)t}$, which are likely to depend on υ_{it}, are simply absent from the expression of the moment. Therefore, identification of δ is obtained through variation of external R&D which is "as good as exogenous"; within this framework, taking the network as given is innocuous. It is useful to draw a parallel between this approach and GMM methods that are typical of the dynamic panel literature (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). In the latter, following a transformation of the model aimed at removing fixed effects, endogenous lagged variables of interest are instrumented by "sufficiently past" further lags. Similarly, here I instrument the possibly endogenous first spatial lag of R&D by "sufficiently distant" farther spatial lags of R&D.

In terms of practical implementation, it must be noted that it is generally possible to find – for each firm i – multiple other firms k located at distance D. Each of them corresponds to a moment condition expressible as (14); all of these moments can be empirically operationalized in different ways, for example as independent conditions in a GMM problem.[21] However, in this analysis I opt for a simple approach: I aggregate all those moment conditions linearly, taking the summation of (14) over all firms in year t. This results in a single moment condition:

$$\mathbb{E}\left[\sum_{k=1}^{N} h_{(ik)t}^{D} \log S_{kt} \upsilon_{it}\right] = 0 \tag{15}$$

which amounts to selecting the log-R&D of all "indirectly connected" firms k located at distance D, giving each an equal weight in the summation. As I illustrate in the next section, in order to identify δ I employ moment condition (12) and variations of it in both simple IV-2SLS and System GMM estimates of the production function.

This still leaves the question about the appropriate value of D to be chosen for estimation open. The previous discussion of Figure 5 suggests selecting "indirect links" located at both distances 2 or 3, or just distance 3. It is important to discuss how the choice of D relates to the data employed in the analysis. For moment condition (15) to be credible, it is necessary that such values of D be large enough to separate out firms with different unobservables. In fact, firms from the BSV sample are large multi-business, multi-technology companies, presenting only weak correlation between their patent and product portfolios (see the description in their article). Therefore, while

[21] In this case, it would be interesting to analyze how the optimal GMM weighting matrix varies as a function of network topology. This is an intriguing topic for further research.

two connected firms are likely to share some characteristics, they are also expected to be different in other respects. Since any spatial cross-correlation that is stronger at low distances must mechanically decay at higher distances, it is likely that common similarities are at best negligible at distance 3. This argument provides an empirical rationale for the indirect evidence given in Figure 5 and the choice of $D = 3$.

3.3 Additional Outcomes

In empirical studies of R&D spillovers, it is customary to assess the effect of other firms' R&D not only on output or productivity, but also on other outcomes and indicators of firm performance and innovation rate. In his seminal study, Jaffe (1986) also measured the effect of spillovers on firms' market value and patent output. BSV follow in his legacy. Under their shared theoretical framework, especially under the maintained hypothesis of R&D as a strategic complement, spillovers stimulate R&D efforts and increase the number and quality of inventions. The effects on productivity can be both indirect (thanks to new or better patents/products) and direct (because of the immediate applicability of spilled knowledge within the production process). This ultimately results in better firm performance and increased market value.

I follow suit and measure the effect of the R&D performed by "connections" on outcomes other than output or productivity by extending the empirical specifications by BSV. Their market value specification is a standard semi-logarithmic Tobin's q model (Jaffe, 1986; Hirsch and Seaks, 1993) augmented as follows:

$$\log\left(\frac{V_{it}}{A_{it}}\right) = \tilde{\alpha}_i + \sum_{q=1}^{\tilde{Q}} \tilde{\beta}_q W_{itq} + \tilde{\delta} \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tilde{\tau}_t + \tilde{v}_{it} \tag{16}$$

where V_{it} is the market value of a firm measured at time t and A_{it} is the replacement value of its assets. The set of controls $\{W_{itq}\}_{q=1}^{\tilde{Q}}$ includes a sixth degree polynomial of the ratio $S_{i(t+1)}/K_{it}$ to control for differences in R&D intensity,[22] current and lagged industry-level sales, as well as the three additional spillover variables (Jaffe's, BSV's business stealing, and the geographic measure). Since this is a linear model, I address endogeneity through the same IV strategy outlined above for the production function.

[22] In semi-logarithmic Tobin's q models, a higher degree polynomial of S/K is a customary approximation to some unknown non-linear function of the R&D stock to capital ratio. Note that the R&D stock is taken at year $t+1$: like BSV, I let market value depend on current R&D investment, as investors may price in the future expected return of R&D.

The effect of connections' R&D on firms' innovation rate is estimated instead via a count model for citation-weighted patents P_{it}:

$$P_{it} = \exp\left(\breve{\alpha}_s + \sum_{q=1}^{\breve{Q}} \breve{\beta}_q Z_{itq} + \breve{\gamma}\log S_{it} + \breve{\delta}\sum_{j=1}^{N} g_{(ij)t}\log S_{jt} + \breve{\tau}_t + \breve{\upsilon}_{it}\right) \tag{17}$$

which, to account for values of $P_{it} = 0$, is specified as a negative binomial model and estimated via maximum likelihood (Hausman et al., 1984; Blundell et al., 1995).[23] The set of controls $\{Z_{itq}\}_{q=1}^{\breve{Q}}$ includes a term for the lagged dependent variable ($\log P_{i(t-1)}$) as well as the three extra spillover variables. To address endogeneity in a non-linear model I adapt my IV strategy to a control function approach. Specifically, I regress the endogenous variables on the excluded instruments and the other controls, and I include sixth-degree polynomials of the resulting residuals into (17).

3.4 Standard Errors

It is important to consider that the cross-correlation across unobserved shocks invalidates standard asymptotic properties of any GMM/2SLS or MLE estimator. As a theory of heteroscedasticity-autocorrelation consistent (HAC) estimators in the case of network dependence has not been developed yet, I adopt a transparent clustering approach that is consistent across both linear and non-linear models.[24] Specifically, I follow Bester, Conley, and Hansen (2011), who argue that even in presence of weak dependence between groups, a clustering covariance estimator (CCE) of the estimates' variance would make for valid inferences (provided that some regularity conditions hold and that small sample corrections are applied).[25] This is particularly important with large networks, because if the structure of cross-node dependence is unknown,

[23] To guarantee convergence of the estimation algorithm, it is convenient not to include firm-specific fixed effects. I introduce four-digits industry fixed effects instead, that are indexed by s.

[24] For spatial data, the standard HAC procedure is the one proposed by Conley (1999), originally conceived for cross-sectional data distributed on a regular lattice defined by a system of coordinates (like locations on a map). However, networks are inherently multidimensional, and actual network data may feature competing notions of distance. Thus, an extension of Conley's HAC procedure to networks is not straightforward. In addition, in the context examined in this paper the data are likely to display both spatial and serial correlation, which would result in very complicated Bartlett-like HAC estimators. A clustering approach is well suited to simultaneously address both issues.

[25] In their simulations, Bester, Conley, and Hansen (2011) show that in both cases of time series and spatial dependence, tests based on Bartlett-like HAC estimates of the variance tend to incorrectly reject relevant null hypotheses considerably more often than tests based on their CCE approach. This difference is particularly pronounced in the case of spatial dependence.

any partition of a network into different clusters would result in some cross-cluster dependence. Bester et al. advocate using as few and as large clusters as possible.

I divide the network into "communities" or clusters by running the "Louvain algorithm" (Blondel et al., 2008) on the "pooled" network that is obtained by summing the same edges over the time series. The Louvain algorithm is a popular tool in network analysis used to identify hierarchies of "communities" or clusters. At every level of the hierarchy, connections are dense within groups and sparse between groups. The algorithm can be fine-tuned by varying the "resolution parameter" φ which selects different levels of the hierarchy.[26] To strike a balance between the CCE approach by Bester et al. and standard practices of clustering standard errors, I set $\varphi = 0.6$ so to obtain 20 clusters. Because of serial correlation, all observations of the same firm in the panel enter the same cluster; for estimates that include firms outside the network, each of these constitutes an additional cluster. Appendix D provides a visualization of cluster assignment. Statistical inferences are not substantially altered by the average size or number of clusters.

4 Empirical Results

In this section I present the empirical results of the paper. This section is divided in five parts. In the first part I present the baseline (OLS) estimates of the production function. In the second part, I discuss the results obtained from the proposed IV and System GMM strategies aimed at addressing endogeneity. In the third part I present the results for the firm value equation; in the fourth those for the patent count model. Finally, in the fifth part I discuss the economic relevance of the estimated effects.

4.1 Production Function, OLS

Table 2 displays the results from the estimation of equation (12). Across all estimates I take both firm and year fixed effects; and I cluster standard errors according to the approach based on network "communities" outlined above. Along with the estimate of γ and δ I report those for Capital and Labor. The estimate of spillovers $\hat{\delta} = 0.016$ from column (1) can be interpreted in light of different thought experiments. For example,

[26] A large value of φ defines a partition of few large communities; smaller values of φ break down these clusters and define smaller groups by moving down the hierarchy.

the quantity $\hat{\delta} g_{(ij)t}$ represents the elasticity of output with respect to a 1% increase in the R&D stock of another firm with connection $g_{(ij)t}$. In the case of an average connection $g_{(ij)t} = 0.083$, the implied elasticity is 0.0013. Hypothesizing instead a 1% increase in the R&D stock of *all* of one firm's neighbors, the implied effect on firm i's output is a $\hat{\delta}\bar{g}_{it}$% rise. For a firm connected to the network with average row-sum $\bar{g}_{it} = 0.50$, this corresponds to a 0.008% increase. The estimated elasticity of private R&D, $\hat{\gamma} = 0.045$, appears in comparison to be one order of magnitude larger.

Table 2: Production Function, Ordinary Least Squares Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0455***	0.0438***	0.0568***	0.0554***	0.0515***
	(0.0108)	(0.0105)	(0.0118)	(0.0128)	(0.0142)
R&D Spillovers (δ)	0.0159***	0.0147***	0.0114***	0.0118***	0.0116***
	(0.0023)	(0.0024)	(0.0024)	(0.0025)	(0.0028)
Geographic Spillovers		0.0035	0.0027	0.0023	0.0015
		(0.0021)	(0.0020)	(0.0023)	(0.0019)
Capital	0.2071***	0.2061***	0.2035***	0.2020***	0.2023***
	(0.0143)	(0.0145)	(0.0194)	(0.0213)	(0.0198)
Labor	0.6550***	0.6580***	0.6634***	0.6613***	0.6622***
	(0.0241)	(0.0249)	(0.0351)	(0.0359)	(0.0363)
Jaffe Tech. Proximity		0.1352**	0.0361	0.0026	0.0179
		(0.0581)	(0.0583)	(0.0766)	(0.0805)
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	NO	YES	YES	YES
No. of Communities					
(Community × Year Effects)	0	0	0	10	20
No. of Observations	12503	12503	7607	7607	7607

Notes: The table reports OLS estimates of model (12). Columns 1 and 2 are estimated over the entire original sample of 736 firms in the time interval 1981-2001. Estimates in columns 3, 4 and 5 restrict the sample to firms with at least one nonzero connection ($g_{(ij)t} \neq 0$) in any year t; all observations of these firms are also included for years with no connections. All estimates include firm and year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with $\varphi = 0.8$ (10 communities) in column 4 and $\varphi = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. For estimates not restricted to the network, firms outside the network constitute single clusters. Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

Relative to column (1), in (2) I show the effect of controlling for the Jaffe measure of knowledge spillovers based on technological proximity, as well as for the geographic

R&D intensity measure. The inclusion of both does not dramatically impact the point estimate $\hat{\delta}$, which falls to about 0.015 while remaining statistically significant. The geographic control, on the other hand, seems to have very little economic significance. To control for the possibility that the estimate $\hat{\delta}$ is driven by persistent productivity differences between firms that belong to the network and those that do not, in column (3) I report estimates restricted to the subsample of firms that enter the network at any point in time. This exercise has an interesting implication: while parameter $\hat{\delta}$ is estimated substantially smaller (down to 0.011) yet still statistically significant, the coefficient for the Jaffe measure of spillovers falls more sharply becoming no longer statistically significant.[27] As firms that do not belong to the network are the smallest and least R&D-intensive ones, this result implies that the positive correlation between real sales and the Jaffe measure is largely driven by small firms patenting in the most R&D-intensive technological fields.[28]

In estimates (4) and (5) I also include an additional set of dummy variables, in an initial attempt to control for the fact that connected firms may be subject to similar shocks. Specifically, I absorb community-by-year effects, where communities are constructed by applying the Louvain algorithm with varying resolution parameters. In particular, in column (4) I employ a network partition of 10 communities ($\varphi = 0.8$); while in column (5) the additional dummy variables are based on the same 20 communities also used for clustering standard errors ($\varphi = 0.6$). Increasing the number of clusters does not dramatically affect the point estimate $\hat{\delta}$. This suggests that the correlation between the connections-induced measure of spillovers and one firm's output is in fact driven by the variation in the R&D stock of that firm's linkages.

4.2 Production Function, IV and GMM

I now illustrate the empirical results from the application of the IV strategy aimed at addressing correlated confounders and network endogeneity. I instrument the R&D stock of one firm's direct connections by aggregating the R&D of its "indirect con-

[27] In the analysis by BSV, their "business stealing" measure has no significant effect on productivity.

[28] This finding can be interpreted as a sample selection bias. COMPUSTAT only reports data for public firms; small firms going public are usually successful firms, and those that "make it into the news" are typically from fast developing high-tech sectors (and being in the news is itself endogenous). If a correlation exists between the Jaffe measure and the probability that small firms go public, this would be reflected in a positive bias in the estimate of the Jaffe measure when small public firms are included in the estimation sample. This issue certainly deserves further attention.

nections" located at distance 2 and 3. In light of the theoretical analysis and of the evidence on the spatial autocorrelation of R&D in the network presented in Figure 5, both instruments could be valid in principle. However, the farther instrument constructed at distance $D = 3$ is more likely to be uncorrelated with both unobserved factors and the other input variables of firm i.

Table 3: Production Function, First Stage Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Distance 2 Instrument	0.0043***	0.0044***			
	(0.0002)	(0.0002)			
Distance 3 Instrument		-0.0001*	0.0006***	0.0005***	0.0005***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
Private R&D	0.1598***	0.1774***	0.5022***	0.4612***	0.4357***
	(0.0517)	(0.0488)	(0.1463)	(0.1436)	(0.1504)
Capital	0.1827	0.1895	0.6169**	0.5903**	0.5923**
	(0.1133)	(0.1098)	(0.2622)	(0.2719)	(0.2746)
Labor	-0.0761	-0.0530	-0.6482**	-0.6021**	-0.6541**
	(0.0921)	(0.0913)	(0.2412)	(0.2636)	(0.2631)
Jaffe Tech. Proximity	1.8439***	1.8224***	3.3663**	3.2740**	3.3672**
	(0.4808)	(0.5123)	(1.5953)	(1.3777)	(1.4690)
Fixed Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities					
(Community $\times$ Year Effects)	0	0	0	10	20
F-statistic	255.17	219.47	24.92	32.06	19.18
No. of Observations	7607	7607	7607	7607	7607

Notes: The table reports OLS "first stage" regressions of the spillover variable $\sum_{j \neq i} g_{(ij)t} \log S_{jt}$ on selected instruments and all other right-hand side variables included in the regressions from Table 2. The sample is restricted to firms with at least one nonzero connection ($g_{(ij)t} \neq 0$) in any year t; all observations of these firms are also included for years with no connections. Columns 1 and 2 include, on the right hand side, the distance 2 instrument; columns 2 through 5 include the distance 3 instrument. All estimates include firm and year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with $\varphi = 0.8$ (10 communities) in column 4 and $\varphi = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

In Table 3 I report the results of various first stage regressions associated with model (12). All estimates are restricted to the subsample formed by those firms that ever enter the network. I regress the network-based spillovers variable on the aggregated log R&D stock of indirect connections located at either distance 2 (column 1), distance 2 *and* distance 3 (column 2), distance 3 only (column 3). The estimates from columns (4) and (5) are analogous to those in column (3), but they additionally include community-by-year fixed effects (respectively based on 10 and 20 communities, in analogy with Table 2). Noticeably, both instruments are strongly, positively correlated with the endogenous spillover variable. The measured F-statistics across all first stage estimates are reassuringly high: the lowest F-statistic is larger than 19. As expected, the correlation between the R&D of direct connections and the R&D of distance 2 indirect links appears much larger (by about one order of magnitude) than the analogous correlation with the R&D of distance 3 indirect connections.[29]

Table 4 displays the results from the 2SLS estimates which correspond, column-by-column, to the first stage regressions reported in Table 3. I focus the discussion on the parameter of interest δ. By instrumenting the spillover variable with the R&D of indirect connections located at distance 2 (column 1), δ is estimated at around 0.013, a figure slightly larger than the corresponding OLS estimate. When including both instruments (column 2) the result is again similar. By instrumenting only for the R&D of indirect connections located at distance 3 (column 3), the result is instead different: the point estimate of δ is substantially higher, hovering around 0.020. Interestingly, the inclusion of community-by-year effects results in even larger estimates: $\hat{\delta} \simeq 0.023$ with 10 communities (column 4) and $\hat{\delta} \simeq 0.021$ with 20 communities (column 5). All estimates of δ are statistically significant at the 5% level. The parameters associated with the other variables are estimated similarly with respect to the OLS case.

A possible concern about these results is that because the other input variables (capital, labor, private R&D) are also endogenous, the corresponding elasticities may be estimated incorrectly, thus affecting the IV estimates of δ. Recall that Proposition 3 from the analytical framework ensures that, if $D = C + 2L + 1$, the instrument is also uncorrelated with the set of estimated inputs, which would result in consistent estimates of δ. However, this may not hold in practice for all eligible distances D. To

[29]These first stage regressions are appropriate linear projections for the sake of 2SLS estimation, but do not provide consistent estimates of the patterns of spatial correlation of R&D. In fact, OLS is by construction an inconsistent estimator of any spatially autoregressive model. This may explain why the coefficient associated with the distance 3 instrument is estimated negative and statistically significant at the 10% level in column (2), while it is is not significantly different from zero in the corresponding reduced form regression reported in Appendix E. Moran's I statistic is the appropriate means for consistent estimation of spatial correlation patterns.

Table 4: Production Function, Two Stages Least Squares Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0560***	0.0562***	0.0510***	0.0489***	0.0464***
	(0.0117)	(0.0118)	(0.0111)	(0.0127)	(0.0131)
R&D Spillovers (δ)	0.0127***	0.0125***	0.0204**	0.0230**	0.0211**
	(0.0029)	(0.0030)	(0.0084)	(0.0088)	(0.0095)
Geographic Spillovers	0.0027	0.0027	0.0030	0.0026	0.0018
	(0.0019)	(0.0019)	(0.0018)	(0.0021)	(0.0018)
Capital	0.2025***	0.2027***	0.1969***	0.1944***	0.1956***
	(0.0200)	(0.0199)	(0.0225)	(0.0244)	(0.0234)
Labor	0.6642***	0.6640***	0.6692***	0.6677***	0.6685***
	(0.0357)	(0.0356)	(0.0373)	(0.0382)	(0.0392)
Jaffe Tech. Proximity	0.0314	0.0324	0.0041	-0.0364	-0.0167
	(0.0549)	(0.0553)	(0.0545)	(0.0669)	(0.0744)
Spillovers IV(s)	$D = 2$	$D = 2, 3$	$D = 3$	$D = 3$	$D = 3$
Fixed Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	0	0	0	10	20
No. of Observations	7607	7607	7607	7607	7607

Notes: The table reports IV-2SLS estimates of model (12). All estimates are restricted to firms with at least one nonzero connection ($g_{(ij)t} \neq 0$) in any year t; all observations of these firms are also included for years with no connections. Models in columns 1 and 2 employ the distance 2 instrument; models in columns 2 through 5 employ the distance 3 instrument. All estimates include firm and year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with $\varphi = 0.8$ (10 communities) in column 4 and $\varphi = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

overcome the problem, I assume that the residual shock υ_{it} presents an AR(1) time dependence structure with "innovation" ε_{it}:

$$\upsilon_{it} = \rho \upsilon_{i(t-1)} + \varepsilon_{it}, \quad \rho \in (0, 1) \tag{18}$$

and I estimate the model by performing System GMM on the ρ-differenced version of (12), following Blundell and Bond (1998, 2000).[30] Specifically, I treat conventional

[30]As noted by Wooldridge (2009) and discussed in more detail by Ackerberg, Caves, and Frazer (2015) the ρ-differenced "dynamic panel" approach for the estimation of production functions bears

inputs, internal R&D and external R&D as endogenous; I employ third and higher lags of both their levels and first differences as instruments (in the differences and level equations, respectively). Conversely, the other right-hand side spillover and demand side controls are either lagged quantities or factors outside the control of firms, thus I treat them as predetermined and independent of ε_{it}. Following a customary approach, I recover the structural parameters of interest from the steady-state representation of the ρ-differenced equation; these estimates are displayed in Table 5. In Appendix E I illustrate the estimation procedure in more detail; in addition I report and discuss the System GMM estimates of the ρ-differenced model.

