Knowledge Spillovers through Networks of Scientists

Paolo Zacchia*

June 2016

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 where each link is a function of the relative proportion of two firms' inventors who have patent coauthors 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 neighbors. An empirical concern is that the resulting estimates may reflect unobserved, simultaneous drivers of both R&D and firm performance. I address this problem with an innovative IV strategy, motivated by a game-theoretic model of firm interaction. I instrument the R&D choices of one firm's neighbors with those of firms that are sufficiently distant in the network. With the resulting spillover estimates, I calculate that the marginal social return of R&D amounts to approximately 24% of the marginal private return.

JEL Classification Codes: C31, D85, O31, O33

Keywords: R&D Spillovers, Coauthorship Networks, Indirect Friends

*IMT School of Advanced Studies, Lucca Italy

Theories of *knowledge spillovers* have been central in economic analysis at least since 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 to 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 use coauthorship of 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 coauthors. This metric changes over time, as scientists move across firms or acquire new coauthors.

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 spillover that rely, for instance, on technological similarities between firms (Jaffe, 1986, 1989). Within the network, these measures are not significantly and robustly correlated with relevant firm-level outcomes.

It is arduous, however, to assign a causal interpretation to these findings. As in the case of other types of studies on spillovers and externalities between economic agents, these correlations may simply reflect the existence of common unobserved confounders simultaneously driving R&D, innovation, and firm performance. For example, a sudden technological breakthrough in a specific technological niche where few connected firms operate may facilitate at the same time follow-up discoveries as well as cost savings. An additional hypothesis is that the R&D of similar, connected firms is driven by parallel changes in their R&D costs, which are themselves correlated with firm profitability. This corresponds to the classical case of industries at an earlier stage of their life cycle, which are simultaneously characterized by rapid product innovation and high costs. In both scenarios the unobservability of these confounders would bias, in either direction, standard estimates of R&D spillovers. This problem corresponds to the one of "correlated effects" per the classification by Manski (1993) of identification problems in the estimation of spillover effects.

Thanks to the characteristics of the network that I observe, I am able to formulate a novel empirical strategy that addresses the problem of common confounders. The basic intuition is straightforward. Unobserved factors that correlate across a pair of connected firms – call them i and j – may bias standard estimates of spillovers as long as their R&D expenditures also reflect those factors. Suppose that a third firm k, which is not connected to i, shares some of these unobservables with firm j but not with i. Crucially, it is not required 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. If shared external circumstances affect R&D investment, as hypothesized, 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 firm i's unobservables. I argue that this type of relationships within triads of nodes is commonplace in networks, as evidenced by specific statistical regularities.

To formalize this intuition, 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 exogenously correlated through the network. Consequently, equilibrium R&D also co-varies across neighboring nodes, and the resulting correlation is endogenously amplified by the strategic anticipation of other firms' investment choices. However, under reasonable assumptions that allow for flexible patterns of cross-correlation in the shocks as well as for 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. Since they are still correlated with the choices of direct friends, the choices of firms that lie at that bound would serve as valid instruments. Empirically, I find that that there is no significant cross-correlation in R&D choices at three degrees of separation. This motivates the use of instrumental variables based upon the R&D of "indirect friends" of third degree 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 investment with the R&D choices of indirect friends of third degree, I obtain larger point estimates of spillover effects, for both the productivity and the patent production outcomes. That such difference is only apparent when applying the third-degree instrument in isolation is remarkably consistent with the proposed model. I interpret these findings as evidence that R&D is indeed driven by common confounders across connected firms. The negative bias of OLS estimates lends support to the idea that such correlated factors are in large part related to R&D costs. In light of my results, I estimate the marginal social return of R&D to be about 24% of the private return.

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 of 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 (2016)[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) solved a longstanding issue in the literature, by separating positive R&D spillovers (based on a measure of technological proximity) from the negative *business stealing* effect induced by other firms' R&D. They 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 paper, I provide for

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

[2] König et al. (2016) 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. In particular, their identifying moment conditions might suffer from problems of correlated confounders.

the first time a measure of cross-firms spillovers based on the observation of an actual social network of inventors: specifically, the coauthorship network. This measure aims at capturing all forms of individual interaction between inventors which result in collaborative projects. While spillovers might occur even through less solid, harder to observe types of interactions, the proposed measure has the advantage of generality. In fact, collaborative cross-firm projects characterize spillover mechanisms previously 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. Thus, it is related to the research on 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 where 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 watch others doing increasingly so. Serafinelli (2013) 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]Moretti (2011) lists knowledge spillovers as one of the micro-level determinants of agglomeration economies. There, in fact, are several complementarities between the literature that documents and looks for the causes of agglomeration economies, and the studies on R&D spillovers. Jaffe et al. (1993, 2000) as well as 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) attempt to distinguish a geographic component of spillovers by placing more weight on R&D performed by other firms in the same state. Lychagin et al. (2010) do a similar exercise by exploiting within-firm variation of their R&D activity at a finer geographic level.

[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 described 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. Unlike some of the papers cited above, this model does not give rise to an empirical reduced form equation with a spatially autoregressive dependent variable (corresponding to Manski's "endogenous effect").

papers, 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 motive for concern about studies on peer effects. To the best of my knowledge, this approach is new also in the spatial econometrics literature. It can be viewed as a spatial extension of the GMM procedure for dynamic panels by Blundell and Bond (1998), when the error term has finite serial memory (e.g. it is of the mobile average type).[7]

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 coauthorship-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 implications. Finally, Section 5 indulges in some concluding remarks. A set of appendices accompanies this paper, to complement 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 corre-

According to the classification of spatial models by Elhorst (2014), the resulting equation is instead analogous to a Spatial Durbin Error Model (SDEM), as it includes both an analogue of Manski's "exogenous effect" and spatially correlated errors. The main feature of my model is that both the above are interdependent, causing an endogeneity problem.

[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.

[7]As it is widely known, similar GMM methods for dynamic panels suffer from two main weaknesses. First, the finite memory assumptions is often unrealistic, as the data evidence autoregressive patterns in time. Second, these approaches show problems of the weak instruments type. However, in this paper both issues are less concerning. In fact, spatial correlograms make the case for spatial-MA type of processes in the network. In addition, the chosen instruments are sufficiently strong, perhaps due to the fact that networks have a higher dimension than time series.

lated shocks. The objective of the model is to formalize the intuition motivating the identification strategy of the paper. In a discussion at the end of this section I relate the main empirical predictions to stylized facts about social networks. The model is static and network formation is not explicitly modeled. Furthermore, the micro-level determinants of R&D spillovers are for now taken as given, to be introduced in the next section. In the appendices I provide mathematical proofs and I discuss a stylized dynamic version of the model, showing how the main results can be extended to it.

1.1 Model Setup

An economy consists 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 that is 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. Thanks to the formal and informal exchange of information that happens in the network in the form of spillovers, one firm's knowledge depends not only on R&D that is performed in-house, but also on R&D from other firms that are connected in the network.

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:

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

where the index $g_{ij} = g_{ji} \in [0, 1)$ reflects the relative intensity of spillover effects occurring between any two firms i and j,[8] with $g_{ii} = 0$ for every firm i (a standard normalization). Parameters $\gamma \in (0, 1)$ and $\delta \in (0, 1)$ represent, respectively, the relative contribution of in-house R&D and of knowledge spillovers to one firm's knowledge capital. Notice that R&D is a strategic complement, consistenly with the empirical framework of the paper. 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.[9]

[8] In line of principle, the model allows the intensity of spillovers to be asymmetric

[9] Whether R&D is in reality more of a strategic complement or a strategic substitute is a con-

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

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

where X_{iq} for $q=1,\dots,Q$ is any conventional input (like capital or labor), $\beta_q\in(0,1)$ its associated elasticity parameter, and $A\in\mathbb{R}_+$ is total factor productivity. Moreover, output is affected by a stochastic shock $\omega_i\in\mathbb{R}$. This shock reflects any technological factor that is specific of that firm and affects in either direction its productivity or profitability. Examples may include circumstances like the temporal progression in the learning curve, or production inputs whose quality and quantity is difficult to observe, like say the effort and the effectiveness of the firm management. In more abstract terms the shock ω_i can also be thought as a reduced-form representation, within a supply-side model, of demand-specific factors such as changes in the tastes of consumers for different varieties of goods.

In order to introduce spatial dependence in the unobservables it is necessary to develop a notion of *distance* between firms in the spillovers network. In particular, define $d_{ij}\in\mathbb{N}$ as the *minimum path length*[10] between firms i and j. Notice that this notion is meaningful only for sufficiently sparse networks, that is for interaction structures such that zero-valued connections ($g_{ij}=0$) are sufficiently frequent across pairs of firms. This concept allows me to formulate the following assumption.

Assumption 1. *The unobserved shocks ω_i present positive spatial correlation extending up to C degrees of distance. That is:*

$$\mathbb{C}\mathrm{ov}\left(\omega_i,\omega_j\right)\begin{cases}\geq 0 & \text{if } d_{ij}\leq C\\ =0 & \text{if } d_{ij}>C\end{cases}$$

for any two firms i,j.

---

troversial matter: it is a notoriously hard dichotomy to test. It 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 on the assumed functional form, and standard econometric techniques are not the best means to assess curvature parameters beyond first derivatives.

[10]Path length is the total number of intermediate connections that indirectly connect two nodes in a network (a *path*). Two nodes can be linked via several paths of different length, but usually only the shortest ones among them are of any interest. The minimum path length is popularly referred to as the "degree of separation" between two nodes.

Assumption 1 formalizes the notion that firms, which are sufficiently close in the network defined by the spillover relationships, also share similar technological factors and circumstances affecting their performance. A concrete, famous case is the ICT and computer industry. Close firms operating in that sector have been enjoying for decades parallel trends in the development of increasingly faster computers: the so-called "Moore's Law." Another appropriate example is the pharmaceutical sector, where firms developing new drugs enjoy common advantages based on the findings of basic research. Assumption 1 also implies that firms that are "too distant" in the network, for example those operating in very different technological areas, do not share similar characteristics. Thus, their shocks ω are independent.