Column (1) of Table 5 displays the baseline results from System GMM estimation of the model, obtained by instrumenting the spillover variable with appropriate "GMM style" lags. The elasticities of "private" inputs are estimated differently with respect to their OLS or 2SLS counterparts: as expected, the labor elasticity shrinks while the capital and private R&D elasticities are larger. By contrast, the estimate of δ (around 0.010) is not statistically significant from zero. Note that under the standard System GMM assumptions, δ is identified. Yet, the network's characteristics may exacerbate problems of weak instruments type that are typical of this procedure: since one firm's connections may be very different after three years, standard GMM instruments may have low predictive power. To address this I substitute the spillover instruments in the level equation with those implied by the following moments:

$$\mathbb{E}\left[\begin{pmatrix} \sum_{k=1}^{N} h_{(ik)t}^{D} \log S_{kt} \\ \sum_{k=1}^{N} h_{(ik)(t-1)}^{D} \log S_{k(t-1)} \end{pmatrix} \varepsilon_{it}\right] = 0 \tag{19}$$

which extend (15). These instruments have the effect of improving the statistical efficiency of the estimates. Column (2) reports the results from the inclusion of moments (19) with $D = 2$: $\hat{\delta}$ increases to 0.014, but it is still not statistically significant. Moving to $D = 3$ (column 3) results in $\hat{\delta} \simeq 0.017$, now statistically significant at the 10% level. With the inclusion of community-by-year effects (columns 4-5) δ is estimated at around 0.021 – about the same value as the corresponding 2SLS estimates – and is significant at the 5% level. Observe that the estimates of the capital, labor and private R&D elasticities from columns 4 and 5 take realistic and conventional values.

important analogies with semi-parametric control function methods *á la* Olley and Pakes (1996) and Levinsohn and Petrin (2003). The two approaches make different assumptions about the unobserved shock ω_{it}, but nevertheless result in analogous moment conditions.

Table 5: Production Function, System GMM Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.1234**	0.0786**	0.0959**	0.0880***	0.0668*
	(0.0470)	(0.0359)	(0.0367)	(0.0306)	(0.0366)
R&D Spillovers (δ)	0.0098	0.0142	0.0167*	0.0212**	0.0207**
	(0.0099)	(0.0092)	(0.0092)	(0.0088)	(0.0087)
Geographic Spillovers	0.0160	0.0189*	0.0067	0.0036	0.0063
	(0.0136)	(0.0104)	(0.0123)	(0.0124)	(0.0113)
Capital	0.2731***	0.2580**	0.2603**	0.2743***	0.2947**
	(0.0806)	(0.0961)	(0.0912)	(0.0936)	(0.1054)
Labor	0.5555***	0.6096***	0.6136***	0.5992***	0.6049***
	(0.0836)	(0.0990)	(0.0934)	(0.1036)	(0.1172)
Jaffe Tech. Proximity	-0.0555*	-0.0391	-0.0590**	-0.0783***	-0.0682**
	(0.0313)	(0.0277)	(0.0249)	(0.0247)	(0.0281)
Lags s of GMM-style IVs	$s \geq 3$	$s \geq 3$	$s \geq 3$	$s \geq 3$	$s \geq 3$
Diff. Eq. Spillover IVs	Standard	Standard	Standard	Standard	Standard
Level Eq. Spillover IVs	Standard	$D = 2$	$D = 3$	$D = 3$	$D = 3$
Time Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities					
(Community $\times$ Year Effects)	0	0	0	10	20
No. of Instruments	978	962	962	1142	1342
No. of Observations	7185	7185	7185	7185	7185

Notes: The table reports System GMM estimates of model (12). All estimates are restricted to firms with at least one nonzero connection ($g_{(ij)t} \neq 0$) in any year t; all observations of these firms are also included for years with no connections. The ρ-differenced production function is estimated by one step System GMM: lags s of the endogenous variables in levels are employed as instruments in the differenced equation; lags s of the same variables in first differences are used as instruments in the level equation. In columns (2) through (5) the standard GMM instruments of the spillover variables for the level equation are substituted by moments (19) for the given value of D. All estimates include year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with $\varphi = 0.8$ (10 communities) in column 4 and $\varphi = 0.6$ (20 communities) in column 5. Structural parameters are recovered from the steady-state representation of the ρ-differenced model; standard errors are calculated by the Delta Method (see Appendix E for details). Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

To summarize, both 2SLS and System GMM estimates of the production function obtained with distance 3 instruments point to a value of δ in a neighborhood of 0.021. For a connected pair of firms with a link of average strength, this implies a 0.0013%

increase in firm sales following a 1% increase in external R&D. For a connected firm with an average row sum of connections, the implied elasticity is about 0.0105 given a 1% increase in the R&D stock of all its linkages. This estimate of δ is about twice as large as the corresponding value obtained by performing simple OLS on the same subsample. The difference is only apparent when employing the distance 3 instrument in isolation, suggesting that the distance 2 instrument might be itself correlated with the unobserved characteristics of firms. In light of Proposition 2, this fact is consistent with the hypothesis that the spatial correlation of R&D is driven by exogenous factors rather than by the endogenous reflection of shocks: $(C, L) = (2, 0)$.

Several non-mutually exclusive hypotheses can be put forward to rationalize the finding that OLS estimates of δ are negatively biased. Perhaps, the simplest explanation is that there is measurement error in (external) R&D or in network connections. Instrumental variables may address this problem; however, the distance 2 instrument does not seem to do so in this context, as one would expect under "classical" measurement error. More structural explanations would point to correlated effects or network endogeneity. Common shocks can induce a negative bias of OLS if they extend to factors that negatively affect the variation of R&D. If, for example, *own* unobserved productivity υ_{it} strongly correlates with *external* R&D costs ϖ_{jt}, and firms' R&D is very responsive to costs, it could follow that $\mathbb{E}\left[\log S_{jt} \cdot \upsilon_{it}\right] < 0$ (see Appendix B for additional discussion). The negative bias can also be explained in terms of network endogeneity, if firms with relatively stronger connections are also relatively less productive. For example, smaller firms may substitute in-house R&D with learning from the R&D of larger firms, by attracting many inventors that are connected to them.

4.3 Market Value

The estimates of the market value model (16) are displayed in Table 6. In columns (1) and (2) I report OLS estimates, respectively conducted on the whole sample and on the network subsample. In columns (3), (4) and (5) I show instead the results of 2SLS estimates performed on the subsample, employing the distance 2 instrument, both instruments and just the distance 3 instrument, respectively. As in the case of the production function, restricting the sample to the network implies a reduction in the main spillover parameter $\tilde{\delta}$ (from 0.034 to 0.026). The IV strategy, however, results in a rebound of such estimate up to about 0.038. Unlike the production function case, however, this increase is uniform across all IV estimates. In fact, employing the distance 3 instrument in isolation results in a less precise, albeit similar, estimate of $\tilde{\delta}$ (with a 7.7% p-value). For a connected firm with average row-sum $\bar{g}_{it} = 0.5$, $\tilde{\delta} \simeq 0.038$ implies about a 0.017 elasticity of all its connections' R&D on its own Tobin's q.

Table 6: Market Value (Tobin's q), OLS and 2SLS Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Industry-level Sales	0.1743***	0.1966**	0.1944**	0.1943**	0.1946**
	(0.0515)	(0.0714)	(0.0700)	(0.0700)	(0.0703)
R&D Spillovers ($\tilde{\delta}$)	0.0338***	0.0264***	0.0386***	0.0387***	0.0374*
	(0.0074)	(0.0080)	(0.0100)	(0.0098)	(0.0200)
Geographic Spillovers	-0.0099	-0.0079	-0.0083	-0.0083	-0.0083
	(0.0065)	(0.0060)	(0.0061)	(0.0061)	(0.0058)
Jaffe Tech. Proximity	0.1088	-0.1164	-0.1696	-0.1698	-0.1643
	(0.1126)	(0.1640)	(0.1553)	(0.1555)	(0.1611)
BSV Business Stealing	-0.0344	0.1058	0.1020	0.1020	0.1024
	(0.0505)	(0.0963)	(0.0967)	(0.0967)	(0.0976)
S/K polynomial: F-statistic	24.19	25.21	26.31	26.35	24.09
(p-value)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Spillovers IV(s)	OLS	OLS	$D = 2$	$D = 2, 3$	$D = 3$
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	YES	YES	YES	YES
No. of Observations	12309	7481	7481	7481	7481

Notes: The table reports various estimates of model (16). All estimates from column 2 through 5 are restricted to firms with a nonzero connection ($g_{(ij)t} \neq 0$) in any year t; all observations of these firms are also included for years with no connections. Columns 1 and 2 report OLS estimates; columns 3, 4 and 5 report 2SLS estimates that employ different combinations of exogenous instruments, specifically: only the distance 2 instrument (3), both the distance 2 and 3 instruments (4), and only the distance 3 instrument (5). All estimates include firm and year fixed effects. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. For estimates not restricted to the network, firms outside the network constitute single clusters. Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

4.4 Patent Count

The results for the patent count model (17) are shown in Table 7, which is organized along the same lines of Table 6. Specifically, column (1) reports baseline estimates on the entire sample; column (2) those restricted to the network subsample, while the results from the control function approach, following the usual sequence of instrument combinations, are given in columns (3), (4) and (5). The main spillover parameter $\breve{\delta}$ is estimated at between 0.03 and 0.04 in columns (1) through (4). The estimates from column (5) however, obtained via a control function approach only based on the distance 3 instrument, register a much larger point estimate of $\breve{\delta} \simeq 0.114$. For a connected firm with an average row-sum of connections, this corresponds to a 0.057 elasticity of patent output following a 1% increase of all connections' R&D. By contrast, the elasticity of private R&D is typically estimated at between 0.070 and 0.075. Observe how, in these estimates, the geography-based measure of spillovers is both statistically and economically significant, while the Jaffe measure loses again all its economic and statistical significance once the analysis is restricted to the network.

Table 7: Cit. Weighted Patent Count, Maximum Likelihood Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Cit. Weighted Patents $(t-1)$	0.3850***	0.4127***	0.4082***	0.4100***	0.3936***
	(0.0177)	(0.0187)	(0.0182)	(0.0182)	(0.0190)
Private R&D ($\breve{\gamma}$)	0.0657*	0.0755**	0.0745**	0.0756**	0.0699**
	(0.0337)	(0.0330)	(0.0332)	(0.0333)	(0.0326)
R&D Spillovers ($\breve{\delta}$)	0.0321***	0.0305***	0.0394***	0.0362***	0.1141***
	(0.0068)	(0.0058)	(0.0085)	(0.0084)	(0.0190)
Geographic Spillovers	0.0390***	0.0443***	0.0433***	0.0441***	0.0244*
	(0.0110)	(0.0126)	(0.0124)	(0.0125)	(0.0126)
Jaffe Tech. Proximity	0.3269***	0.0662	0.0638	0.0643	0.0517
	(0.0467)	(0.0779)	(0.0774)	(0.0776)	(0.0768)
Industry Dummies	YES	YES	YES	YES	YES
Only Network	NO	YES	YES	YES	YES
Control Function IV(s)	None	None	$D = 2$	$D = 2, 3$	$D = 3$
No. of Observations	11866	6941	6941	6941	6941

Notes: The table reports maximum likelihood estimates of model (17). All estimates from column 2 through 5 are restricted to firms with at least one nonzero connection ($g_{(ij)t} \neq 0$) in any year t; all observations of these firms are also included for years with no connections. A sixth-degree polynomial of the predicted residuals from "control function" regressions is included in estimates reported in columns 3, 4 and 5. Specifically, $\sum_{j \neq i} g_{(ij)t} \log S_{jt}$ is regressed on the Z_{itq} controls, 4-digits industry effects, year fixed effects, and: the distance 2 instrument (column 3), both the distance 2 and 3 instruments (column 4), only the distance 3 instrument (column 5). All estimates include 4-digits industry dummies and year fixed effects. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in multiple years enter the same cluster. For estimates not restricted to the network, firms outside the network constitute single clusters. Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

4.5 Social Returns of R&D

To quantify the economic relevance of the estimated spillovers, it is useful to evaluate the Marginal Social Return (MSR) of R&D in its relationship with the corresponding Marginal Private Return (MPR). For any firm i, the MPR of R&D is defined as the marginal increase of output following a marginal increase of its *private* R&D stock. The MSR, on the other hand, extends this definition by considering how output varies at the margin along with the R&D of *all* other firms. Under the simplifying hypothesis of a homogeneous percentage increase of all firms' R&D stock ($dS_i/S_i = dS_j/S_j$ for all i and j), the MSR and MPR relative to firm i are easily obtained from (2):

$$\mathrm{MSR}_i = (\gamma + \delta \bar{g}_i) \frac{Y_i}{S_i} \geq \gamma \frac{Y_i}{S_i} = \mathrm{MPR}_i \tag{20}$$

this illustrates how, in general, the MSR exceeds the MPR by an amount that depends on the strength of connections. To calculate aggregate values for the MPR and MSR, I take the average of both expressions in (20) by pooling firms with nonzero connections ($\bar{g}_{it} > 0$) over all years, and using estimates $\hat{\gamma} \simeq 6.68\%$ and $\hat{\delta} \simeq 2.07\%$ from column (5) of Table 5. This exercise results in a network-wide MPR evaluated at around 102%, and in a corresponding aggregate MSR approximately equal to 114%: about 112% of the MPR.[31] These are realistic and economically significant values, comparable with evaluations from other studies (see Hall et al., 2010).[32]

5 Conclusion

In this paper I propose a new method for evaluating R&D spillovers. By aggregating information on patent collaboration relationships between individuals that work for different organizations, I construct a network of firms that are reciprocally connected through their respective R&D teams. I evaluate the dependence of firm productivity, market value and patent production rate from the R&D performed by firms connected

[31] Note that these are the calculated returns from the R&D *stock*. To estimate the returns from annual R&D *expenditures*, one should divide these figures by the steady-state flow/stock ratio. Under the typical assumption of a 0.20 steady-state ratio, one obtains approximately a 20% private return and a 23% social return from yearly R&D expenditures. Within the network subsample pooled over years, the average flow/stock ratio is 0.183 with 0.118 standard deviation.

[32] Note that excluding firms without connections from the calculation results in a higher MSR by construction, because those firms do not receive spillovers. However, connected firms are larger and more R&D intensive, and they represent a disproportionately larger share of the U.S. economy.

in the network, weighted by the intensity of mutual links. Concerned by the possibile presence of common confounders that simultaneously drive R&D choices, the intensity of cross-firm connections as well as firm-level outcomes, I employ a novel identification strategy based on the network topology. In particular, I instrument the R&D choices of one firm's direct connections with those of sufficiently distant indirect links. Under conditions specified by a formal model of firms' interaction, appropriately constructed instrumental variables predict the intensity of spillovers received by one firm, but are otherwise unrelated to its performance and innovation outcomes.

Estimates based on this definition of connections register sizable spillovers of connected firms' R&D on the productivity, market value, and patent output measures. These results, unlike those based on more traditional metrics of R&D spillovers, are robust to different specifications, and to the restriction of the sample to the largest and most R&D intensive firms. In conformity with the prediction of the theoretical model, the application of the identification strategy that I propose shows that, when instrumenting peers' R&D with the R&D of sufficiently distant indirect links, point estimates of spillover effects on both productivity and patent output increase substantially. This suggests that unobserved factors driving, on the one hand R&D and/or network connections, and firm outcomes on the other hand, might do so in opposite directions. I use the estimates of spillovers obtained from the proposed methodology to evaluate the relative importance of the marginal social returns to R&D relative to the private returns, finding that the former are about 112% of the latter among firms that are connected to the network.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. *Econometrica 83*(6), 2411–2451.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

Arellano, M. and S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economics Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Azoulay, P., J. S. Graff Zivin, and J. Wang (2010). Superstar Extinction. *The Quarterly Journal of Economics 125*(2), 549–589.

Bester, C. A., T. G. Conley, and C. B. Hansen (2011). Inference with dependent data using cluster covariance estimators. *Journal of Econometrics 165*(2), 137–151.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment 2008*(10), P10008.

Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica 81*(4), 1347–1393.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear Social Interactions Models. *Journal of Political Economy 123*(2), 444–496.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Blundell, R. and S. Bond (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews 19*(3), 321–340.

Blundell, R., R. Griffith, and J. Van Reenen (1995). Dynamic Count Data Models of Technological Innovation. *The Economic Journal 105*(429), 333–344.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Branstetter, L. G. and M. Sakakibara (2002). When Do Research Consortia Work Well and Why? Evidence from Japanese Panel Data. *American Economic Review 92*(1), 143–159.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer Effects and Social Networks in Education. *The Review of Economic Studies 76*(4), 1239–1267.

Christakis, N. A. and J. H. Fowler (2013). Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine 32*(4), 556–577.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Conley, T. G. and C. R. Udry (2010). Learning About a New Technology: Pineapple in Ghana. *The American Economic Review 100*(1), 35–69.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of Social Interactions through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics 2*(2), 241–275.

Elhorst, P. J. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer.

Graham, B. (2015). Methods of Identification in Social Networks. *Annual Review of Economics 7*, 465–85.

Graham, B. (2017). An Econometric Model of Network Formation with Degree Heterogeneity. *Econometrica*. Forthcoming.

Griffith, R., R. Harrison, and J. Van Reenen (2006). How Special Is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing. *American Economic Review 96*(5), 1859–1875.

Griliches, Z. (1964). Research Expenditures, Education, and the Aggregate Agricultural Production Function. *The American Economic Review 54*(6), 961–974.

Griliches, Z. (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics 10*(1), 92–116.

Griliches, Z. (1992). The Search for R&D Spillovers. *The Scandinavian Journal of Economics 94*, 29–43.

Griliches, Z. (1998). R&D and Productivity: The Unfinished Business. In *R&D and Productivity: The Econometric Evidence*. University of Chicago Press.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper No. 8498.

Hall, B. H., J. Mairesse, and P. Mohnen (2010). Measuring the Returns to R&D. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*. Elsevier.

Hausman, J. A., B. H. Hall, and Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica 52*(4), 909–938.

Hirsch, B. T. and T. G. Seaks (1993). Functional Form in Regression Models of Tobin's *q*. *The Review of Economics and Statistics 75*(2), 381–385.

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value. *American Economic Review 76*(5), 984–1001.

Jaffe, A. B. (1989). Characterizing the technological position of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy 18*(2), 87–97.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty (2000). Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *American Economic Review 90*(2), 215–218.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics 108*(3), 577–598.

Kelejian, H. H. and I. R. Prucha (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *The Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics 104*(2), 219–257.

König, M., X. Liu, and Y. Zenou (2014). R&D Networks: Theory, Empirics and Policy Implications. *CEPR Discussion Paper No. DP9872*.

Kranton, R., M. D'Amours, and Y. Bramoullé (2014). Strategic Interaction and Networks. *American Economic Review 104*(3), 898–930.

Levinsohn, J. and A. Petrin (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies 70*(2), 317–341.