Firm i's cost schedule for investing in R&D is also a function of a random variable, $\varpi_i \in \mathbb{R}$, that is spatially correlated in the network:

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

that is, the cost borne by one firm for each additional effective unit of R&D S_i increases with larger values of ϖ_i. Cost factors are correlated in the network, denoting for example common developments in the supply of labor endowed with specific technological skills, or in financing opportunities. This may also represent circumstances such as radical discoveries that pave the way for successive, easier incremental innovations (which in practice increases the return of R&D). The restriction on the statistical properties of the random variable ϖ_i are expressed as follows.

Assumption 2. *The cost factors ϖ_i present positive spatial correlation extending up to C degrees of distance. That is:*

$$\mathbb{C}\mathrm{ov}\left(\varpi_i, \varpi_j\right) \begin{cases} \geq 0 & \text{if } d_{ij} \leq C \\ = 0 & \text{if } d_{ij} > C \end{cases}$$

for any two firms i, j. Moreover, $\mathbb{C}\mathrm{ov}\left(\omega_i, \varpi_i\right) \in \mathbb{R}$ for any firm $i = 1, \ldots, N$.

Assumption 2 formalizes the idea that the spatial cross-correlation of R&D costs is bounded, in terms of network distances, similarly as the productivity shock ω (that the bounding distance C is identical is a simplifying assumption of little consequence). Notice that I am imposing no restriction on the joint probability distribution of (ω_i, ϖ_i); in particular, the covariance of the two shocks can be of either sign. The case for a positive covariance is intuitive: in high-tech industries, increased level of profitability are associated with higher R&D costs. However, some specific situations might be best described by a negative covariance. Moore's Law in ICT, for example, dictates that as processors become more powerful and computers are improved in their quality, development costs are reduced.

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

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

for any firm $i = 1, \ldots, N$. Notice 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 dependent, in equilibrium, 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 some fairly general assumptions about the structure of this set.

Assumption 3. *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 3 states the obvious consideration that firms are aware of their own circumstances (shocks ω_i and ϖ_i). Moreover, 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.

The timing of the game is the following.

1. Nature draws $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)$ and $\boldsymbol{\varpi} = (\varpi_1, \ldots, \varpi_N)$ from a common knowledge joint p.d.f. $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{\varpi})$. Every firm i observes its own information set Ω_i.
2. Firms simultaneously make their R&D and conventional input choices.
3. Payoffs (profits) are paid out.

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: $(\mathrm{S}_i, \mathrm{X}_i) : \Omega_i \to (S_i; X_{i1}, \dots, X_{iQ}) \in \mathbb{R}_{++}^{Q+1}$ for every firm $i = 1, \dots, N$. Denote the vector of all other firms' *R&D* strategies as $\mathrm{S}_{-i} = \{(S_1, \dots, S_N) \setminus S_i\} \in \mathbb{R}_{++}^{N-1}$. A Bayes-Nash equilibrium is a profile of strategies $(\mathrm{S}^*, \mathrm{X}^*) = [(\mathrm{S}_1^*, \mathrm{X}_1^*), \dots, (\mathrm{S}_N^*, \mathrm{X}_N^*)]$ 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 $i = 1, \dots, N$. The following result characterizes the equilibrium.

Proposition 1. *Denote the set of spillover weights as $\mathcal{G} = \{g_{ij} : i, j = 1, \dots, N\}$. If*

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

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

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

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

where $b_i(\mathcal{G})$ the Bonacich-Katz network centrality measure with attenuation factor θ for $i = 1, \dots, N$, while $g_i^*(\Omega_i, \mathcal{G})$ is a firm-specific function of both its information set Ω_i and network topology. Function $g_i^*(\Omega_i, \mathcal{G})$ is spatially recursive and can be bounded by an expression that is linear in the spillover weights:

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

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

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 with the "certain" component of peers' R&D, itself a function $b_i(\mathcal{G})$ 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, of how random shocks to both productivity and R&D costs of all firms in the network (including itself) would affect its own net marginal return of R&D. In equilibrium, in fact, all the shocks may affect the R&D investment of peers, 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 of R&D. In theory, an explosive equilibrium could be catalyzed by a single, very connected firm in the network. 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.

Before characterizing the main result about the spatial cross-correlation of R&D choices, it is worth analyzing the following property of the equilibrium.

Corollary 1. *The model may 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.*

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(\log S_i^*, \log S_j^*\right) &> 0\\ \mathbb{C}\mathrm{ov}\left(\omega_i, \log S_j^*\right) &< 0\\ \mathbb{C}\mathrm{ov}\left(\omega_j, \log S_i^*\right) &< 0\end{aligned} \tag{6}$$

The apparently counterintuitive situation in which, in equilibrium, R&D is positively correlated through the network – but the productivity shocks of some firms are negatively correlated with the R&D of its connections – is actually allowed by the model. It must be remarked that (6) may hold for some pair of connected firms but not for others, depending on *i.* network topology, *ii.* the information structure of the game, and *iii.* the distribution of shocks $\mathcal{F}(\boldsymbol{\omega}, \boldsymbol{\varpi})$. In Appendix A I show that (6) holds for one specific, exemplifying set of restrictions imposed on the model.

This scenario may arise if the spatial cross-correlation of the cost factor ϖ drives that of R&D disproportionately more than the cross-correlation of productivity shocks ω. As R&D costs affect R&D investment negatively, in such circumstances the covariance between one firm's (i) productivity residual and the R&D of some its connections (j) is dominated by a negative component which is driven by the term $\mathbb{C}\mathrm{ov}\left(\omega_i, \varpi_j\right)$. This intuition can be related to a classical stylized fact in the analysis of industry life cycles: higher development costs, as well as higher product prices, are associated to those early stages of an industry where innovation and productivity growth are most rapid (Gort and Klepper, 1982; Klepper, 1996). This discussion is important in light of the empirical results of the paper, as it provides a rationale for OLS estimates of spillovers being negatively biased. This fact cannot be accounted for by standard supply-side models of production featuring spillovers and strategic complementarities, short of introducing some spatial dependence across firm costs.

The next result motivates the empirical strategy of this paper.

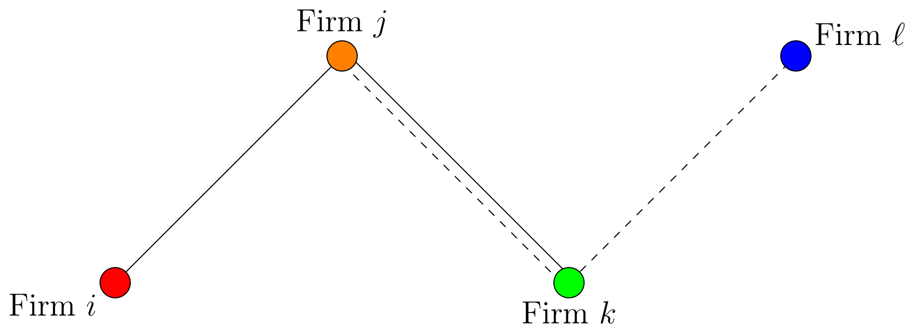
Proposition 2. *Suppose that Assumptions 1-3 hold. It follows that*

$$\mathbb{C}\mathrm{ov}\left(\omega_i, \log S_j^*\right) = 0 \text{ if } d_{ij} > C + L \tag{7}$$

$$\mathbb{C}\mathrm{ov}\left(\log S_i^*, \log S_j^*\right) = 0 \text{ if } d_{ij} > C + 2L \tag{8}$$

that is, the unobserved shock of one firm and the equilibrium strategy of another 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.*

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 also their combined effect 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 (7) and (8), appropriate "predicting" firms are located at any distance 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 made of two open triads[11] that partially overlap on each other, 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 across firms that are reciprocally connected – but not otherwise, as it only reflects private shocks. Consequently, under 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^*), since the two are correlated but the former is independent from firm i's R&D (S_i^*). However, the R&D of firm ℓ, (S_ℓ^*) is not a valid predictor, as it is uncorrelated with that of firm j. Similarly, S_i^* exogenously predicts S_j^* under 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 more complex cases. If $C = 2$ and $L = 0$ firms are still unable to observe the shocks of others, but now the cross-correlation of R&D extends up to two degrees of distance as it reflects the primitive cross-correlation of the shocks. Hence, under 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

[11]An *open triad* is a network, or a subsection of a network, that is made of three nodes – of which two are not connected to one another, while being both connected to the third node. In Graph 1 the two semi-overlapping open triads are represented by a solid and a dashed line, respectively.

shocks, which can be observed between connected firms. Notice 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. Notice 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 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 distant at least three degrees of separation are still valid predictors as per (7).

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

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

$$\mathbb{C}\text{ov}\left(\log X_{iq}^*, \log S_j^*\right) = 0 \text{ if } d_{ij} > C + 2L, \text{ for } q = 1, \dots, Q \tag{9}$$

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 (8) and (9).

1.3 Discussion

I have characterized a general framework that can help identifying, in any network, exogenous sources of variation in the characteristics of connected nodes. In particular, Proposition 2 establishes a connection between the empirical cross-correlation of the strategic spillovers variable and the appropriate level of distance at which to select the "predictors" of one node's links. However, the practical applicability of this framework depends on network topology. Specifically, the network should be neither too tight nor too sparse, and display an appropriate number of "tetrads" like the one in Graph 1. As evidenced later, though, given the characteristics of the data at hand this does not appear to be a relevant concern for the empirical analysis performed in this work.

The main results, in any case, are not trivial. The recent findings of stylized facts about the so-called "three degrees of influence" in networks (Christakis and Fowler, 2013), an expression referring to the typical maximum extent of cross-correlation of nodes' characteristics,[12] currently lack a unified explanation. Economists have only recently started to consider the problem, and to investigate both the economic mechanisms driving it and their empirical implications (Graham, 2014). However simple and stylized, this model offers a possible framework to explain these stylized facts through a combination of both interacting exogenous factors and endogenous influences. In addition, a network approach incorporating spatially correlated heterogeneity can be informative for the empirical analyses of peer effects. This set of studies, in fact, face the challenge of how to appropriately account for the endogeneity problem induced by common confounders (Angrist, 2014).