Li, G.-C., R. Lai, A. D'Amour, D. M. Doolin, Y. Sun, V. I. Torvik, A. Z. Yu, and L. Fleming (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). *Research Policy 43*(6), 941–955.

Lychagin, S., J. Pinkse, M. E. Slade, and J. Van Reenen (2016). Spillovers in Space: does Geography Matter? *Journal of Industrial Economics 64*(2), 295–335.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies 60*(3), 531–542.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan. (First Edition).

Mas, A. and E. Moretti (2009). Peers at Work. *American Economic Review 99*(1), 112–145.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-level Production Functions. *American Economic Review 94*(1), 656–690.

Moretti, E. (2011). Local Labor Markets. *Handbook of Labor Economics 4*, 1237–1313.

Olley, S. G. and A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica 64*(6), 1263–1297.

Serafinelli, M. (2016). Good Firms, Worker Flows and Productivity. Mimeo.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature 49*(2), 326–365.

Thompson, P. and M. Fox-Kean (2005). Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *The American Economic Review 95*(1), 450–460.

Waldinger, F. (2011). Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *Review of Economics Studies 72*(2), 838–861.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economic Letters 104*(3), 112–114.

Appendix A Analytical Model: Proofs

Proof of Proposition 1. The problem of the individual firm can be expressed as

$$\max_{(S_i, X_{i1}, \dots, X_{iQ})} \left\{ \left(\prod_{q=1}^{Q} X_{iq}^{\beta_q} \right) S_i^{\gamma} \mathbb{E} \left[\left(\prod_{j=1}^{N} S_j^{g_{ij}} \right)^{\delta} \middle| \Omega_i \right] e^{\omega_i} - \sum_{q=1}^{Q} \xi_q X_{iq} - e^{\varpi_i} S_i \right\}$$

where the term in square brackets represents the uncertainty about the R&D investment choices of other firms. Firms respond in equilibrium to network externalities; as these depend on correlated shocks all firms make use of their available information in order to accurately predict them. Consider that the $Q+1$ First Order Conditions relative to S_i and $(X_{i1}, \dots, X_{iQ})$ are sufficient to characterize a maximum, since the problem is concave in all its choice variables. The FOCs are, respectively:

$$\frac{\partial \mathbb{E}\left[\pi_i(\cdot)\middle| \Omega_i\right]}{\partial S_i} = \gamma \left(\prod_{q=1}^{Q} X_{iq}^{\beta_q} \right) S_i^{\gamma-1} \mathbb{E} \left[\left(\prod_{j=1}^{N} S_j^{g_{ij}} \right)^{\delta} \middle| \Omega_i \right] e^{\omega_i} - e^{\varpi_i} = 0 \tag{A.1}$$

$$\frac{\partial \mathbb{E}\left[\pi_i(\cdot)\middle| \Omega_i\right]}{\partial X_{iq'}} = \beta_{q'} \left(\prod_{q=1}^{Q} X_{iq}^{\beta_q} \right) X_{iq'}^{-1} S_i^{\gamma} \mathbb{E} \left[\left(\prod_{j=1}^{N} S_j^{g_{ij}} \right)^{\delta} \middle| \Omega_i \right] e^{\omega_i} - \xi_{q'} = 0 \tag{A.2}$$

with (A.2) taken for $q' = 1, \dots, Q$. Combining (A.1) with each of the Q conditions expressed in (A.2), one gets

$$X_{iq} = \frac{\beta_q}{\gamma \xi_q} e^{\varpi_i} S_i \tag{A.3}$$

for $q = 1, \dots, Q$. These relationships state that the vector of equilibrium input choices is uniquely determined for every firm given their optimal R&D decisions and ϖ_i, thus motivating (5). Intuitively, R&D is a sufficient statistic of equilibrium externalities (actually each of the $Q+1$ choice variables can be considered as such, but singling out S_i is more convenient). Therefore, in order to demonstrate existence and uniqueness of the Bayes-Nash equilibrium under the conditions stated in the text, it is sufficient to show the existence of a fixed point of the R&D equilibrium choices S_i.

To this end, substitute the Q relationships in (A.3) into (A.1), obtaining:

$$S_i = \left(\mathbb{E} \left[\prod_{j=1}^{N} S_j^{\delta g_{ij}} \middle| \Omega_i \right] e^{\mu + \omega_i - \left(1 - \sum_{q=1}^{Q} \beta_q\right) \varpi_i} \right)^{\frac{\vartheta}{\delta}} \tag{A.4}$$

where $\mu \equiv \log \gamma + \sum_{q=1}^{Q} \beta_q (\log \beta_q - \log \xi_q - \log \gamma)$. Note that expression (A.4) is a mapping from the space $\boldsymbol{\Omega}$ of all information sets that can be available to players and the set of other players' strategies S_{-i} onto the set of positive real numbers, which I denote as $S_i^* : \boldsymbol{\Omega} \times \mathrm{R}_{++}^{N-1} \rightarrow \mathbb{R}_{++}$. By definition, a fixed point of the vector-valued function $\mathrm{S}^* = (S_1^*, \ldots, S_N^*)$ is a Bayes-Nash equilibrium of the game. Clearly, there is a one-to-one relationship between $\mathrm{S}^* = (S_1^*, \ldots, S_N^*)$ and the associated logarithmic function $\log \mathrm{S}^* = (\log S_1^*, \ldots, \log S_N^*)$. It turns out that it is more convenient to show existence and uniqueness of the equilibrium in its logarithmic form.

Denote the space spanned by $\log \mathrm{S}^* = (\log S_1^*, \ldots, \log S_N^*)$ as $\mathcal{S}$, and endow it of the max-norm $\|\log \mathrm{S}^*\| = \max_i \|\log S_i^*\|_\infty$. Define the operator $H : \mathcal{S} \to \mathcal{S}$ as:

$$H_i\left(\log S_1^*, \ldots, \log S_N^*\right) = \frac{\vartheta}{\delta}\left\{\mu + \log \mathbb{E}\left[\prod_{j=1}^{N}\left(S_j^*\right)^{\delta g_{ij}}\middle|\,\Omega_i\right] + \omega_i - \left(1 - \sum_{k=1}^{K}\beta_k\right)\varpi_i\right\}$$

for $i = 1, \ldots, N$; this is well-defined as there is a one-to-one relationship between S_i^* and its logarithm. The operator is based on the "manipulated" First Order Conditions of the restricted game (A.4), hence it is consistent with expected utility maximization. Consider that, for any two $(\log \mathrm{S}^*, \log \mathrm{Z}^*) \in \mathcal{S}^2$, it is:

$$\begin{aligned}
\|H(\log \mathrm{S}^*), H(\log \mathrm{Z}^*)\| &= \max_i \left|H_i(\log \mathrm{S}^*) - H_i(\log \mathrm{Z}^*)\right| \\
&= \max_i \frac{\vartheta}{\delta}\left|\log \mathbb{E}\left[\prod_{j=1}^{N}\left(S_j^*\right)^{\delta g_{ij}}\middle|\,\Omega_i\right] - \log \mathbb{E}\left[\prod_{j=1}^{N}\left(Z_j^*\right)^{\delta g_{ij}}\middle|\,\Omega_i\right]\right| \\
&\leq \max_i \frac{\vartheta}{\delta}\left|\log \mathbb{E}\left[\left(S_i^*\right)^{\delta(\max_i \bar{g}_i)}\middle|\,\Omega_i\right] - \log \mathbb{E}\left[\left(Z_i^*\right)^{\delta(\max_i \bar{g}_i)}\middle|\,\Omega_i\right]\right| \\
&\leq \max_i \frac{\vartheta}{\delta}\left|\log \mathbb{E}\left[S_i^* \middle|\, \Omega_i\right]^{\delta(\max_i \bar{g}_i)} - \log \mathbb{E}\left[Z_i^* \middle|\, \Omega_i\right]^{\delta(\max_i \bar{g}_i)}\right| \\
&= \vartheta\left(\max_i \bar{g}_i\right)\max_i \left|\log \mathbb{E}\left[S_i^* \middle|\, \Omega_i\right] - \log \mathbb{E}\left[Z_i^* \middle|\, \Omega_i\right]\right| \\
&= \left\|\vartheta \max_i \bar{g}_i\right\| \left\|\log S_i^* - \log Z_i^*\right\|
\end{aligned}$$

where $\bar{g}_i = \sum_{j=1}^{N} g_{ij}$; the fourth line (second inequality) follows from the observation that, because $\delta\bar{g}_i < 1$ for all $i = 1, \ldots, N$, by Jensen's inequality:

$$\mathbb{E}\left[S_i \middle|\, \Omega_i\right]^{\delta(\max_i \bar{g}_i)} - \mathbb{E}\left[S_i^{\delta(\max_i \bar{g}_i)}\middle|\,\Omega_i\right] \geq 0$$

a difference which is monotonically increasing in $S_i > 0$ for all $\Omega_i \in \mathbf{\Omega}$. Consequently, H is a contraction with Lipschitz constant

$$\vartheta \max_i \bar{g}_i = \frac{\delta}{1 - \gamma - \sum_{q=1}^{Q}\beta_q}\max_i\left(\sum_{j=1}^{N} g_{ij}\right) < 1$$

which is smaller than 1 under the conditions stated in the text. In such a circumstance, by the Contraction Mapping Theorem both $\log \mathrm{S}^*$ and S^* have a fixed point, implying that the game has a unique Bayes-Nash equilibrium.

It still needs to be demonstrated that equilibrium R&D S_i^* is expressible as (4). To this end, rewrite that equation as $\log S_i^* = \mu\vartheta\delta^{-1} b_i + s_i^*(\Omega_i, \mathcal{G})$ for some generic $b_i > 0$, and substitute it into (A.4) for all $j \neq i$, thus obtaining:

$$S_i^* = \left(\left[\prod_{j=1}^{N} \exp\left(\mu\vartheta g_{ij} b_j\right) \right] \mathbb{E} \left[\left(\prod_{j=1}^{N} \exp\left(g_{ij}\delta \cdot s_j^*(\Omega_j, \mathcal{G}) \right) \right) \middle| \Omega_i \right] \exp\left(\mu + \tilde{\omega}_i\right) \right)^{\frac{\vartheta}{\delta}}$$

taking logarithms and rearranging terms this becomes

$$\log S_i^* = \mu\vartheta\delta^{-1} \left(1 + \vartheta \sum_{j=1}^{N} g_{ij} b_j \right) + s_i^*(\Omega_i, \mathcal{G}) \tag{A.5}$$

it easy to see that this expression conforms to the definition of the contraction operator H, and that $s_i^*(\Omega_i, \mathcal{G})$ has the form given in the text. For (A.5) to be consistent also with (4), it must be shown that

$$1 + \vartheta \sum_{j=1}^{N} g_{ij} b_j = b_i = b_i^*(\mathcal{G}; \vartheta)$$

for every firm $i = 1, \dots, N$. Rewrite the first equality above in matrix form:

$$\iota + \vartheta \mathbf{G}\mathbf{b} = \mathbf{b}$$

where $\mathbf{G}$ is the *adjacency matrix* with g_{ij} entries and $\mathbf{b} = (b_1, \dots, b_N)^{\mathrm{T}}$. Since matrix $(\mathbf{I} - \vartheta\mathbf{G})$ is invertible almost surely, a solution for $\mathbf{b}$ exists almost always and reads as:

$$\mathbf{b} = (\mathbf{I} - \vartheta\mathbf{G})^{-1} \iota$$

where:

$$(\mathbf{I} - \vartheta\mathbf{G})^{-1} = \sum_{\ell=1}^{\infty} \vartheta^{\ell} \mathbf{G}^{\ell}$$

the series converges under the conditions stated in the text. The solution is exactly the vector of Katz-Bonacich centrality measures with attenuation parameter ϑ.

Proof of Proposition 2. The proof is constructive, and it is intuitive given basic concepts of graph theory. For any pair of firms i and j such that $d_{ij} = D > C + L$, take any of their shortest paths of length D. Then, order the intermediate connections along the chosen path: $\ell = 0, \dots, D$ where (without loss of generality) $i = 0$ and $j = D$. By Assumption 2 and the definition of path in a network, $\{\omega_\ell, \varpi_\ell\} \notin \Omega_j$ if $\ell < L$. Thus, the shortest path connecting $\{\omega_i, \varpi_i\}$ with any element of set Ω_j has length $D - L$. Since $D - L > C$, $\{\omega_i, \varpi_i\}$ and all the elements of Ω_j are orthogonal by

Assumption 1, implying $\mathbb{C}\mathrm{ov}\left(\omega_i, \log S_j^*\right) = 0$ because of equation (4). If this is true for the shortest path connecting i and j, it must be so for any other path, thereby establishing (6). By analogous reasoning, suppose that $d_{ij} = D > C + 2L$, and take the shortest path between i and j as defined earlier. In addition to the considerations above, $\{\omega_\ell, \varpi_\ell\} \notin \Omega_i$ if $\ell > L$, hence the shortest path connecting any element of set Ω_i with any element of set Ω_j has length $D - 2L > C$. Consequently, the optimal R&D choices of firms i and j are in equilibrium functions of mutually independent sets of random variables, which implies (7). Finally, the same result cannot be established for firms that are distant $C + 2L$ or less, which is reflected in (8).

Proof of Proposition 3. Recall from the First Order Conditions that in equilibrium, $\log X_{iq}^* = \log S_i^* - \log \beta_q - \log \gamma - \log \xi_q + \varpi_i$ for $q = 1, \ldots, Q$, which can be rewritten as:

$$\log X_{iq}^* = x_i^*(\mathcal{G}) + s_i^*(\Omega_i; \mathcal{G}) - \varpi_i$$

for some firm-specific function of the network topology $x_i^*(\mathcal{G})$. The stochastic properties of equilibrium inputs are driven by the term $s_i^*(\Omega_i; \mathcal{G}) - \varpi_i$; but since ϖ_i is always listed in Ω_i by Assumption 2, $\log X_{iq}^*$ must be orthogonal to any combination of random variables that is also orthogonal to $\log S_i^*$. Hence, an analysis similar to the one made above would demonstrate that $\log X_{iq}^*$ and $\log S_j^*$ are independent for all Q conventional inputs as long as $d_{ij} = D > C + 2L$, which proves (9).

Appendix B Additional Discussion of the Model

In this appendix I extend the analytical framework of the paper and analyze in more depth some of its properties. This appendix is divided in three independent parts. In the first part I explore models of network formation between multi-technology firms, showing that their predictions are largely consistent with Assumption 1. In the second part I illustrate how the model may deliver equilibria that are empirically associated to a negative bias of OLS estimates of δ. In the third part I sketch a dynamic version of the model in which firms accumulare knowledge capital over time, examining under what circumstances the main result of Proposition 2 is still valid.

B.1 Network Formation

In what follows I illustrate models of network formation between firms operating in different technological areas. These models are at least in part stochastic – reflecting the fact that connections between firms may be driven by factors outside the control of firm managers, such as freely moving inventors. In some of these models I allow for connections to be caused by both random factors and firms' decisions reflecting cost-benefit evaluations. As my objective is to illustrate how network formation relates to the cross-correlation of firms' productivity shocks conditional on network distance, it is appropriate to introduce more primitive concepts that drive both firms' similarities and their association in the network. To this end, I provide a micro-foundation of the firm-specific productivity shock ω_i in terms of "fundamental technologies."

Let $\mathcal{U}$ be a set of U "fundamental technologies" that firms are capable to recombine for production purposes. Index such technologies as $u = 1, \ldots, U$. Each of the N firms is exogenously endowed of knowledge about some of the elements in $\mathcal{U}$. Specifically, let $l_{iu} = 1$ denote that firm i knows about fundamental technology u, otherwise $l_{iu} = 0$ (with $\sum_{u=1}^{U} l_{iu} > 0$ for every firm i). Then one can think of the firm-level productivity shock ω_i as:

$$\omega_i = \sum_{u=1}^{U} l_{iu}\nu_u \tag{B.1}$$

where $(\nu_1, \ldots, \nu_U)$ is a set of "fundamental technology shocks," each of them drawn from some unspecified distribution but all of them independent from one another, that represent how much a specific fundamental technology translates into productivity advantages. This amounts to characterize ω_i as a "spatial moving average" process in the space of fundamental technologies. Therefore, the cross-correlation of ω_i between any two firms depends on the overlap over their knowledge of different fundamental technologies. Note that I could as well have specified (B.1) as a non-linear function of the shocks ν_u, possibly allowing for interaction terms. The linearity assumption is made for convenience and is of little consequence. The crucial assumption is that the "fundamental" technologies are independent of one another (hence their name).

I am now ready to illustrate alternative models of network formation, which are discussed in sequence. In all models, connections in the network depend on the extent to which two firms overlap in the space of fundamental technologies. For simplicity I abstract from cost shocks ϖ_i. However, the discussion about the cross-correlation of ω_i in the network may be easily extended to that of ϖ_i, so long as the latter is made dependent on fundamental technologies as well.

Pure Technological Association. In the simplest model of network formation that I present here, cross-firm connections are determined by the following rule:

$$g_{ij} = \frac{\sum_{u=1}^{U} l_{iu} l_{ju} \nu_u^2}{\sum_{u=1}^{U} \left(l_{iu} + l_{ju}\right) \nu_u^2} \equiv \widehat{g}_{ij} \in [0, 1] \tag{B.2}$$

that is, cross-firm connections are equal to the ratio between the sum of the squared "fundamental" shocks ν_u that are common to both firms, and the sum of all squared shocks pertaining to at least one of the two firms. I call this ratio "similarity factor" and I denote it as $\widehat{g}_{ij}$. Note the analogy between $\widehat{g}_{ij}$ and the measure of connection (10) introduced in this article. Rule (B.2) has the following interpretation: bilateral connections between firms solely depend on the similarity of their technologies, and so of their inventors' specializations. It does not allow for other determinants.

The model can be enriched, for example, by making g_{ij} a function of $\widehat{g}_{ij}$ (say, its square root), or by introducing some random component (to the extent that $g_{ij} \neq 0$ if $\widehat{g}_{ij} \neq 0$). However, any of these variations would not affect the main property of this model: by rule (B.2), Assumption 1 would be satisfied for $C = 1$, as firms with similar technologies (that is, $\mathbb{C}\text{ov}\left(\omega_i, \omega_j\right) \neq 0$) would always be connected, while firms with unrelated technologies ($\mathbb{C}\text{ov}\left(\omega_i, \omega_j\right) = 0$) would never be.

Random Technological Association. Consider instead the case of an unweighted network where $g_{ij} \in \{0, 1\}$, and the following rule that determines whether any two firms are connected:

$$g_{ij} = \mathbb{1}\left[\widehat{g}_{ij} - \eta_{ij} \geq 0\right] \tag{B.3}$$

where $\eta_{ij} \in [0, 1]$ is some random shock specific to the (i, j) pair. This rule can be interpreted as follows: two firms are connected if their similarity $\widehat{g}_{ij}$ is larger than some "cost" η_{ij} associated with the connection. Such a cost, in turn, may be interpreted in multiple ways. For example, it is the cost of letting information "leak" if g_{ij} represents spontaneous inventor interactions; in the case of formal R&D joint ventures instead, it may represent the cost of managing the relationship between the two firms. This model is similar to the one by Graham (2017), but with some differences. Graham's model explicitly distinguishes between firm-specific shocks, pair-specific shocks and firm characteristics; here the similarity factor $\widehat{g}_{ij}$ accounts for all possible "advantages" from the connection, while other firms characteristics, such as input factors, are not to be included as they are themselves endogenous to the model's primitives.

This model may be crudely extended to the case of a weighted network as follows:

$$g_{ij} = \mathbb{1}\left[\widehat{g}_{ij} - \eta_{ij} \geq 0\right] \cdot \widehat{g}_{ij} \tag{B.4}$$

that is, connections are equal to the similarity factor so long as the "cost" η_{ij} is low enough (one can think to make the intensity of the connection g_{ij} a function of η_{ij}, or to include an additional random factor that affects, but this is beyond the point of this analysis). Note that, in general, Assumption 1 will not be exactly met in such a model: however unlikely, it is very well possible that two firms with nonzero similarity end up very far apart from one another. The "however unlikely" part of this sentence, however, should be adequately stressed. In fact, the probability that two firms i and j with nonzero similarity, and so with $\mathbb{C}\text{ov}\left(\omega_i, \omega_j\right) \neq 0$, end up distant more than 2 degrees away from one another, is typically very small.