In Section 3 the core result from this analysis, informed by some descriptive evidence about the empirical cross-correlation of R&D, is exploited to define an empirical strategy based on Instrumental Variables which addresses the chief endogeneity concern in this context: namely, the potential presence of common factors driving both R&D choices and the outcomes of connected firms.

2 Networks and Data

This section is divided in three parts. In the first part, I describe in abstract terms how I characterize the existence of spillover relationships between firms. In particular, I focus on linkages between their R&D-performing teams, on the basis of observable previous collaborations on patents. I formalize the metrics of *connection* that I empirically measure. In the second part, I describe the resulting dynamic network of R&D intensive firms selected from a specific panel of companies listed on the U.S.

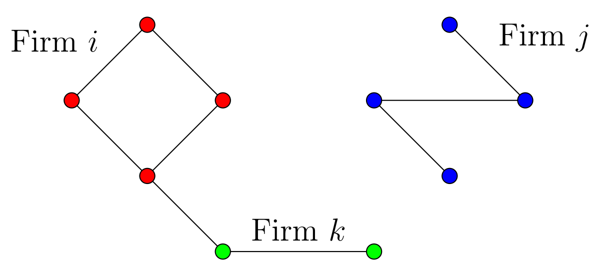
[12] In their work, Christakis and Fowler conduct a systematic investigation of the cross-correlation of adolescents' health habits across linkages of friendship networks. They document descriptively how most individual characteristics do not manifest any significant correlation across indirect friends of third degree or higher, with some variables registering some small third-degree correlation. For the most part, adolescent behavior displays "two degrees of influence," like – in the completely different context of this paper – firm investment in R&D.

stock market. In the third part I provide some relevant descriptive statistics relative to the variables employed in the empirical analysis, as well as estimates of the spatial cross-correlation of R&D in the network which motivate the choice of the instruments.

2.1 The Measures of Connection

Assume that there are three R&D intensive firms whose scientists are related to each others even beyond the borders of their respective organizations. Denote as M_i, M_j and M_k the sets of inventors belonging to each firm, with $M = M_i \cup M_j \cup M_k$. I define an existing *coauthorship relationship* between any two elements of 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 into a patent application listing both their names. Absent such a relationship, it is $p^t_{mn} = 0$. One could visualize the resulting network as a graph where each elements of M is a node, and nodes are linked by edges if $p = 1$.

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



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

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 is the connection between two R&D teams, the stronger are the spillovers occurring 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 function f of the fraction of inventors of either firm who are connected to inventors in 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 $f(1) = 1$. 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. The latter implicitly assumes that the spillover relationship is symmetric between any two firms, regardless of the relative size of their R&D departments.[13] In addition, it must be stressed that a connection measure essentially captures the relative number of personal professional relationships that have been established in the past, in terms of patent coauthorships; it is silent about the relative importance of a single linkage.[14] In Appendix D I explore alternative definitions of connections based on departures from these assumptions.

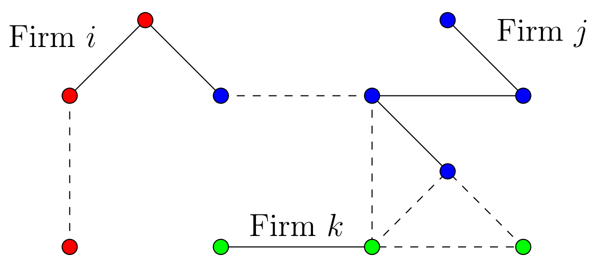
Connection measures between two firms can change over time. Their dynamics are the result of conceptually different types of events that are in principle observable, although I am not able to do so with the available data. Said events are: *i.* cross-firm R&D collaborations, such as joint ventures, resulting say in collaborative 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.[15]

[13] 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. One possibility, for example, is that just few "insiders" may be sufficient to grasp most of the secrets of one firm's R&D activity.

[14] A departure from this assumption is to consider that connections between two inventors that involve prolonged relationships over the years, relationships that result in many jointly filed patents, can be more relevant than others. Similarly, connections involving superstar inventors who issue many patents, of which some have been extremely well cited, can be of exceptional importance.

[15] New entrants increase the denominator of (10), but can potentially generate new cross-firm

Graph 3 extends the previous example by the advancing of one time period to $t = 1$, and examining the consequences of various changes in the underlying network of coauthorship. New linkages between inventors, due to newly appearing joint patents, are represented by dashed lines. In the example, some inventors of firm j have been observed to patent jointly with researchers from firm k, including an entrant inventor from that company. There is also a new entrant in firm i, but he is not connected to anyone elsewhere. 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 $c^f_{(ik)1} = 0$.