To see this, suppose that the shocks η_{ij} are all independently drawn from some distribution $B\left(\eta\right)$: the probability that any firm k links up to another firm ℓ is given by $B\left(\widehat{g}_{k\ell}\right)$. Thus, the probability that firms i and j end up at distance higher than 2 is:

$$\mathbb{P}\text{r}\left(d_{ij} > 2 | \, \widehat{g}_{ij} \neq 0\right) = \left[1 - B\left(\widehat{g}_{ij}\right)\right] \prod_{k \in \mathcal{K}_{ij}} \left[1 - B\left(\widehat{g}_{ik}\right) B\left(\widehat{g}_{jk}\right)\right] \tag{B.5}$$

where $\mathcal{K}_{ij} \subset \mathcal{I}$ is the set of all firms with nonzero similarity with both firms i and j ($\mathcal{K}_{ij} = \left\{k \in \mathcal{I} : \widehat{g}_{ik} \neq 0 \wedge \widehat{g}_{jk} \neq 0\right\}$). It is easy to establish regularity conditions under which as the set of firms expands (N grows larger), and $\mathcal{K}_{ij}$ does too for all pairs (for this to be true it is enough that "new" firms are associated with a random subset of $\mathcal{U}$) the probability in (B.5) goes to zero for every appropriate pair (i, j). By analogous reasoning, it can be argued that the probability of any two similar firms ending up more than three degrees of separation apart is even smaller. Thus, Assumption 1 can be thought to "asymptotically" hold for $C = 2$ (or, to play it safe, for $C = 3$).

Structural Transitivity. Suppose that rule (B.4) is augmented as follows:

$$g_{ij} = \mathbb{1}\left[\widehat{g}_{ij} + \kappa \sum_{k=1}^{N} g_{ik} g_{jk} - \eta_{ij} \geq 0\right] \cdot \widehat{g}_{ij} \tag{B.6}$$

where parameter κ corresponds to the presence of "preference for triadic closure" or "structural transitivity" in models of network formation: the hypothesis that any two agents are more likely to form mutual connections if they are both connected to some third party. In this article's context, $\kappa > 0$ can be interpreted as the possibility that inventors from two different firms are more likely to interact if both groups do so with inventors of other firms; this can be motivated on lower costs of information leakage or some underlying social dynamics. Models of network formation that include structural transitivity are known to feature multiple equilibria; the econometric identification of the preference for triadic closure is an area of active research (Graham, 2015).

The presence of structural transitivity in network formation would clearly reinforce any mechanism of assortativity on unobserved characteristics, such as "fundamental technologies." To see how this relates to the identification strategy proposed in this paper, consider the previous discussion about how Assumption 1 would hold asymptotically under random technological association. If connections are formed in equilibrium according to rule (B.6) with $\kappa > 0$, then as both $\mathcal{I}$ and $\mathcal{K}_{ij}$ expand not only the conditional probability that $d_{ij} = 2$ for any two firms i and j would approach one, but also the unconditional probability that $d_{ij} = 1$ $(g_{ij} \neq 0)$ increases. By analogous reasoning, in presence of structural transitivity the probability that at least two "similar" connections of firms i and j are connected increases faster as the network grows larger, implying larger terms of the type $B\left(\widehat{g}_{ik}\right) B\left(\widehat{g}_{jk}\right)$ and thus even faster convergence of the probability in (B.5) to zero. Thus, structural transitivity does not challenge the identification strategy proposed in this article; in fact, if the tendency to triadic closure is a feature of network formation independent of firm unobservables, the credibility of the empirical strategy would be strengthened.

Forward-looking Network Formation. It is difficult to modify the game presented in Section 1 to allow for "strategic" (instead of random) network formation in the first stage. The reason is that while input choices are taken by individual firms, connections are bilateral: the initial stage has to be modeled as a set of $\frac{1}{2}N\left(N-1\right)$ decisions about pair-specific connections, with possible interdependencies and multiple equilibria. It is nevertheless possible to show that such an exercise would not affect the conceptual framework in a sensible way. To appreciate this, suppose that after nature has drawn the technology shocks and firms' unobservables, any two firms i and j cooperatively form a weighted link g_{ij}, by applying backward induction, as follows:

$$g_{ij} = \underset{g_{ij}\in[0,1]}{\arg\max} \left\{ \pi_i \left[\left(\mathrm{S}^*, \mathrm{X}^*\right), g_{ij}, \mathcal{G}^*_{-ij} \right] + \pi_j \left[\left(\mathrm{S}^*, \mathrm{X}^*\right), g_{ij}, \mathcal{G}^*_{-ij} \right] - r\left(g_{ij}, \widehat{g}_{ij}\right) \right\} \qquad \text{(B.7)}$$

where $\pi_i\left(\cdot\right)$ is the profit function of firm i (and symmetrically for j), $\left(\mathrm{S}^*, \mathrm{X}^*\right)$ is the vector of Bayes-Nash equilibrium input choices expressed in Proposition 1, $\mathcal{G}^*_{-ij}$ is the equilibrium set of connection decisions taken by all other firm pairs, and $r\left(g_{ij}, \widehat{g}_{ij}\right)$ is a "link cost function" that is *increasing* in the intensity of the connection g_{ij} (the stronger the link, the higher the cost to maintain it), and *decreasing* in the similarity factor $\widehat{g}_{ij}$ (the more similar are two firms' technologies, the easier for them to link up). This cost function is assumed dependent on firms' similarity factor for simplicity, so to avoid modifying the benefit part of the trade-off (joint firm profits do not depend on how "similar" ω_i and ω_j are, as per the setup of the model in Section 1).

Denote the "combined profits in absence of a connection" as follows:

$$\Pi^*_{ij} \equiv \pi_i \left[\left(\mathrm{S}^*, \mathrm{X}^*\right), g_{ij} = 0, \mathcal{G}^*_{-ij} \right] + \pi_j \left[\left(\mathrm{S}^*, \mathrm{X}^*\right), g_{ij} = 0, \mathcal{G}^*_{-ij} \right]$$

(note that Π^*_{ij} is a function of ω_i and ω_j, so it is a random variable). Clearly, firms i

and j are jointly better off by being connected ($g_{ij} > 0$) so long as:

$$\Pi_{ij}^{*} - r\left(g_{ij} = 0, \widehat{g}_{ij}\right) \geq 0$$

or, by further assuming that $r(\cdot)$ is locally invertible at $(0, \widehat{g}_{ij})$ with respect to $\widehat{g}_{ij}$:

$$\widehat{g}_{ij} - r^{-1}\left(g_{ij} = 0, \Pi_{ij}^{*}\right) \geq 0 \tag{B.8}$$

where the expression $r^{-1}\left(g_{ij} = 0, \Pi_{ij}^{*}\right)$ is a decreasing function of Π_{ij}^{*}, but it does not depend on how similar ω_i and ω_j are (if the two shocks are correlated but take small values, combined profits decrease; vice versa two unrelated firms may both receive high draws of ω_i). Note the equivalence between (B.8) and (B.3), and by extension the analogy with (B.4): cross-firm connections appear in those firm pairs for which the similarity factor $\widehat{g}_{ij}$ is larger than the realization of some random variable which is independent of $\widehat{g}_{ij}$ itself. Therefore, the previous discussion on how Assumption 1 holds asymptotically can also be extended to a setup of forward-looking cooperative network formation. Because optimal input choices in the ensuing subgame would still be expressed as the result of Proposition 1, Propositions 2 and 3 would still be valid (albeit asymptotically) implying little threat for the proposed empirical strategy.

B.2 Negative Bias Prediction

A possible explanation of the negative bias associated with OLS of δ is that this is in fact a structural feature of the data generation process. For this to be consistent with the proposed conceptual framework, it is necessary that the model can give rise to equilibria in which, for two connected firms i and j ($g_{ij} \neq 0$), the following three relationships hold simultaneously in equilibrium.

$$\begin{aligned} \mathbb{C}\mathrm{ov}\left(\log S_i^{*}, \log S_j^{*} \middle| d_{ij} = 1\right) &> 0 \\ \mathbb{C}\mathrm{ov}\left(\omega_i, \log S_j^{*} \middle| d_{ij} = 1\right) &< 0 \\ \mathbb{C}\mathrm{ov}\left(\omega_j, \log S_i^{*} \middle| d_{ij} = 1\right) &< 0 \end{aligned} \tag{B.9}$$

This apparently counterintuitive circumstance cannot be rationalized by standard supply-side models of production featuring spillovers *and* strategic complementarities, short of introducing some spatial dependence across firm costs. That is, it may arise if the spatial cross-correlation of the cost factor ϖ drives that of R&D disproportionately more than how the cross-correlation of productivity shocks ω does. In fact, since R&D costs affect R&D investment negatively, it might be that the covariance between one firm's (i) productivity residual and the R&D of some its connections (j) is dominated by a negative term which is driven directly by the moment $\mathbb{C}\mathrm{ov}(\omega_j, \varpi_j)$, and indirectly by $\mathbb{C}\mathrm{ov}(\omega_i, \omega_j)$. Intuitively, if there are common shocks in ω_i and ϖ_i, if cost factors co-evolve together with firms' technologies, and finally if R&D expenditures are quite (negatively) dependent on R&D costs ϖ_i, then ω_i and $\log S_j$ are negatively correlated.

A Parametric Example. The set of inequalities in (B.9) are expressed with respect to a specific pair of firms i and j; for this prediction to be the source of the negative OLS bias of δ, it must be more general. In what follows, I outline a specific structure of the model such that the relationships in (B.9) hold simultaneously for all potentially connected pairs in the network. Specifically, consider the case where the joint distribution of $(\omega_i, \omega_j, \varpi_i, \varpi_j)$ is such that:

$$\mathbb{C}\text{orr}\left(\omega_i, \omega_j\right) = \mathbb{C}\text{orr}\left(\varpi_i, \varpi_j\right) = g_{ij} \in [0, 1)$$

for all pairs of firms (i, j), $i \neq j$, $i, j = 1, \ldots, N$. Furthermore, suppose that firms only observe their private shocks: $\Omega_i = \{\omega_i, \varpi_i\}$ for $i = 1, \ldots, N$; and that $\mathcal{F}\left(\boldsymbol{\omega}, \boldsymbol{\varpi} | \mathcal{G}\right)$ is a multivariate normal distribution. In this specific case it particularly easy to see (but the result is more general) that equilibrium R&D is a linear function of private shocks: $s_i^*\left(\Omega_i, \mathcal{G}\right) = a_i^0 + a_i^1\left(\omega_i - \varkappa\varpi_i\right)$ for $a_i^0, a_i^1 > 0$ and $\varkappa \equiv 1 - \sum_{q=1}^{Q} \beta_q \in (0, 1)$. Denote $\varsigma_\ell \equiv \mathbb{V}\text{ar}\left(\omega_\ell\right)$ and $\varrho_\ell \equiv \mathbb{C}\text{ov}\left(\omega_\ell, \varpi_\ell\right)$ for $\ell = i, j$, and assume further that

$$\begin{aligned}
\mathbb{V}\text{ar}\left(\varpi_i\right) &= \varrho_i^2 \varsigma_i \\
\mathbb{V}\text{ar}\left(\varpi_j\right) &= \varrho_j^2 \varsigma_j \\
\mathbb{C}\text{orr}\left(\omega_i, \varpi_j\right) &= g_{ij} \\
\mathbb{C}\text{orr}\left(\omega_j, \varpi_i\right) &= g_{ij}
\end{aligned}$$

so that the following variance-covariance matrix

$$\mathbb{V}\text{ar}\begin{pmatrix} \omega_i \\ \omega_j \\ \varpi_i \\ \varpi_j \end{pmatrix} = \begin{pmatrix} \varsigma_i & \cdots & \cdots & \cdots \\ g_{ij}\sqrt{\varsigma_i\varsigma_j} & \varsigma_j & \cdots & \cdots \\ \varrho_i & g_{ij}\varrho_i\sqrt{\varsigma_i\varsigma_j} & \varrho_i^2\varsigma_i & \cdots \\ g_{ij}\varrho_j\sqrt{\varsigma_i\varsigma_j} & \varrho_j & g_{ij}\varrho_i\varrho_j\sqrt{\varsigma_i\varsigma_j} & \varrho_j^2\varsigma_j \end{pmatrix}$$

is a positive semidefinite matrix for $\varsigma_\ell \geq 1$ and $0 \leq g_{ij} \leq \frac{1}{4}\left(1 + \varsigma_\ell^{-1}\right)^{\frac{1}{2}}$ with $\ell = i, j$ (note that these are realistic values of g_{ij} given the descriptives in Figure 2), which makes it a legitimate characterization of the random vector $(\omega_i, \omega_j, \varpi_i, \varpi_j)$. Under all these hypotheses, the inequalities in (B.9) can be expressed after some calculation as the following expressions:

$$\begin{aligned}
\mathbb{C}\text{ov}\left(a_i^1\left(\omega_j - \varkappa\varpi_j\right), a_j^1\left(\omega_j - \varkappa\varpi_j\right)\right) &\propto \left(\varkappa\varrho_i - 1\right)\left(\varkappa\varrho_j - 1\right) > 0 \\
\mathbb{C}\text{ov}\left(\omega_i, a_j^1\left(\omega_j - \varkappa\varpi_j\right)\right) &\propto 1 - \varkappa\varrho_j < 0 \\
\mathbb{C}\text{ov}\left(\omega_j, a_i^1\left(\omega_i - \varkappa\varpi_i\right)\right) &\propto 1 - \varkappa\varrho_i < 0
\end{aligned}$$

which all hold simultaneously as long as $\varrho_i, \varrho_j > \varkappa^{-1}$. While this example is stylized it illustrates quite well that, for (B.9) to hold, both $\mathbb{C}\text{ov}\left(\omega_i, \varpi_i\right)$ and $\mathbb{C}\text{ov}\left(\omega_j, \varpi_j\right)$ must be large enough, so that cost shocks predominate the variation of external R&D.

B.3 Dynamic Model (Sketched)

Here I sketch a two-period version of the model in which firms accumulate R&D stocks by making yearly investments in R&D (flows). The purpose is to illustrate under what circumstances the main result of Proposition 2 also applies in the more general case of a dynamic model. For simplicity, in what follows I omit conventional inputs. I discuss two possible scenarios corresponding to different economic assumptions. In the first of the two scenarios, firms commit in advance to a future sequence of R&D investment (flows). In the second scenario, firms are able to revise their R&D investment choices in every period. I analyze the two scenarios in sequence.

Pre-commitment

Firms might have compelling reasons to commit to a long-term plan of R&D investment. One reason might be financial: say, for example, that venture capital support is conditional on long-term projects. More structural reasons are probably related to the actual nature of R&D activity: highly risky, characterized by large fixed costs and requiring many years to yield (potentially high) rewards. Thus, it might be optimal for firms to commit in advance to long-term R&D plans.

Under commitment, the firm's objective function reads as:

$$
\begin{aligned}
\pi_i\left(R_{i1}, R_{i2}; \dots\right) = R_{i1}^{\gamma}\mathbb{E}&\left[\left(\prod_{j=1}^{N} R_{j1}^{g(ij)1}\right)^{\delta}\middle|\,\Omega_{i1}\right] e^{\omega_{i1}} \\
&+ \psi\left(\zeta R_{i1} + R_{i2}\right)^{\gamma}\mathbb{E}\left[\left(\prod_{j=1}^{N}\left(\zeta R_{j1} + R_{j2}\right)^{g(ij)2}\right)^{\delta} e^{\omega_{i2}}\middle|\,\Omega_{i1}\right] \\
&- e^{\varpi_{i1}} R_{i1} - \psi\mathbb{E}\left[e^{\varpi_{i2}}\middle|\,\Omega_{i1}\right] R_{i2}
\end{aligned}
$$

where the first term represents revenue in $t = 1$, the second term is revenue in $t = 2$, and the last two terms denote costs over the two periods. Here $R_{it} \in \mathbb{R}_{++}$ is the R&D investment (flow) in period t, $\zeta \in [0, 1]$ is the depreciation parameter, while $\psi \in [0, 1]$ is the discount factor. The R&D stock for $t = 1$ is identical to the flow: $R_{i1} = R_{i1}$. For $t = 2$ instead, it is given by the current investment plus the past depreciated flow: $S_{i2} = \zeta R_{i1} + R_{i2}$. Note how connections weights are allowed to vary over time.

Suppose that the game rules are the same as in the one-period case: first nature draws types, then firms observe their own information set, so to make *simultaneous* choices of R_{i1} and R_{i2} for both periods. Now the Bayes-Nash equilibrium is technically expressed as fixed point of $(\mathrm{R}_1, \mathrm{R}_2) = (R_{11}, \dots, R_{N1}, R_{21}, \dots, R_{N2})$. However, there is clearly a one-to-one mapping between a fixed point of R&D flows and a fixed point of both periods' R&D stocks, which are a linear function of flows.

The First Order Conditions are sufficient for a maximum; with some manipulation they can be expressed in terms of R&D stocks as follows:

$$\frac{\partial \pi_i\left(R_{i1}, R_{i2}; \dots\right)}{\partial R_{i1}} = \gamma S_{i1}^{\gamma-1} \mathbb{E}\left[\left(\prod_{j=1}^{N} S_{j1}^{g(ij)1}\right)^{\delta} \middle| \Omega_{i1}\right] e^{\omega_{i1}} + \psi\zeta \mathbb{E}\left[e^{\varpi_{i2}} \middle| \Omega_{i1}\right] - e^{\varpi_{i1}} = 0$$

$$\frac{\partial \pi_i\left(R_{i1}, R_{i2}; \dots\right)}{\partial R_{i2}} = \psi\gamma S_{i2}^{\gamma-1} \mathbb{E}\left[\left(\prod_{j=1}^{N} S_{j2}^{g(ij)2}\right)^{\delta} e^{\omega_{i2}} \middle| \Omega_{i1}\right] - \psi \mathbb{E}\left[e^{\varpi_{i2}} \middle| \Omega_{i1}\right] = 0$$

consequently, the R&D stocks of both periods S_{i1} and S_{i2} is an implicit function of the information set at time 1, Ω_{i1}. Hence, the results from Proposition 2 (and thus of Proposition 3, in the extension of the model that includes conventional inputs) apply in this case as well, with reference to the values of C_1 and L_1 valid at $t = 1$. I omit the proof that the equilibrium is unique under proper conditions as this is a tedious extension of the proof from the one-period case.

Observe how the dynamics of the networks do not matter towards the determination of the equilibrium's stochastic properties: only the information set Ω_{i1} and the cross-correlation of the shocks at the time when the decisions are taken affect the cross-correlation of R&D stocks. This implies that if any new links are generated on $t = 2$, thereby altering cross-firm distances in the network, the spatial correlation of R&D stocks in period 2 would still reflect period 1 circumstances, regardless of any potential serial dependence in the shocks $(\omega_{it}, \varpi_{it})$ over time.

Dynamic R&D Programming

The dynamic programming extension of the problem differs in that the decisions about R_{i2} are based on the information set available at time $t = 2$ *and* on the observation of first period choices, which might reveal information about $(\boldsymbol{\omega}_2, \boldsymbol{\varpi}_2)$. In this case, the First Order Conditions read as:

$$\frac{\partial \pi_i\left(R_{i1}, R_{i2}; \dots\right)}{\partial R_{i1}} = \gamma S_{i1}^{\gamma-1} \mathbb{E}\left[\left(\prod_{j=1}^{N} S_{j1}^{g(ij)1}\right)^{\delta} \middle| \Omega_{i1}\right] e^{\omega_{i1}} + \psi\zeta \mathbb{E}\left[e^{\varpi_{i2}} \middle| \Omega_{i1}\right] - e^{\varpi_{i1}} = 0$$

$$\frac{\partial \pi_i\left(R_{i1}, R_{i2}; \dots\right)}{\partial R_{i2}} = \gamma S_{i2}^{\gamma-1} \mathbb{E}\left[\left(\prod_{j=1}^{N} S_{j2}^{g(ij)2}\right)^{\delta} e^{\omega_{i2}} \middle| \Omega_{i2}; S_{11}, \dots, S_{N1}\right] - e^{\varpi_{i2}} = 0$$

to assess whether the results from Proposition 2 still hold, I distinguish two cases.

1. Past R&D flows do *not* reveal information about current shocks:

$$\mathbb{E}\left[\left(\prod_{j=1}^{N} S_{j2}^{g(ij)2}\right)^{\delta} e^{\omega_{i2}} \middle| \Omega_{i2}; S_{11}, \dots, S_{N1}\right] = \mathbb{E}\left[\left(\prod_{j=1}^{N} S_{j2}^{g(ij)2}\right)^{\delta} e^{\omega_{i2}} \middle| \Omega_{i2}\right]$$

a circumstance that arises *if* shocks are uncorrelated across periods *or* firms do not observe period 1 choices of other sufficiently distant firms. In this case the results from Proposition 2 are still valid, provided that the network grows over time and no connections are severed. The intuition is that in each period, the game is similar to the static model analyzed in the text. The main difference is that optimal R&D flows also incorporate the *expected* future marginal productivity of R&D, itself a function of the *current* information set. Hence, the logic expressed by the proof of Proposition 2 still applies. However, a problem arises if some connections are severed over time. If at time t a link is lost between any two firms i and j ($g_{(ij)s} \neq 0$, $g_{(ij)t} = 0$ for $s \leq t$) then the cross-correlation between ω_i and $\log S_{jt}$ might be nonzero even if i and j are now located at distance higher than $C_t + L_t$, due to the past connection (similarly if intermediate links between i and j are lost). This is a minor concern in the case of the network in this article, as it tends to tighten and become denser over time.

2. Past R&D flows do reveal information about current shocks, as firms are able to recover past shocks of all other firms in the network and use them to predict current shocks (provided that shocks are serially correlated). This circumstance would invalidate Proposition 2, because the model would be similar to a complete information game in periods later than $t = 1$. However, this scenario is not realistic given the evidence provided in Figure 5. In order to rationalize this fact, I make four, not mutually excludable hypotheses.

    (a) The unobserved shocks are serially uncorrelated, which is unlikely.

    (b) For the most part, firms pre-commit to R&D investment plans.

    (c) Between periods, firms do not actually observe the choices of "sufficiently distant" firms. A variation of this idea is that it is too costly for firms to gather and use such "distant" information, as it does not have a first order impact on their outcomes.

    (d) The pattern of cross-firm R&D complementarities is more complex than in the ultimately simplistic expression of "knowledge capital" from (1). Suppose that the R&D stock S_{it} of a firm can be split into several "projects," and that some projects are complementary across (connected) firms while others are not. In this circumstance, firms would not respond to the choices of "sufficiently distant" firms – even if their shocks are known – because these might not affect, in equilibrium, the relevant "projects" of connected firms. This is an intriguing piece of intuition towards further development of the theoretical framework presented in this work.

In either scenario, the model has a unique equilibrium, provided that the spillovers parameter δ is sufficiently small. The proof is once again omitted.

Appendix C Data and Connection Measures

In this appendix I provide details on the dataset construction, with emphasis on the calculation of the connection measures.

C.1 BSV Data

The main panel of firms has been reconstructed by Bloom et al. (2013) by selecting firms from COMPUSTAT with entries in the "Segment" complementary dataset; the latter breaks down the sales of firms by four-digit sector. All variables employed in the estimation of the production function, market value and patent outcome models are constructed according to standard methodologies; in particular, monetary values are deflated using appropriate price indices (see the online appendix BSV for details). COMPUSTAT firm-level identifiers are matched to patents through the NBER patent dataset developed until 2006 (Hall et al., 2001). All the observed patents for each firm i in the entire time interval under analysis are broken down into 426 patent classes defined by the USPTO. Following Jaffe, BSV calculate the $TECH$ weights as the uncentered correlation of two firms' technological allocation of patents:

$$TECH_{ij} = \frac{\left(T_i T_j'\right)}{\left(T_i T_i'\right)^{\frac{1}{2}} \left(T_j T_j'\right)^{\frac{1}{2}}}$$

where $T_i = (T_1, \ldots, T_{426})$ is the vector that collects the shares of patents of each firm across the 426 patent classes. Note that these weights are constant over time. The Jaffe measure of technological proximity is constructed as the average of all other firms' R&D stock weighted by the $TECH$ measures, $Spilltech_{it} = \sum_j TECH_{ij} S_{jt}$. The BSV measure of "business stealing" is similarly constructed, with so-called SIC weights that measure the uncentered correlation of two firms' allocation of sales by sector, and $Spillsic_{it} = \sum_j SIC_{ij} S_{jt}$. To facilitate comparisons, I employ the same variables in my estimates; they enter logarithmically into (12).

C.2 Measures of Connection

To calculate the measures of connection, I need information on i) the disambiguated identity of all the actual inventors who signed all the patents attributed to the firms, ii) their patent collaboration relationships; iii) the time interval in which each inventor is associated to a firm. I obtain information on i) and ii) thanks to the dataset by Li et al. (2014), which I match to the NBER patent data through unique USPTO patent identifiers. I rely on the work by Li et al. for the quality of their disambiguation algorithm; see their paper and data documentation for additional details. Still, I have no direct information about iii). In order to associate individuals to firms, I use indirect information extrapolated from all patents granted between 1976 and 2006.

The dataset allows to identify all assignees associated with the patents of an individual inventor. Thus, by characterizing the time interval in which every individual is observed to collaborate on the patents assigned to a specific firm, I can approximate the actual spell of an inventor-firm association. Specifically, let $\underline{v}_{mi}$ be the first year (application year) when inventor m is observed patenting for firm i; similarly, $\overline{v}_{mi}$ be the last year. The assignment rule between inventor m and firm i in year t is:

$$v_{(mi)t} = \begin{cases} 1 & \text{if } t \in [\underline{v}_{mi} - 1, \overline{v}_{mi} + 1] \\ 0 & \text{otherwise} \end{cases}$$

the linkage is extended one year in the past relative to $\underline{v}_{mi}$ and one year in the future relative to $\overline{v}_{mi}$. This choice is motivated on the presumption that every collaboration does not begin immediately the year the first patent is being applied for, and does not terminate immediately after the last patent. Clearly, this rule might miss those years when an inventor, while not producing patents, is still part of the organization. This would be a problem (by generating issues of measurement error) if these idle inventors were connected to individuals in other firms; on the other hand, it is arguable that idle inventors are not the most active in the process of knowledge creation and exchange. Such a restricted time window essentially captures the size of a firm's R&D-performing team, whether it is composed by regular employees or, say, academic collaborators. It is a reassuring fact that the results are very robust to perturbations of this assignment rule (such additional results are available upon request).

One can collect all the binary indicators $v_{(mi)t}$ in a matrix $\mathbf{V}_t$ which has N rows (number of firms in the data) and M_t columns (the number of inventors at time t). In order to calculate the connection measures, one should first obtain the binary and symmetric adjacency matrix $\mathbf{P}_t$ of connected inventors at time t. This is a matrix of dimension $M_t \times M_t$ where $p_{(mn)t} = p_{(mn)t} = 1$ if the two inventors m and n have at least one joint patent at $t+1$. Define $\mathcal{B}(\cdot)$ as a boolean operator that, when applied to a matrix, returns another matrix whose entries are equal to 1 for positive original entries and 0 otherwise. One can easily calculate the asymmetric $N \times N$ matrix that counts the reciprocal connections between inventors across firms at time t as follows:

$$\mathbf{K}_t = \mathbf{V}_t \cdot \mathcal{B}\left(\mathbf{P}_t \mathbf{V}_t^{\mathrm{T}}\right) = \mathcal{B}\left(\mathbf{V}_t \mathbf{P}_t\right) \cdot \mathbf{V}_t^{\mathrm{T}}$$

and obtain the numerator of the expression within parentheses in (10) for every pair of firms as $k_{(ij)t} + k_{(ji)t}$. Observe how the diagonal elements of $\mathbf{K}_t$ correspond to the total number of inventors assigned to one firm in year t. Hence, the denominator of the aforementioned argument of (10) can be obtained as $k_{(ii)t} + k_{(jj)t}$. Therefore, for any appropriate function $f(\cdot)$ the measures of connections are calculated as follows.

$$c^f_{(ij)t} = c^f_{(ji)t} = f\left(\frac{k_{(ij)t} + k_{(ji)t}}{k_{(ii)t} + k_{(jj)t}}\right)$$

C.3 Geographic Control and Measures of Proximity

The measure of "geographic spillovers" included in the analysis accounts for the relative spatial proximity of two R&D teams. In analogy with the main spillovers variable, this additional variable is based on R&D weights called measures of *proximity*; unlike connections $g_{(ij)t}$, however, these proximity metrics identify inventors as "linked" not because of joint past patents, but if they are "neighbors" in space. Specifically, in each year I assign every inventor to a 2006 Core Based Statistical Area (CBSA) by matching the latter to inventors' ZIP codes reported on patents. I make some assumptions to address cases of inventors with multiple ZIP codes or few observed patents in a given sequence of years (details are available upon request). In analogy with connections, I calculate proximity measures $w_{(ij)t}$ as follows (note that the denominator in the expression below also counts inventors outside US CBSAs, including foreign).

$$w_{(ij)t} = \sqrt{\frac{(\#\text{inventors of firms } i \text{ and } j \text{ overlapping on the same CBSAs at } t)}{(\#\text{ inv.s of firm } i \text{ at } t) + (\#\text{ inv.s of firm } j \text{ at } t)}}$$

Figure C.1 below displays the yearly distributions of proximity metrics $w_{(ij)t}$ conditional on $w_{(ij)t} > 0$; it is similar to that of connections $g_{(ij)t}$ from Figure 3 but it is less asymmetric. The mean value of $w_{(ij)t}$ is 0.051 with 0.035 standard deviation; the maximum value of $w_{(ij)t}$ is 0.5. The actual geographic spillover variables employed in estimations is constructed as $\sum_{j \neq i} w_{(ij)t} \log S_{jt}$.

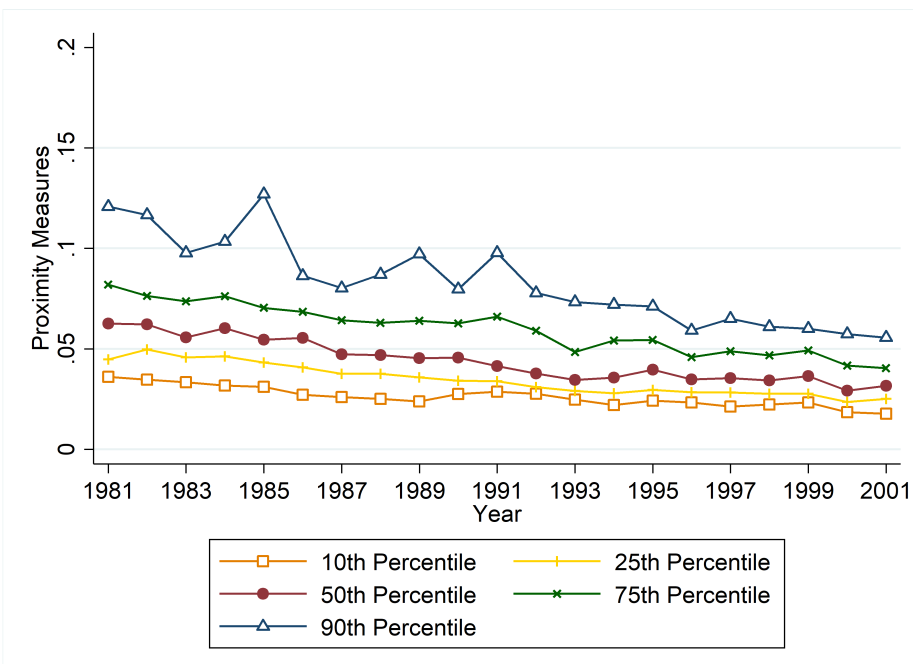


Figure C.1: Distribution of the proximity weights $w_{(ij)t}$ over time

Appendix D Graphical Description of the Network

This appendix collects some visual representations of the network in the form of graphs. For ease of comparison, all nodes (firms) are placed in the same position and have the same size across all figures. Node size is proportional to the total strength of links $\bar{\bar{g}}_i$ summed over all the years. In addition, in Figures from D.1 to D.5 nodes are distinguished by various shades of the orange color; in particular, darker shades denote larger values of network centrality associated with a specific node. The names of selected firms are apposed to some of the largest, most central nodes in the network. A brief introduction or commentary for each of these graphs is given in the list below.

- Figure D.1 displays the network in 1985.
- Figure D.2 displays the network in 1990.
- Figure D.3 displays the network in 1995.
- Figure D.4 displays the network in 2000.
- Figure D.5 displays the "pooled" network, which results from aggregating all edges (connections) over time across all firm pairs.
- Figure D.6 displays the communities obtained through the Louvain algorithm with $\varphi = 1.0$ resolution. There are in total 6 communities: the semiconductor-electronics-ICT, the mechanical, the biotech-pharmaceutical and the chemical industries are identifiable as separate clusters; moreover there are two smaller, mixed groups whose nodes are dispersed across the graph.
- Figure D.8 displays the communities obtained through the Louvain algorithm with $\varphi = 0.8$ resolution. There are in total 10 communities.
- Figure D.8 displays the communities obtained through the Louvain algorithm with $\varphi = 0.6$ resolution. There are in total 20 communities. This partition is used to cluster standard errors in all empirical estimates presented in this work.