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

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 narrative 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 D I present the empirical results from applying alternative definitions of connections; including ones based on the pure ratio of cross-connected inventors.

linkages, thus determining the tightening of connections. Similarly, the exit of scientists would decrease the denominator of (10), as well as the numerator of it if the exiting inventors were playing the role of connecting firms to each other.

2.2 Firm-level Network

In the empirical analysis, I combine two different data sources. I construct the dynamic firm-level network on the basis of 707 firms from the unbalanced panel employed by Bloom et al. (henceforth BSV). Their sample consists of mostly manufacturing, R&D intensive firms listed in the U.S. stock market and belonging to the COMPUSTAT database, that are observed in the time interval of 1976-2001. This panel is representative of the bulk of private R&D performed in the US, which is concentrated among the largest and most productive firms. The dataset assembled by BSV includes information on accounting measures, various indicators of innovation performance, as well as the Jaffe-type measures that they use in their paper to disentangle different types of externalities.

I match the firm-level identifiers to the NBER patent data in order to obtain the list of patents assigned to each firm in the time interval under analysis. Subsequently, I match the official USPTO patent numbers to the Harvard Patent Network Dataverse (HNPD), which is a dataset that allows to identify with great accuracy the individual inventors signing each patent application. This is made possible by applying a specific disambiguation algorithm based upon the information contained in patents registered at the USPTO; specifically, some formulation of inventors' names and their ZIP codes of residence (Li et al., 2014). Ultimately, this results in the selection of 1,315,060 patents signed by 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 signed patents. Specifically, for two inventors m and n, I assign $p^t_{mn} = 1$ if at time $t+1$ the USPTO has received at least one patent *application* signed *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.[16] Similarly, in order to assign inventors to firms one has to extrapolate facts on the basis of limited available information. I use the sequence of patents signed by inventor m and assigned to firm f 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.

[16]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.

I calculate the connection measure for each pair of firms and year. In total, 460 firms out of 707 display at least one positive connection with another firm in any year from 1981 to 2001.[17] The number of firms that are actually connected in any year varies with time: some of the initially unconnected firms would eventually develop bonds. Similarly, the 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. Because of this, one never observes all the 460 firms of the dynamic network in each cross section. Figure 1 shows how many connected firms appear in each year, as well as the total number of yearly observed bilateral connections.

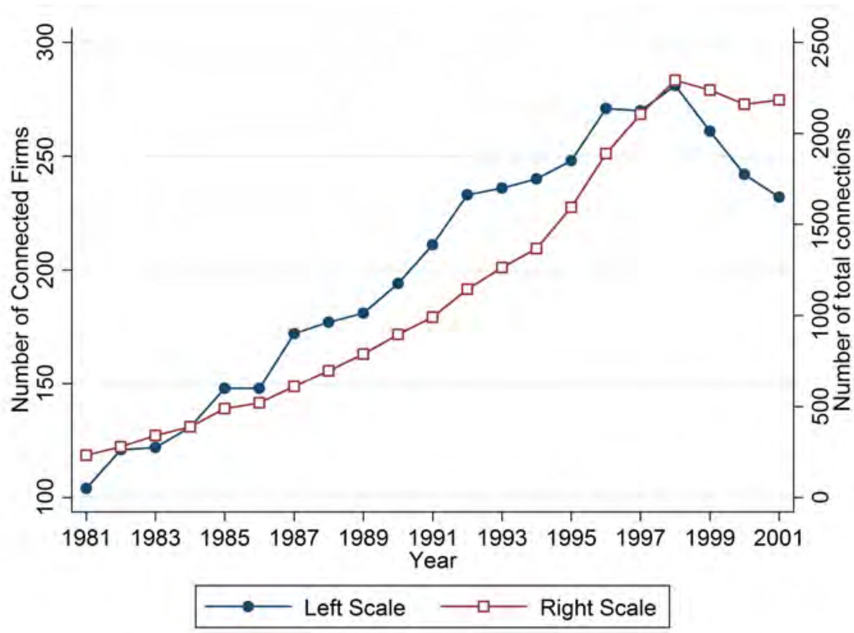


Figure 1: Connected Firms and Total Connections over time

Figure 1 displays a steady rise in the total number of connected firms between 1981 and 1998, to be followed by a drop from 1998 to 2001 because of losses of singleton connections by smaller firms. However, the total number of linkages, and thus the overall density of the network, remains quite stable during the final years of the sample. Another way to appreciate this temporal evolution is to visualize the actual network, in the form of graphs, as it looks like in different years. Selected graphs (for the years 1985, 1990, 1995 and 2000) are reported in Appendix E.