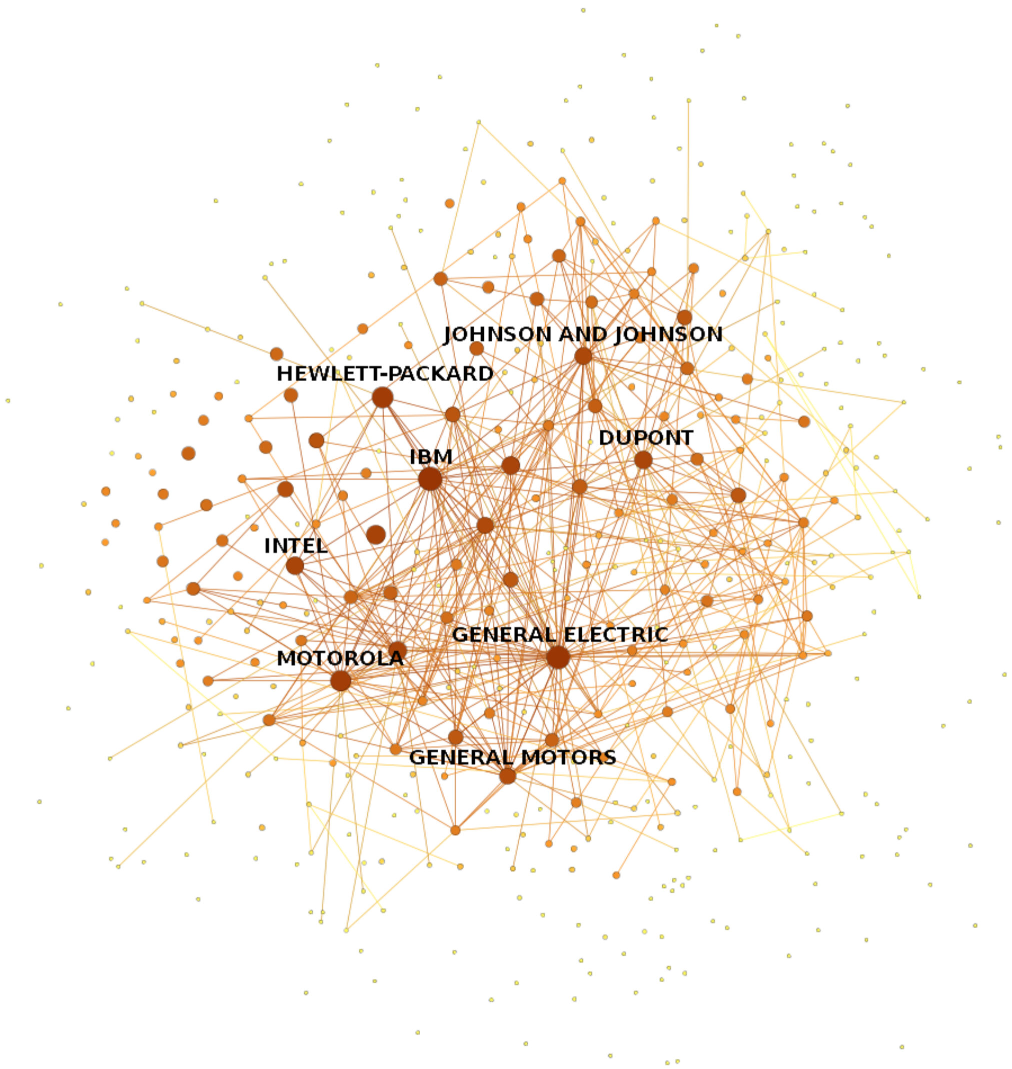


Figure D.1: The Network in 1985

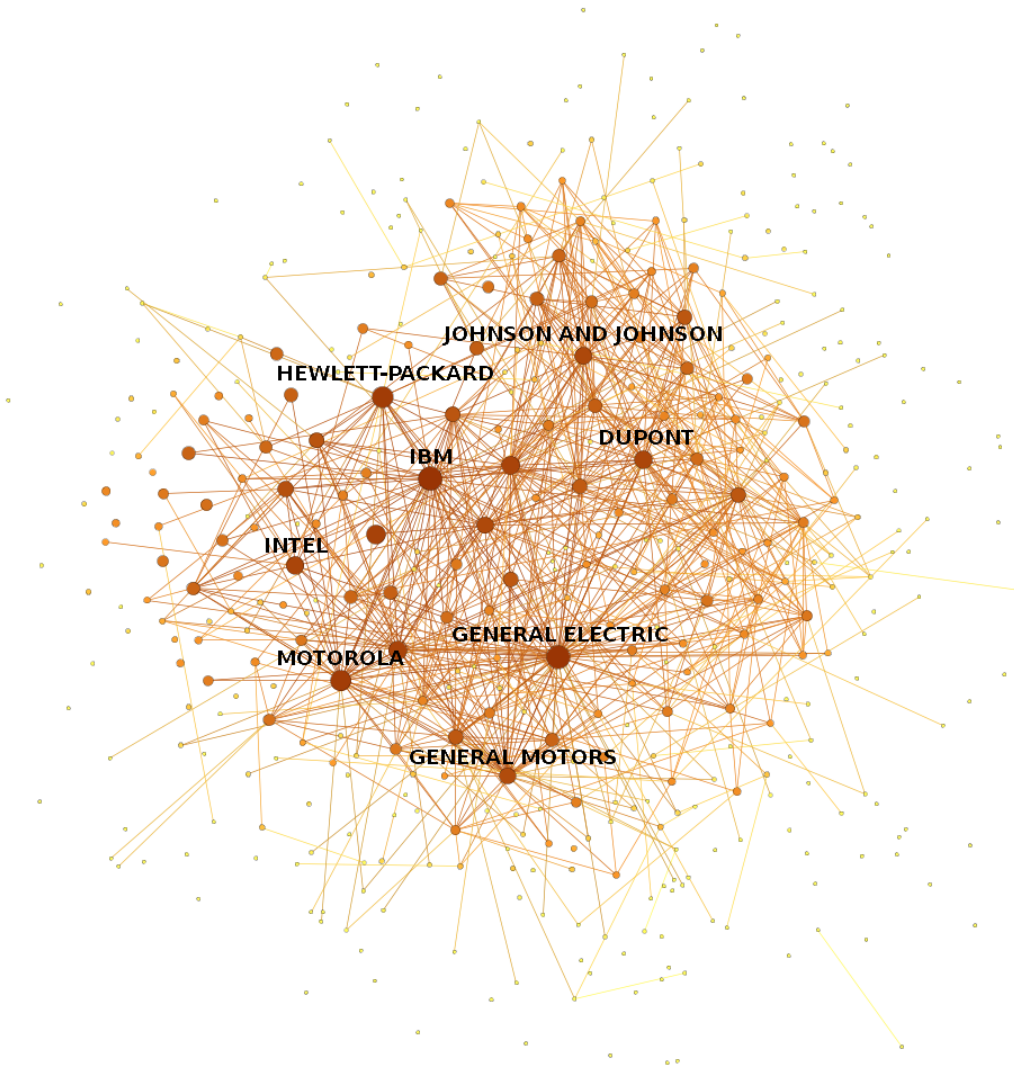


Figure D.2: The Network in 1990

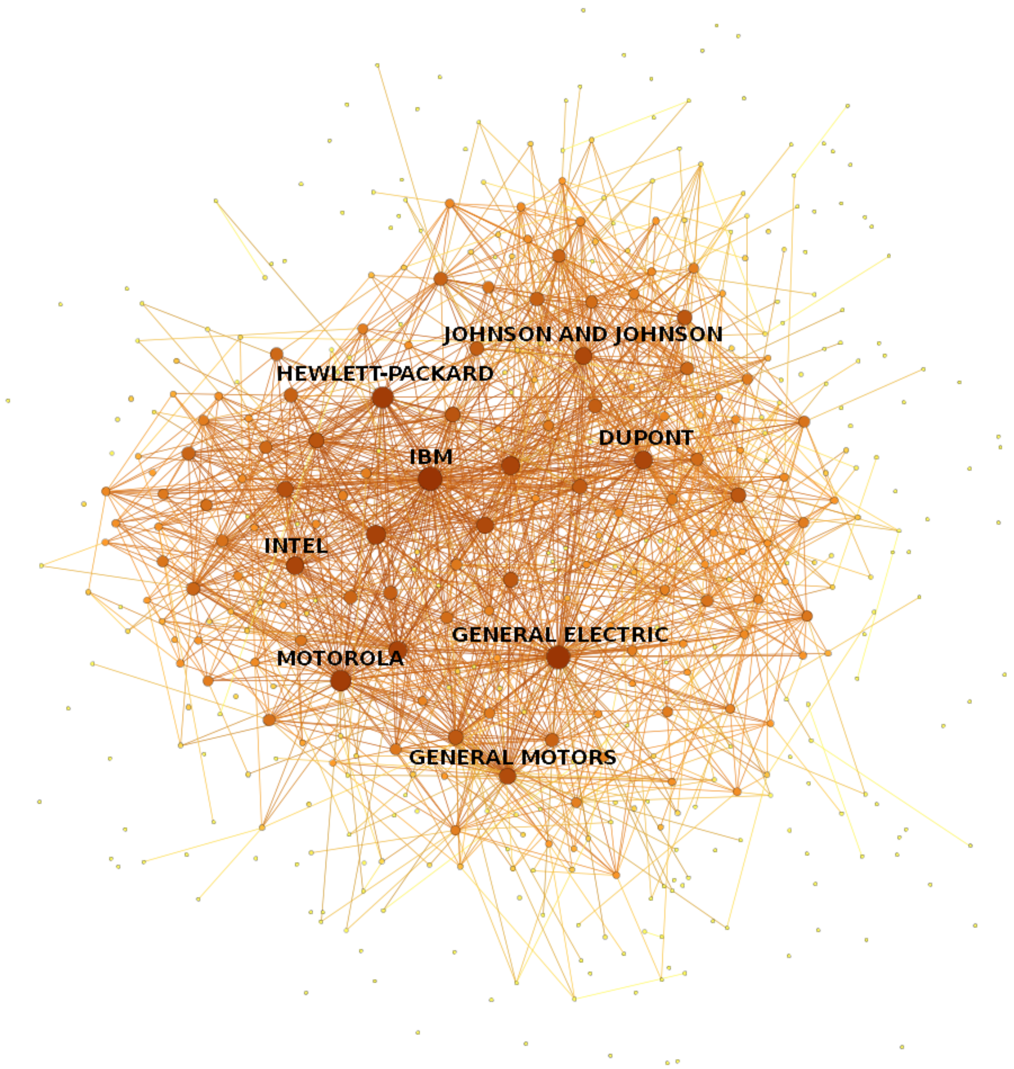


Figure D.3: The Network in 1995

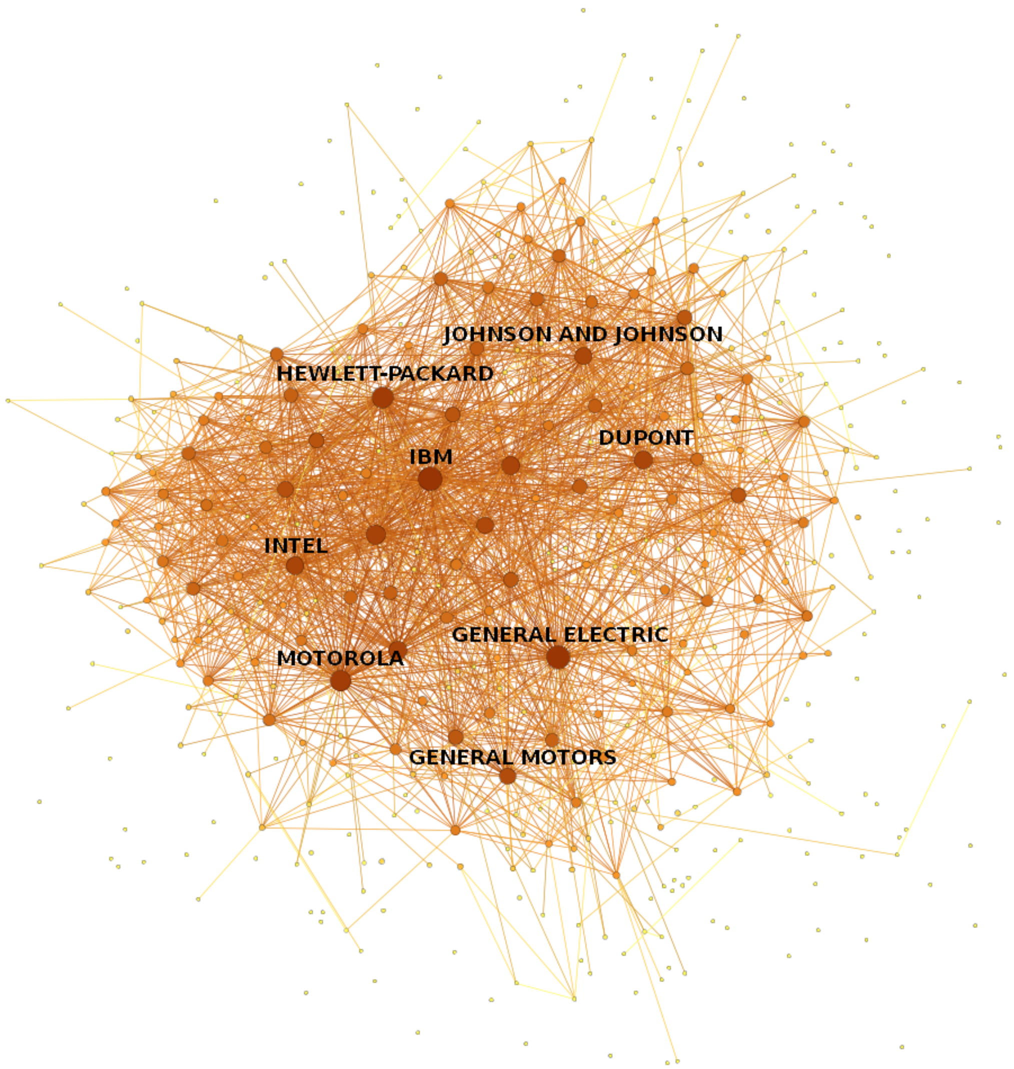


Figure D.4: The Network in 2000

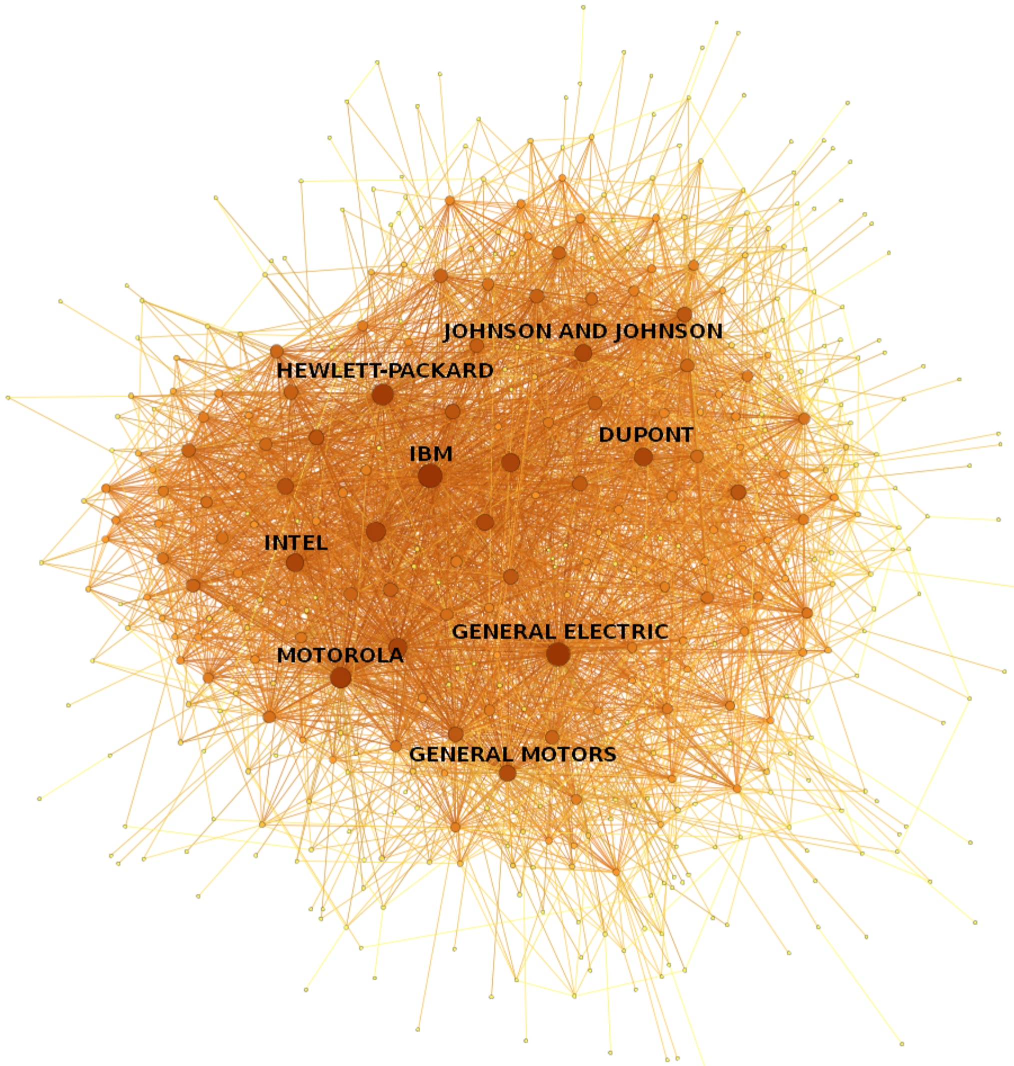


Figure D.5: The "Pooled" Network

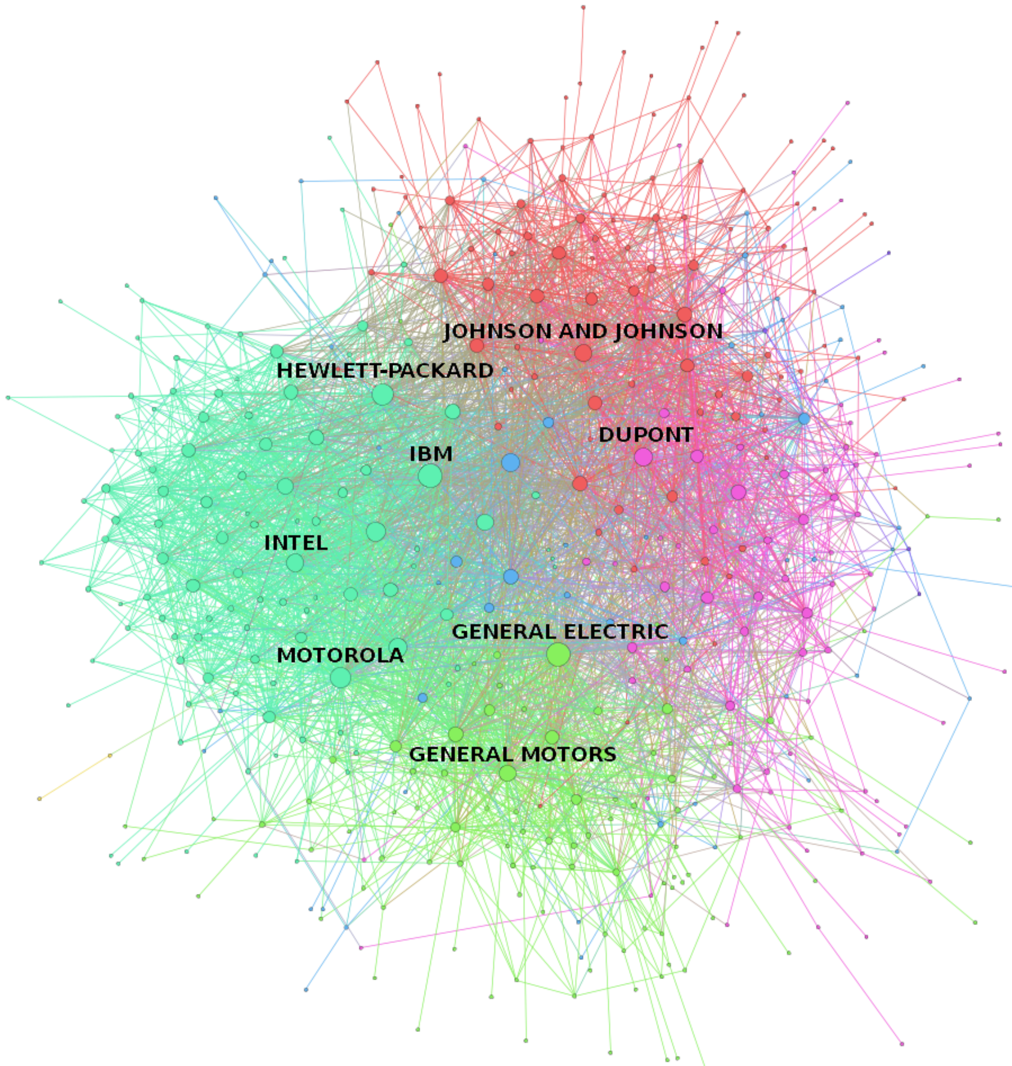


Figure D.6: Network Communities, Resolution $\varphi = 1$

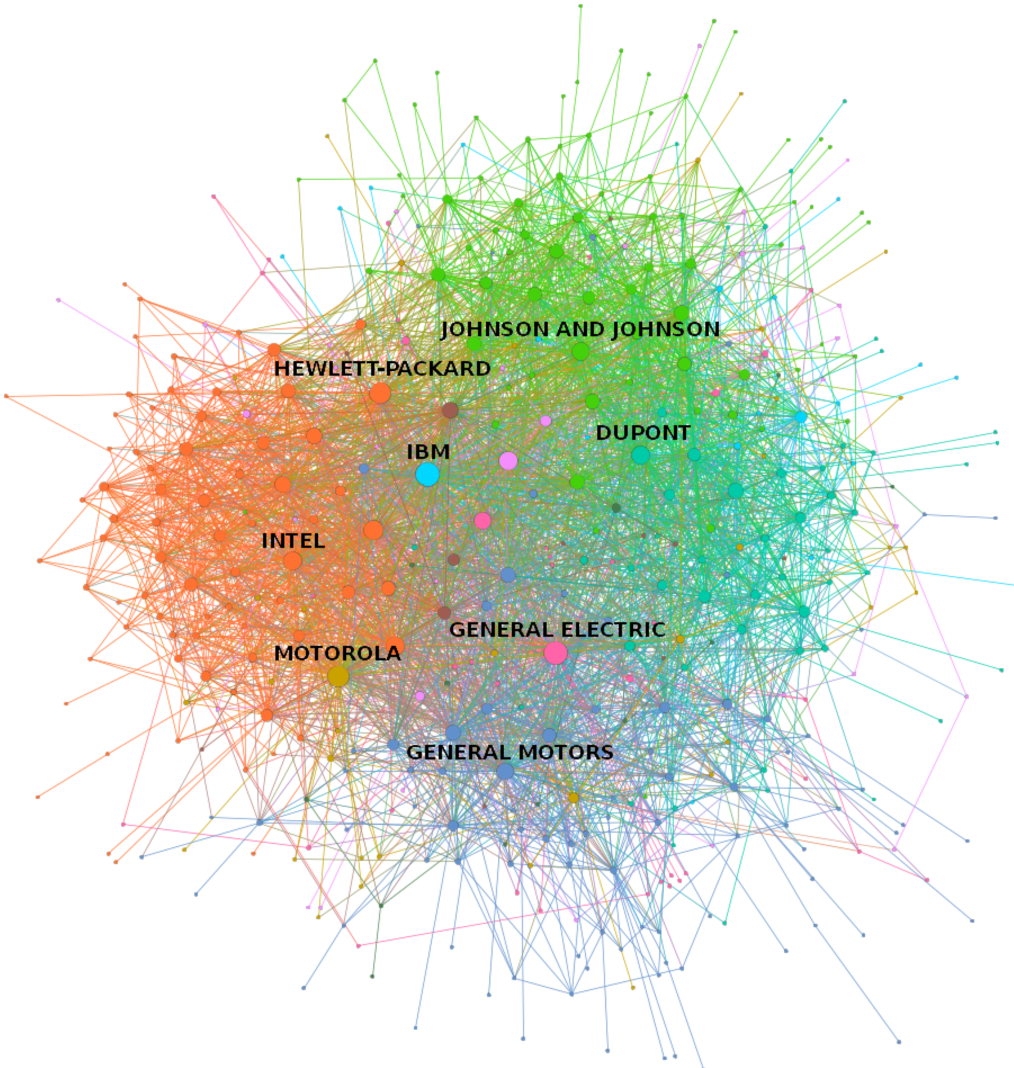


Figure D.7: Network Communities, Resolution $\varphi = 0.8$

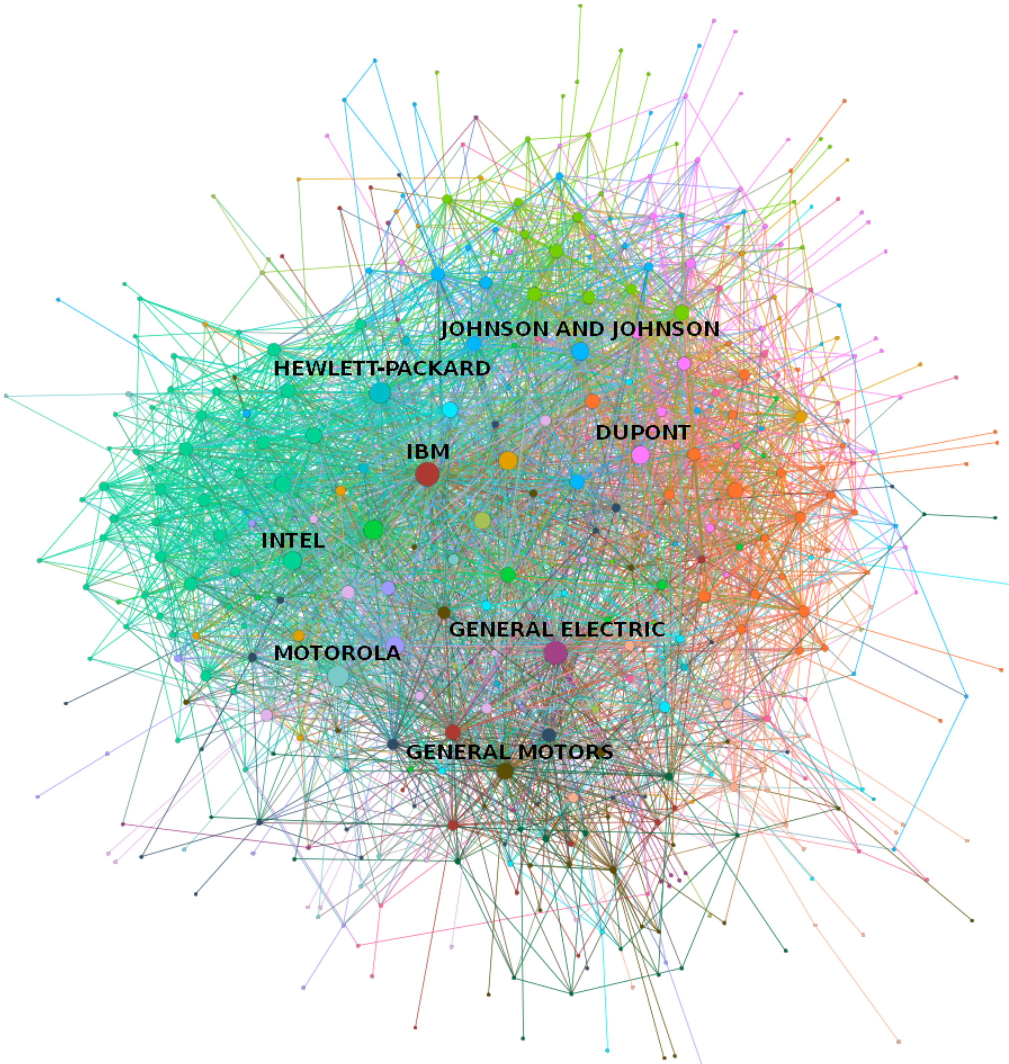


Figure D.8: Network Communities, Resolution $\varphi = 0.6$

Appendix E Additional Empirical Results

In this appendix I provide additional empirical estimates of the production function model (12). The appendix is split in five parts. In the first part, I report the reduced form estimates associated with the 2SLS results discussed in section 4. In the second part I describe the System GMM estimation procedure in detail; in addition I discuss the estimates of the ρ-differenced model. In the third part I report 2SLS and System GMM estimates of the production function extended to firms outside the network. In the fourth part I discuss some preliminary evidence about estimates performed on selected subsets of the data (restricted either by firm size or by industry classification). In the fifth part I discuss connection measures that are alternative to the square root metric, and I report empirical estimates obtained from employing these alternatives in model (12) in substitution of the standard weights.

E.1 Reduced Form Estimates

Table E.1: Production Function, Reduced Form Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0580***	0.0576***	0.0613***	0.0595***	0.0556***
	(0.0116)	(0.0113)	(0.0115)	(0.0125)	(0.0137)
Dist. 2 Instrument ($\times 10^{-3}$)	0.0543***	0.0502***			
	(0.0125)	(0.0156)			
Dist. 3 Instrument ($\times 10^{-3}$)		0.0035	0.0115**	0.0119**	0.0111**
		(0.0060)	(0.0046)	(0.0044)	(0.0047)
Geographic Spillovers	0.0022	0.0022	0.0022	0.0019	0.0010
	(0.0019)	(0.0019)	(0.0022)	(0.0026)	(0.0022)
Capital	0.2048***	0.2047***	0.2095***	0.2080***	0.2081***
	(0.0196)	(0.0197)	(0.0200)	(0.0218)	(0.0206)
Labor	0.6632***	0.6627***	0.6560***	0.6538***	0.6546***
	(0.0353)	(0.0351)	(0.0352)	(0.0358)	(0.0364)
Jaffe Tech. Proximity	0.0549	0.0554	0.0729	0.0390	0.0545
	(0.0576)	(0.0573)	(0.0653)	(0.0838)	(0.0874)
Fixed Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities					
(Community $\times$ Year Effects)	0	0	0	10	20
No. of Observations	7607	7607	7607	7607	7607

Notes: The table reports reduced form estimates associated with first stage and 2SLS estimates from Tables 3 and 4, respectively (see the notes of these tables for details). Instrumental variables are scaled by a 10^{-3} factor for better interpretation of the associated coefficients.

Table E.1 reports reduced form estimates of model (12), where instead of the spillover variable, selected distance D instruments are included on the right-hand side. These are scaled by a 10^{-3} factor to facilitate the interpretation of the associated estimates. As expected, the instruments are positively associated with the dependent variable. Most notably, when both distance 2 and distance 3 instruments are included (column 2), the coefficient of the latter is estimated not statistically different from zero. This should clear concerns about its corresponding first stage estimate from Table 3, which is negative and weakly significant (see the discussion in footnote 28).

E.2 Estimates of the ρ-differenced Model

I describe next the System GMM approach to production function estimation, which is based on Blundell and Bond (2000). Under assumption (18) one obtains:

$$\log Y_{it} = \alpha_i (1-\rho) + \rho \log Y_{i(t-1)} + \sum_{q=1}^{Q} \beta_q \log X_{itq} + \gamma \log S_{it} + \delta \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} -$$
$$- \rho \left(\sum_{q=1}^{Q} \beta_q \log X_{i(t-1)q} + \gamma \log S_{i(t-1)} + \delta \sum_{j=1}^{N} g_{(ij)(t-1)} \log S_{j(t-1)} \right) + \tau_t - \rho\tau_{t-1} + \varepsilon_{it}$$

by subtracting from both sides of (12) the ρ-multiplied, lagged version of (12) itself. The above can be rewritten as:

$$\log Y_{it} = \alpha_i' + \lambda \log Y_{i(t-1)} + \sum_{q=1}^{Q} \left(\phi_{q0} \log X_{itq} + \phi_{q1} \log X_{i(t-1)q} \right) + \chi_0 \log S_{it} +$$
$$+ \chi_1 \log S_{i(t-1)} + \theta_0 \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \theta_1 \sum_{j=1}^{N} g_{(ij)(t-1)} \log S_{j(t-1)} + \tau_t' + \varepsilon_{it} \quad \text{(E.1)}$$

implying the so-called "common factor" restrictions $\lambda = \rho$; $\phi_{q0} = \beta_q$ and $\phi_{q1} = -\rho\beta_q$ for $q = 1, \ldots, Q$; $\chi_0 = \gamma$; $\chi_1 = -\rho\gamma$; $\theta_0 = \delta$; $\theta_1 = -\rho\delta$ in addition to $\alpha_i' = \alpha_i (1-\rho)$ and $\tau_t' = \tau_t - \rho\tau_{t-1}$. Since these restrictions are typically rejected by the data (as the time feedback between inputs, especially stock variables, and output is more nuanced) following standard practice I recover the structural parameters of (12) from estimates of (E.1) and the following steady-state representation of the model:

$$\log Y_i = \frac{\alpha_i'}{1-\lambda} + \sum_{q=1}^{Q} \underbrace{\frac{\phi_{q0} + \phi_{q1}}{1-\lambda}}_{=\beta_q} \log X_{iq} + \underbrace{\frac{\chi_0 + \chi_1}{1-\lambda}}_{=\gamma} \log S_i + \underbrace{\frac{\theta_0 + \theta_1}{1-\lambda}}_{=\delta} \sum_{j=1}^{N} g_{ij} \log S_j + \frac{\tau'}{1-\lambda} \quad \text{(E.2)}$$

and I calculate standard errors through the Delta Method.

To illustrate how I estimate the parameters of (E.1), I let $X_{i1} = K_{it}$ be the capital input and $X_{i2} = E_{it}$ be the labor input. As per the standard System GMM approach, I use the following moment conditions for the equation in differences, for appropriate lags s:

$$\mathbb{E}\left[\begin{pmatrix} \log K_{i(t-s)} \\ \log E_{i(t-s)} \\ \log S_{i(t-s)} \\ \sum_{j=1}^{N} g_{(ij)(t-s)} \log S_{j(t-s)} \end{pmatrix} \cdot \Delta\varepsilon_{it}\right] = 0$$

the following moment conditions for the equation in levels, for appropriate lags s:

$$\mathbb{E}\left[\Delta\begin{pmatrix} \log K_{i(t-s)} \\ \log E_{i(t-s)} \\ \log S_{i(t-s)} \\ \sum_{j=1}^{N} g_{(ij)(t-s)} \log S_{j(t-s)} \end{pmatrix} \cdot (\alpha_i' + \varepsilon_{it})\right] = 0$$

as well as moments of the following kind for the predetermined variables:

$$\mathbb{E}\left[\begin{pmatrix} \log X_{itq} \\ \log X_{i(t-1)q} \end{pmatrix} \varepsilon_{it}\right] = 0 \quad \text{for } q = 3, \ldots, Q.$$

In addition, I include year dummies and, for selected estimates, community-by-year dummies. As specified in the text of the article I set $s \geq 3$; furthermore, for selected estimates, I substitute the spillover instruments in the level equation with moments (19), which are based on the proposed distance D instruments.

Because of the assumption that I make about the spatial cross-correlation of the error term (which extend to ε_{it}) and the chosen clustering scheme that I implement for the calculation of standard errors, a couple of considerations are in order.

- Standard GMM overidentification statistics are not consistent, either because the assumptions that motivate them are violated or because the actual degrees of freedom of the model are too few. Consequently I do not report them.

- After recovering the structural parameters along with their standard errors via (E.2), I perform statistical inferences assuming that t-ratios follow a Student's T distribution with a number of degrees of freedom equal to that of the estimated ρ-differenced model (with 20 clusters, this corresponds to 19 degrees of freedom). The statistical motivation is in the spirit of Bester et al. (2011). Since clusters are large, a Central Limit Theorem argument can be invoked to argue that the contribution of each cluster to the variance-covariance matrix of the estimates of (E.1) follows asymptotically a multivariate normal distribution. The use of the Delta Method, whose asymptotics rest on the fact that the original estimates of (E.1) are consistent, effectively applies the same linear transformation to each of these cluster contributions. If the probability limits of the variances of the RHS variables of (E.1) are uniform across clusters (as in Bester et al., Assumption 3) the resulting t-ratios are by construction Student T distributed with a number of degrees of freedom that equals the number of clusters minus one.

Table E.2: Production Function, Estimates of the ρ-differenced Model, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Sales $(t-1)$	0.8710***	0.8367***	0.8351***	0.8382***	0.8369***
	(0.0164)	(0.0157)	(0.0190)	(0.0198)	(0.0171)
Private R&D (t)	0.0570	0.0481	0.0536	0.0559	0.0511
	(0.0333)	(0.0332)	(0.0323)	(0.0341)	(0.0383)
Private R&D $(t-1)$	-0.0411	-0.0353	-0.0378	-0.0417	-0.0402
	(0.0277)	(0.0284)	(0.0274)	(0.0301)	(0.0336)
R&D Spillovers (t)	0.0052*	0.0041	0.0064**	0.0054**	0.0052*
	(0.0030)	(0.0025)	(0.0027)	(0.0024)	(0.0029)
R&D Spillovers $(t-1)$	-0.0039	-0.0017	-0.0037	-0.0020	-0.0018
	(0.0036)	(0.0033)	(0.0032)	(0.0030)	(0.0033)
Capital (t)	0.1214***	0.1343***	0.1348***	0.1384***	0.1365***
	(0.0316)	(0.0308)	(0.0306)	(0.0322)	(0.0340)
Capital $(t-1)$	-0.0862**	-0.0921**	-0.0919**	-0.0940**	-0.0884**
	(0.0366)	(0.0380)	(0.0380)	(0.0384)	(0.0400)
Labor (t)	0.6253***	0.6219***	0.6245***	0.6114***	0.6182***
	(0.0386)	(0.0426)	(0.0433)	(0.0389)	(0.0402)
Labor $(t-1)$	-0.5537***	-0.5224***	-0.5233***	-0.5145***	-0.5195***
	(0.0327)	(0.0333)	(0.0368)	(0.0348)	(0.0320)
Lags s of GMM-style IVs	$s \geq 3$	$s \geq 3$	$s \geq 3$	$s \geq 3$	$s \geq 3$
Diff. Eq. Spillover IVs	Standard	Standard	Standard	Standard	Standard
Level Eq. Spillover IVs	Standard	$D = 2$	$D = 3$	$D = 3$	$D = 3$
Time Effects	YES	YES	YES	YES	YES
Only Network	YES	YES	YES	YES	YES
No. of Communities					
(Community $\times$ Year Eff.)	0	0	0	10	20
COMFAC Wald statistic	6.75	7.43	10.84	13.19	7.88
(p-value)	(0.0015)	(0.0009)	(0.0001)	(0.0000)	(0.0006)
No. of Instruments	978	962	962	1142	1342
No. of Observations	7185	7185	7185	7185	7185

Notes: This table reports estimates of model (E.1). The associated estimates of the structural parameters are given in the corresponding columns of Table 5 in the text, see that for estimation details. In addition, in this table I report Wald tests about the common factor restrictions implied by the model (COMFAC) for the displayed coefficients. The Wald statistics are assumed to be F distributed with denominator degrees of freedom equal to the degrees of freedom of the model.

Table E.2 shows the estimates of model (E.1) from which the parameter estimates of Table 5 are calculated. To economize on space, the coefficients for the geographic spillovers variable and the Jaffe measure are not reported. Some features of the results are common to all estimates. First, the coefficient of the lagged dependent variable is estimated very precisely always taking values higher than 0.83. This denotes that the data are highly persistent, which is a common feature of firm-level data. Second, the common factor restrictions are always rejected with infinitesimal associated p-values. This is due to the fact that the effect of stock variables on output is distributed over time (note that the only variable for which the structural estimate of Table 5 closely matches the time t coefficient in Table E.2 is the labor input). To illustrate, consider the estimates associated with the spillover variables in column (4): $\hat{\theta}_0 \simeq 0.0055$ for the time t elasticity and $\hat{\theta}_1 \simeq -0.0024$ for the time $t-1$ elasticity. Since the coefficient for the lagged dependent variable is estimated around $\hat{\lambda} \simeq 0.84$, the ratio

$$-\frac{\hat{\theta}_1}{\hat{\lambda}\hat{\theta}_0} \simeq 0.52$$

indicates that much of lagged external R&D also affects future variations of output (or else the ratio would be close to one, as in the case of the labor input). The economic interpretation is that, as it is widely known, the returns to current knowledge build up over time. Finally, it is important to remark that while none of the private R&D coefficients from Table E.2 is statistically significant, the sum $\chi_0 + \chi_1$ is typically significant (obviously, current and past R&D share common variation). For example, the Wald test for $\text{H}_0 : \chi_0 + \chi_1 = 0$ returns a p-value of 2.4% in the case of the estimates from column (3), a p-value of 1.4% for those in column (4), and a p-value of 8.6% in the case of column (5). These tests are not reported in Table E.2 for brevity.

E.3 Estimates Extended to the Whole Sample

Tables E.3 and E.4 illustrate, respectively, 2SLS and System GMM estimates of the production function conducted on the whole sample – that is, not restricted to firms that ever enter the spillover network. The estimates about the geographic spillovers and the Jaffe measure are once again omitted for space reasons. With both methods δ is estimated similarly to the restricted sample case; in fact, by comparing the results from the restricted and non-restricted samples obtained under the same specification, one observes that δ is always estimated larger in the whole sample case. By contrast, γ is always estimated smaller with the whole sample and, in the case of System GMM (which is more prone to issues of statistical power) it is never estimated significantly different from zero. Most likely, this is to be attributed to the fact that firms outside the network are structurally different from those inside it, the former being less R&D intensive and productive than the latter. Estimates restricted to firms that ever enter network are likely to be more robust to this kind of unobserved heterogeneity.

Table E.3: Production Function, Two Stages Least Squares Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0425***	0.0427***	0.0360***	0.0340***	0.0331***
	(0.0104)	(0.0104)	(0.0097)	(0.0106)	(0.0108)
R&D Spillovers (δ)	0.0174***	0.0169***	0.0299***	0.0268***	0.0249***
	(0.0028)	(0.0028)	(0.0073)	(0.0082)	(0.0092)
Capital	0.2047***	0.2050***	0.1979***	0.1952***	0.1959***
	(0.0148)	(0.0148)	(0.0157)	(0.0166)	(0.0161)
Labor	0.6590***	0.6588***	0.6641***	0.6634***	0.6634***
	(0.0251)	(0.0251)	(0.0257)	(0.0262)	(0.0266)
Spillovers IV(s)	$D = 2$	$D = 2, 3$	$D = 3$	$D = 3$	$D = 3$
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	NO	NO	NO	NO
No. of Communities					
(Community × Year Effects)	0	0	0	10	20
No. of Observations	12503	12503	12503	12503	12503

Notes: The estimates in this table replicates those in Table 4, but without restricting the sample to firms that are part of the connections network. See the notes to Table 4 for estimation details.

Table E.4: Production Function, System GMM Estimates, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0769	0.0284	0.0589	0.0331	0.0218
	(0.0483)	(0.0379)	(0.0360)	(0.0330)	(0.0358)
R&D Spillovers (δ)	0.0142	0.0222**	0.0220**	0.0269***	0.0283***
	(0.0097)	(0.0091)	(0.0096)	(0.0088)	(0.0094)
Capital	0.2360***	0.2330***	0.2357***	0.2224***	0.2261***
	(0.0520)	(0.0566)	(0.0554)	(0.0585)	(0.0613)
Labor	0.6261***	0.6785***	0.6763***	0.6896***	0.6985***
	(0.0772)	(0.0776)	(0.0769)	(0.0799)	(0.0852)
Lags s of GMM-style IVs	$s \geq 3$	$s \geq 3$	$s \geq 3$	$s \geq 3$	$s \geq 3$
Diff. Eq. Spillover IVs	Standard	Standard	Standard	Standard	Standard
Level Eq. Spillover IVs	Standard	$D = 2$	$D = 3$	$D = 3$	$D = 3$
Time Effects	YES	YES	YES	YES	YES
Only Network	NO	NO	NO	NO	NO
No. of Communities					
(Community × Year Effects)	0	0	0	10	20
No. of Instruments	978	962	962	1142	1342
No. of Observations	11796	11796	11796	11796	11796

Notes: The estimates in this table replicates those in Table 5, but without restricting the sample to firms that are part of the connections network. See the notes to Table 5 for estimation details.

E.4 Estimates Restricted to Subsets of the Data

In what follows I present empirical results based on selected subsets of the data. Before describing the different partitions and the associated estimates, it must remarked that I refrain from performing System GMM estimation on these smaller subsamples. In fact, System GMM is a very data demanding procedure, which is unlikely to work well upon reducing the number of observations further. However, despite the smaller samples I still calculate standard errors under the conservative clustering approach based on network "communities" that I have employed throughout the analysis.

Table E.5: Production Function Estimates by Firm Size, 1981-2001

	Small Firms		Large Firms	
	(1)	(2)	(1)	(2)
Private R&D (γ)	0.0478*	0.0407*	0.0565**	0.0591**
	(0.0260)	(0.0224)	(0.0211)	(0.0218)
R&D Spillovers (δ)	0.0165**	0.0353**	0.0126***	0.0099
	(0.0065)	(0.0125)	(0.0040)	(0.0160)
Geographic Spillovers	-0.0094	-0.0077	0.0038	0.0037
	(0.0070)	(0.0076)	(0.0033)	(0.0040)
Capital	0.1557***	0.1535***	0.2604***	0.2640***
	(0.0242)	(0.0237)	(0.0299)	(0.0285)
Labor	0.7460***	0.7401***	0.5646***	0.5612***
	(0.0439)	(0.0431)	(0.0446)	(0.0458)
Jaffe Tech. Proximity	0.1638*	0.1560*	-0.0307	-0.0151
	(0.0859)	(0.0768)	(0.1394)	(0.1918)
Estimation	OLS	2SLS	OLS	2SLS
Fixed Effects	YES	YES	YES	YES
Only Network	YES	YES	YES	YES
No. of Communities (Community × Year Effects)	20	20	20	20
No. of Observations	3800	3800	3807	3807

Notes: The table displays separate estimates of model (12) for the "small" and "large" firms in the network, respectively. For both groups, column (1) reports OLS estimates, while column (2) reports 2SLS estimates performed with the $D = 3$ instrument only. All estimates include firm and year fixed effects, as well as community-by-year dummies for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

In Table E.5 I present estimates that are obtained by splitting the sample of firms that *ever enter the network* by firm size. Specifically, I calculate each firm's average number of employees over the time interval when that firm is observed, and I classify a firm as "small" if it falls before the median of average size, and conversely as "large" if its average size equals or surpasses the median. I separately estimate the production function for both groups via either OLS or 2SLS (using the distance 3 instrument). This exercise is restricted to firms in the network because performing it on the whole sample would be largely redundant, as firms in the network are typically larger than those outside of it and the main results are based on the network subsample.