[17] I calculate existing connections in 1981 thanks to information on patents with both USPTO

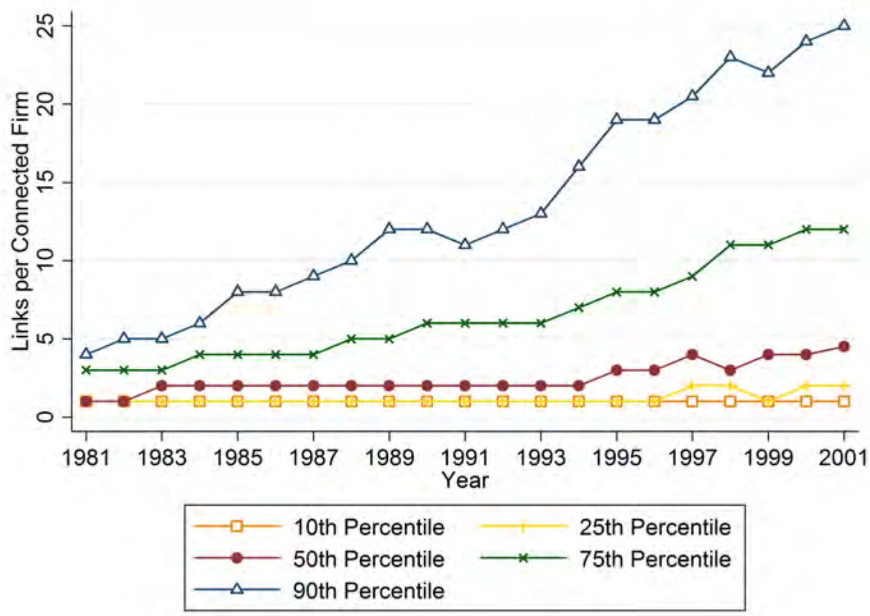


Figure 2: Degree distribution (binary connections) over time

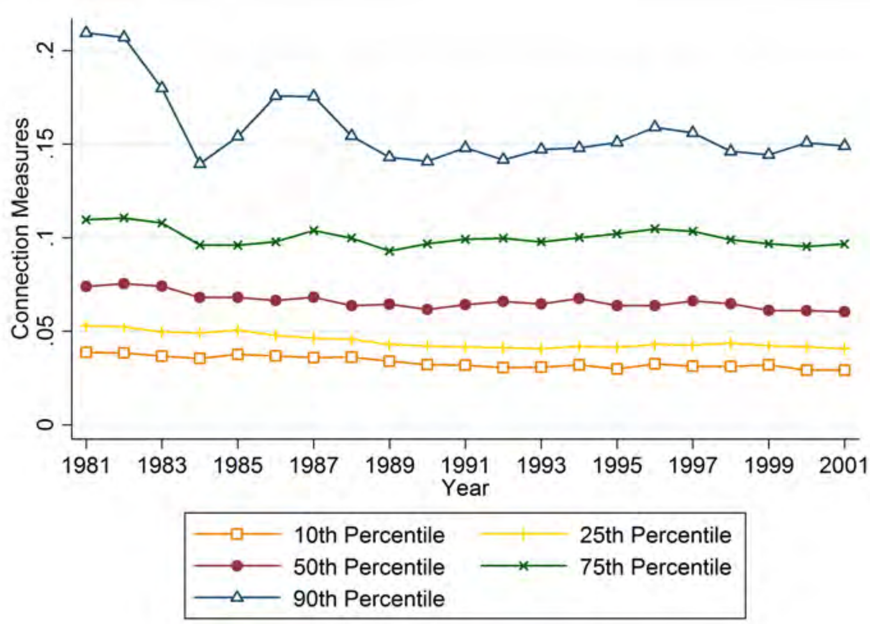


Figure 3: Distribution of the of connections ($g_{(ij)t}$) over time

issue date and application year subsequent to 1975.

Figure 2 reports the yearly degree distributions. Like in many networks, it is a very asymmetric one and it tends to widen over time. The most connected firms in the early eighties have less than 10 links, but several dozens of them around year 2000. Similarly, the average number of connections increases from about 1.5 to about 5. Each of these connection measures equals to 0.083 on average, with a 0.066 standard deviation.[18] The average hides another asymmetric distribution: this is displayed in figure 3 and is quite stable over time. In order to interpret the empirical estimates, one may also want to consider the *total* amount of spillovers that a firms receives from others to which it is connected. A measure that combines the variability in the degree distribution together with the variability in the strength of links is the *row sum* of connections, which is defined as $\bar{g}_{it} = \sum_{j \neq i} g_{(ij)t}$. Its yearly empirical distributions are displayed in figure 4. The aggregate mean and standard deviation of $\bar{g}_{it}$ are respectively 0.44 and 0.18. Apparently, the increase in its spread over time is due to the widening of the degree distribution.