The results evidence that, while OLS estimates of δ are statistically significant for both groups, the 2SLS estimate for small firms is larger and significantly different from zero ($\hat{\delta} \simeq 0.035$), while that for large firms is smaller and not significantly different from zero. Since small firms are less productive (their revenue per employee is on average \$150,000 versus \$191,000 for the large firms, 1996 dollars) this can be seen as preliminary evidence in favor of the hypothesis that the negative bias of OLS estimates is due to spillovers mostly benefiting smaller and less productive firms. Also note that the coefficients associated with the Jaffe measure is positive and statistically significant (albeit weakly) for small firms, which is consistent with the interpretation about the results for the Jaffe measure that has been given in footnote 28 of the text.

I execute a similar exercise for four selections of the sample that are defined by their industry classification. Specifically, I construct four groups of firms that operate in technologically related areas (as informed by primary SIC codes) for which I have enough observations in the data to obtain statistically meaningful estimates. Firms outside the network are allowed. For each group I perform OLS and 2SLS estimates (using the Distance 3 instrument) which are reported in Table E.6. The four selections are described below, along with some comment about the relative empirical estimates.

- BIPHOC: Biotech, Pharmaceutical and Organic Chemistry (primary SIC codes in the range 2820-2880). For this extended industry I do not obtain significant OLS estimates of δ; with 2SLS however, I get $\hat{\gamma} \simeq 0.087$ and $\hat{\delta} \simeq 0.032$, both in the high range (but they are statistically significant at the 10% level only).

- IMACH: Industrial Machinery, excluding office machinery like computers and similar electronic equipment (primary SIC codes in the range 3510-3560). I get significant estimates of δ under both OLS and 2SLS, but twice as large for 2SLS for a staggering $\hat{\delta} \simeq 0.072$ (with $\hat{\gamma} \simeq 0.038$).

- ELELCOM: Electrical, Electronic and Communication sectors, excluding semiconductors as the inclusion of this unique sector affects the results wildly (primary SIC codes in the range 3600-3695, excluding 3674). The OLS estimate of δ is halved and becomes not statistically significant upon moving to 2SLS.

- OLMIN: Optical, Laboratory and Medical Instruments (primary SIC codes in the range 3821-3851). Neither OLS nor 2SLS estimates of δ are economically or statistically significant, but "geographic spillovers" seem to matter in this case.

This exercise shows that, in general, connections-induced spillovers are heterogeneous across industries, and are by no means restricted to frontier technologies.

Table E.6: Production Function Estimates by Industry, 1981-2001

	BIPHOC		IMACH	
	(1)	(2)	(1)	(2)
Private R&D (γ)	0.1101**	0.0876*	0.0445**	0.0385**
	(0.0457)	(0.0477)	(0.0169)	(0.0178)
R&D Spillovers (δ)	0.0038	0.0318*	0.0310***	0.0721***
	(0.0052)	(0.0185)	(0.0096)	(0.0189)
Geographic Spillovers	-0.0043	-0.0048	0.0013	-0.0019
	(0.0066)	(0.0069)	(0.0076)	(0.0082)
Capital	0.2563***	0.2680***	0.2144***	0.1998***
	(0.0480)	(0.0503)	(0.0373)	(0.0360)
Labor	0.6465***	0.6753***	0.7003***	0.6992***
	(0.1139)	(0.1300)	(0.0333)	(0.0358)
Estimation	OLS	2SLS	OLS	2SLS
Fixed Effects	YES	YES	YES	YES
No. of Observations	1012	1012	953	953

	ELELCOM		OLMIN	
	(1)	(2)	(1)	(2)
Private R&D (γ)	0.0594*	0.0625*	0.0712**	0.0712**
	(0.0314)	(0.0329)	(0.0344)	(0.0348)
R&D Spillovers (δ)	0.0206*	0.0111	0.0003	-0.0058
	(0.0110)	(0.0443)	(0.0042)	(0.0133)
Geographic Spillovers	0.0073*	0.0061	0.0175**	0.0171**
	(0.0037)	(0.0070)	(0.0081)	(0.0083)
Capital	0.2017***	0.2061***	0.0852***	0.0867***
	(0.0551)	(0.0612)	(0.0278)	(0.0285)
Labor	0.5791***	0.5773***	0.7896***	0.7921***
	(0.0853)	(0.0866)	(0.0403)	(0.0428)
Estimation	OLS	2SLS	OLS	2SLS
Fixed Effects	YES	YES	YES	YES
No. of Observations	1546	1546	1049	1049

Notes: The table displays separate estimates of model (12) relative to four subsets of the sample, which are distinguished by industry (see the text of the Appendix for detailed description). Subsamples are not restricted to firms belonging to the spillovers network. For each subsample, column (1) reports OLS estimates, while column (2) reports 2SLS estimates performed with the $D = 3$ instrument only. All estimates include firm and year fixed effects but no community-by-year dummies. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

E.5 Alternative Connection Measures

In what follows I illustrate some alternative connection measures $c_{(ij)t}$. After briefly introducing each, I present relevant empirical results obtained by using it in place of the square-root metric (11), along with estimates of the MPR and MSR. Recall that the MSR depends on the distribution of $c_{(ij)t}$, which differs across measures.

- **Linear Connection**. I use a pure linear connection measure $c_{(ij)t}$ (that is, in (10) $f(\cdot)$ is an identity function). This measure does not give disproportionate importance to few connected inventors that are part of two large R&D teams.

Table E.7: Results for the Linear Connection Measure, 1981-2001

	(1)	(2)	(3)	(4)
Private R&D (γ)	0.0385***	0.0543***	0.0485***	0.0806*
	(0.0115)	(0.0143)	(0.0133)	(0.0405)
R&D Spillovers (δ)	0.0632***	0.0579***	0.1546**	0.1689**
	(0.0154)	(0.0147)	(0.0654)	(0.0730)
Geographic Spillovers	0.0023	0.0018	0.0030	0.0127
	(0.0021)	(0.0023)	(0.0023)	(0.0115)
Capital	0.2032***	0.2056***	0.1975***	0.2927**
	(0.0144)	(0.0196)	(0.0213)	(0.1071)
Labor	0.6567***	0.6585***	0.6654***	0.5933***
	(0.0249)	(0.0360)	(0.0378)	(0.1214)
Jaffe Tech. Proximity	0.1271*	0.0334	-0.0112	-0.0628**
	(0.0694)	(0.0855)	(0.0740)	(0.0291)
Estimator	OLS	OLS	2SLS	SGMM
Fixed Effects	YES	YES	YES	(Year)
Only Network	NO	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	20	20	20	20
No. of Observations	12503	7607	7607	7185

Notes: The table displays estimates of model (12) using the alternative connection measure stated in the title. Columns (1) and (2) report OLS estimates, respectively extended to the full sample and restricted to the network subsample. Columns (3) and (4) respectively report 2SLS and System GMM estimates, both employing $D = 3$ instruments. All estimates include firm fixed effects (where applicable), year effects, and community-by-year effects for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

  – Using $\hat{\gamma}$ and $\hat{\delta}$ from column (4) I calculate, among connected firms: average **MPR**; 124%, average **MSR**: 148% (about 119% of the MPR).

- **Second Degree Connections**. With this measure I extend definition of "connected" inventors from the numerator of (10) so to allow for indirect linkages. Specifically, for two inventors in different firms to be "connected" it is no longer necessary that they be past collaborators, but it is enough that there is another inventor who used to be a past collaborator of both. As in (11) I take the square root of the ratio resulting from this alternative definition. Relative to the baseline, this metric downplays connections with scientists who do not develop many bonds *within* their own firm (like occasional inventors). The average value of $c_{(ij)t}$ for this measure is 0.126, with 0.102 standard deviation.

Table E.8: Results for the Second Degree Connection Measure, 1981-2001

	(1)	(2)	(3)	(4)
Private R&D (γ)	0.0372***	0.0520***	0.0470***	0.0713**
	(0.0112)	(0.0138)	(0.0131)	(0.0338)
R&D Spillovers (δ)	0.0046***	0.0042***	0.0078**	0.0065**
	(0.0010)	(0.0010)	(0.0036)	(0.0027)
Geographic Spillovers	0.0022	0.0017	0.0023	0.0058
	(0.0018)	(0.0020)	(0.0019)	(0.0109)
Capital	0.2007***	0.2018***	0.1945***	0.2897**
	(0.0146)	(0.0199)	(0.0239)	(0.1034)
Labor	0.6602***	0.6641***	0.6722***	0.6114***
	(0.0251)	(0.0366)	(0.0402)	(0.1133)
Jaffe Tech. Proximity	0.1193*	0.0203	-0.0136	-0.0687**
	(0.0669)	(0.0764)	(0.0701)	(0.0278)
Estimator	OLS	OLS	2SLS	SGMM
Fixed Effects	YES	YES	YES	(Year)
Only Network	NO	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	20	20	20	20
No. of Observations	12503	7607	7607	7185

Notes: The table displays estimates of model (12) using the alternative connection measure stated in the title. Columns (1) and (2) report OLS estimates, respectively extended to the full sample and restricted to the network subsample. Columns (3) and (4) respectively report 2SLS and System GMM estimates, both employing $D = 3$ instruments. All estimates include firm fixed effects (where applicable), year effects, and community-by-year effects for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

– Using $\hat{\gamma}$ and $\hat{\delta}$ from column (4) I calculate, among connected firms: average **MPR**: 111%; average **MSR**: 118% (about 106% of the MPR).

- **Polynomial Connections**. This measure is analogous to the baseline square root measure, in that it generates some "concavity" that gives relatively more importance to firm pairs with fewer cross-firm inventor connections. However, instead of using the square root function $f(x) = \sqrt{x}$, I construct this measure by using the polynomial function $f(x) = -x^2 + 2x$. Note that this is the only concave polynomial function of second degree that conforms to the requirements on $f(\cdot)$ and also achieves a maximum at 1. At values of x close to 0 this function increases more slowly than the square root function. The average value of $c_{(ij)t}$ for this measure is 0.022, with 0.050 standard deviation.

Table E.9: Results for the Polynomial Connection Measure, 1981-2001

	(1)	(2)	(3)	(4)
Private R&D (γ)	0.0383***	0.0539***	0.0483***	0.0789*
	(0.0115)	(0.0143)	(0.0132)	(0.0401)
R&D Spillovers (δ)	0.0359***	0.0329***	0.0801**	0.0883**
	(0.0081)	(0.0078)	(0.0340)	(0.0386)
Geographic Spillovers	0.0024	0.0019	0.0030	0.0128
	(0.0021)	(0.0022)	(0.0023)	(0.0114)
Capital	0.2029***	0.2051***	0.1975***	0.2920**
	(0.0144)	(0.0196)	(0.0215)	(0.1073)
Labor	0.6569***	0.6590***	0.6655***	0.5948***
	(0.0249)	(0.0359)	(0.0378)	(0.1213)
Jaffe Tech. Proximity	0.1256*	0.0310	-0.0109	-0.0624**
	(0.0692)	(0.0847)	(0.0740)	(0.0288)
Estimator	OLS	OLS	2SLS	SGMM
Fixed Effects	YES	YES	YES	(Year)
Only Network	NO	YES	YES	YES
No. of Communities (Community × Year Effects)	20	20	20	20
No. of Observations	12503	7607	7607	7185

Notes: The table displays estimates of model (12) using the alternative connection measure stated in the title. Columns (1) and (2) report OLS estimates, respectively extended to the full sample and restricted to the network subsample. Columns (3) and (4) respectively report 2SLS and System GMM estimates, both employing $D = 3$ instruments. All estimates include firm fixed effects (where applicable), year effects, and community-by-year effects for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

– Using $\hat{\gamma}$ and $\hat{\delta}$ from column (4) I calculate, among connected firms: average **MPR**: 121%, average **MSR**: 145% (about 120% of the MPR).

- **Squared Connections**. This measure is instead the opposite of the baseline square root measure, as it replaces the latter with the "convex" square function $f(x) = x^2$. Obviously, its economic interpretation is reverse too: this measure gives relatively more importance to spillovers occurring between firms with a lot of inventor cross-connections. While this implication is perhaps unrealistic, it is interesting to examine – for validation purposes – the consequences of choosing this measure on the empirical estimates. In fact, δ is estimated very imprecisely with this metric (the MPR and MRS must be taken with a grain of salt). The average value of $c_{(ij)t}$ for this measure is 0.001, with 0.011 standard deviation.

Table E.10: Results for the Squared Connection Measure, 1981-2001

	(1)	(2)	(3)	(4)
Private R&D (γ)	0.0406***	0.0577***	0.0518**	0.1097**
	(0.0117)	(0.0144)	(0.0183)	(0.0405)
R&D Spillovers (δ)	0.0716	0.0651	2.2043*	0.7093
	(0.0600)	(0.0616)	(1.1804)	(0.4162)
Geographic Spillovers	0.0017	0.0012	0.0040	0.0104
	(0.0023)	(0.0024)	(0.0035)	(0.0118)
Capital	0.2061***	0.2101***	0.1980***	0.3000***
	(0.0148)	(0.0204)	(0.0213)	(0.1042)
Labor	0.6543***	0.6547***	0.6635***	0.5804***
	(0.0250)	(0.0364)	(0.0412)	(0.1228)
Jaffe Tech. Proximity	0.1417**	0.0578	-0.0204	-0.0749**
	(0.0712)	(0.0932)	(0.0940)	(0.0312)
Estimator	OLS	OLS	2SLS	SGMM
Fixed Effects	YES	YES	YES	(Year)
Only Network	NO	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	20	20	20	20
No. of Observations	12503	7607	7607	7185

Notes: The table displays estimates of model (12) using the alternative connection measure stated in the title. Columns (1) and (2) report OLS estimates, respectively extended to the full sample and restricted to the network subsample. Columns (3) and (4) respectively report 2SLS and System GMM estimates, both employing $D = 3$ instruments. All estimates include firm fixed effects (where applicable), year effects, and community-by-year effects for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

– Using $\hat{\gamma}$ and $\hat{\delta}$ from column (4) I calculate, among connected firms: average **MPR**: 169%, average **MSR**: 183% (about 108% of the MPR).

- **Asymmetric "Receiving" Connections**. By abandoning the framework of directed networks one can consider the possibility that spillover relationships are asymmetric between firms. Suppose in particular that the degree of a firm's access to the knowledge of another firm depends only by its *own* share of connected inventors: $k_{(ij)t}/k_{(ii)t}$. By applying the usual square root function to this ratio, the result is the asymmetric "receiving" connection measure $c_{(ij)t} = \sqrt{k_{(ij)t}/k_{(ii)t}}$ (see Appendix C). This metric gives more importance to smaller, well connected firms in the process of ideas exchange. The average value of $c_{(ij)t}$ for this measure is 0.118, with 0.127 standard deviation.

Table E.11: Results for the As. "Receiving" Connection Measure, 1981-2001

	(1)	(2)	(3)	(4)
Private R&D (γ)	0.0378***	0.0532***	0.0532***	0.0498
	(0.0112)	(0.0137)	(0.0130)	(0.0393)
R&D Spillovers (δ)	0.0093***	0.0086***	0.0086**	0.0202***
	(0.0015)	(0.0014)	(0.0037)	(0.0065)
Geographic Spillovers	0.0023	0.0018	0.0018	0.0186
	(0.0020)	(0.0021)	(0.0022)	(0.0137)
Capital	0.2023***	0.2043***	0.2043***	0.2965**
	(0.0145)	(0.0199)	(0.0209)	(0.1101)
Labor	0.6572***	0.6594***	0.6594***	0.6013***
	(0.0250)	(0.0363)	(0.0372)	(0.1275)
Jaffe Tech. Proximity	0.1227*	0.0273	0.0270	-0.0529
	(0.0683)	(0.0817)	(0.0782)	(0.0308)
Estimator	OLS	OLS	2SLS	SGMM
Fixed Effects	YES	YES	YES	(Year)
Only Network	NO	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	20	20	20	20
No. of Observations	12503	7607	7607	7185

Notes: The table displays estimates of model (12) using the alternative connection measure stated in the title. Columns (1) and (2) report OLS estimates, respectively extended to the full sample and restricted to the network subsample. Columns (3) and (4) respectively report 2SLS and System GMM estimates, both employing $D = 3$ instruments. All estimates include firm fixed effects (where applicable), year effects, and community-by-year effects for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

– Using $\hat{\gamma}$ and $\hat{\delta}$ from column (4) I calculate, among connected firms:
average **MPR**: 77%, average **MSR**: 97% (about 126% of the MPR).

- **Asymmetric "Spilling" Connections**. An alternative economic assumption is that spillovers do not depend on the active acquisition of knowledge by well-connected firms, but rather by passive access to spontaneously leaked information. In this scenario it would be more advantageous to have access to as many inventors as possible among "spilling" firms. A connection measure that embodies this idea reads as $c_{(ij)t} = \sqrt{k_{(ji)t}/k_{(jj)t}}$ (see Appendix C). This metric gives more relevance to firms that are well connected to larger ones. However, it gives rise to poor estimates of δ (and so, to an unreliable estimate of the MSR). The average value of $c_{(ij)t}$ for this measure is 0.118, with 0.127 standard deviation.

Table E.12: Results for the As. "Spilling" Connection Measure, 1981-2001

	(1)	(2)	(3)	(4)
Private R&D (γ)	0.0381***	0.0535***	0.0348*	0.0909**
	(0.0115)	(0.0143)	(0.0196)	(0.0352)
R&D Spillovers (δ)	0.0068***	0.0064**	0.0337	0.0106
	(0.0023)	(0.0022)	(0.0211)	(0.0065)
Geographic Spillovers	0.0015	0.0011	0.0007	-0.0003
	(0.0019)	(0.0020)	(0.0028)	(0.0133)
Capital	0.2028***	0.2051***	0.1820***	0.2918**
	(0.0147)	(0.0202)	(0.0338)	(0.1032)
Labor	0.6574***	0.6597***	0.6821***	0.6023***
	(0.0253)	(0.0369)	(0.0460)	(0.1142)
Jaffe Tech. Proximity	0.1215*	0.0228	-0.1374	-0.0810**
	(0.0699)	(0.0877)	(0.1704)	(0.0314)
Estimator	OLS	OLS	2SLS	SGMM
Fixed Effects	YES	YES	YES	(Year)
Only Network	NO	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	20	20	20	20
No. of Observations	12503	7607	7607	7185

Notes: The table displays estimates of model (12) using the alternative connection measure stated in the title. Columns (1) and (2) report OLS estimates, respectively extended to the full sample and restricted to the network subsample. Columns (3) and (4) respectively report 2SLS and System GMM estimates, both employing $D = 3$ instruments. All estimates include firm fixed effects (where applicable), year effects, and community-by-year effects for 20 communities. Standard errors are clustered by the usual set of 20 communities (small sample corrections are applied). Asterisks denote conventional significance levels of t-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

– Using $\hat{\gamma}$ and $\hat{\delta}$ from column (4) I calculate, among connected firms: average **MPR**: 140%, average **MSR**: 146% (about 104% of the MPR).