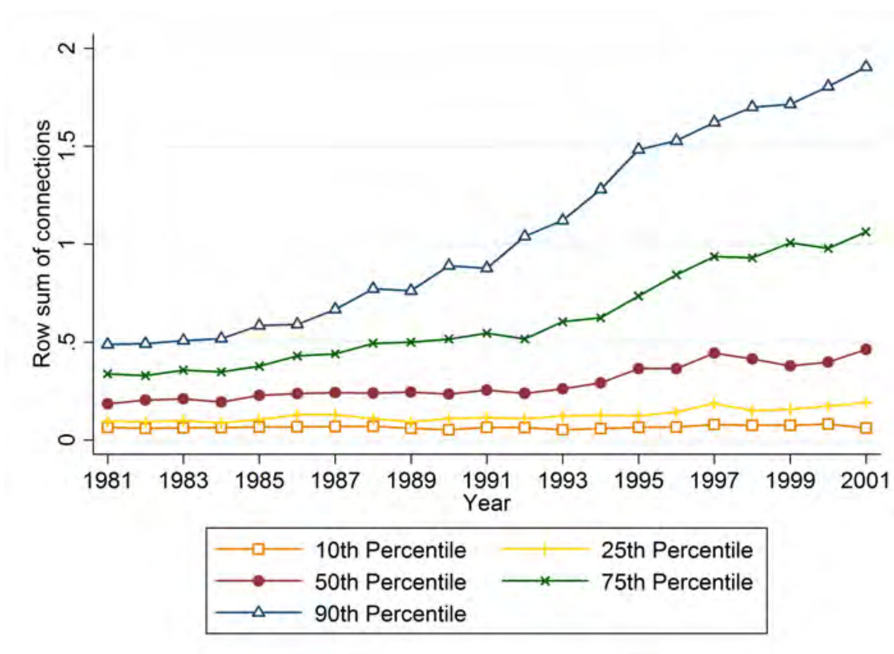


Figure 4: Distribution of the Row-sum of connections ($\bar{g}_{(it)} = \sum_{j \neq i} g_{(ij)t}$) over time

In Figure 5 I report the temporal evolution of the *Network Census*, that is the total count of both open and closed triads. The relative proportion between the two

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

types of triad is a valid measure of the overall network density. Over the time interval 1981-2001, one can count in total 160,365 open triads and 15,623 closed triads (which are about 9.1% of the total). The number of both types of triads grows together with the total number of connections. The network is neither "too dense" nor "too sparse;" in fact, in light of the discussion at the end of Section 1 it displays a density that is appropriate for the empirical strategy that I propose.

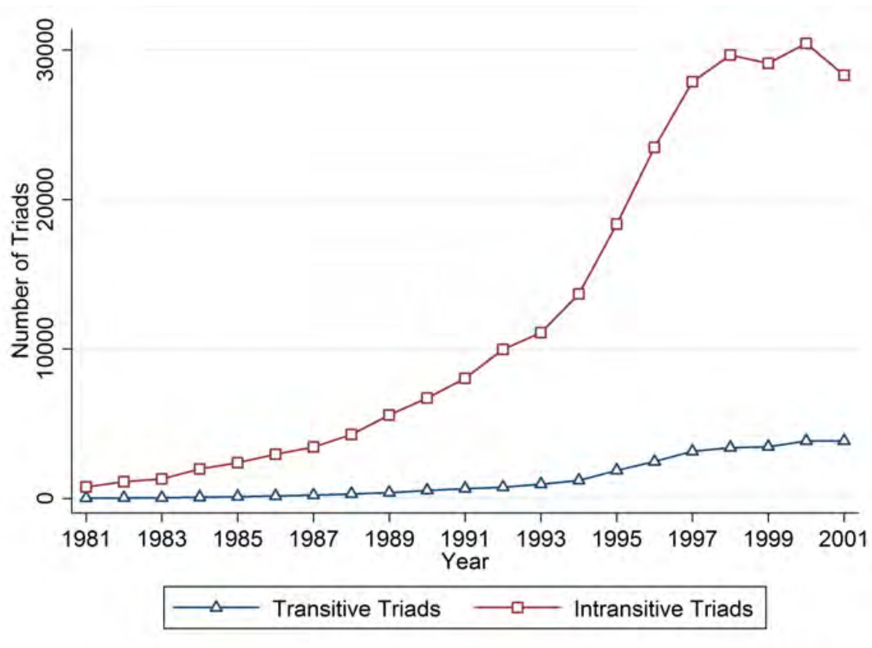


Figure 5: Network Census

2.3 Summary Statistics and Spatial Correlation

In Table 1 I provide some firm-level summary statistics. I divide the sample into five groups: the firms that do not belong to the network, and four groups for those that do. In particular, I calculate the *overall* sum of connections for each firm as $\bar{\bar{g}}_i = \sum_t \bar{g}_{it}$ and assign each firm to a group on the basis of its classification by quartile of $\bar{\bar{g}}_i$. Quartile 1 contains the least connected firms in the network over the time interval; quartile 4 contains the most connected ones.[19] For each group, I provide the mean and standard deviation of specific variables by pooling all the years in the sample.

[19]The four quartile-groups do not exactly contain the same number of observations because of attrition in the unbalanced panel.

For this reason, in addition to real sales (Y_i), other outcome measures (Tobin's q, citation-weighted patents P_i) and number of employees (L_i), I also report the ratio of Y_i to several input or spillover measures. The last row, in particular, represents the ratio of Y_i on the measure of *knowledge capital* defined in (1) and employed in the empirical analysis. Table 1 highlights the fact that the firms that belong to the network – in particular the most connected among them – are larger, more R&D intensive and more productive than the excluded ones.

Table 1: Summary Statistics, 1981-2001

	No Network	Quartile of $\sum_t \bar{g}_{it}$ 1	2	3	4
Y_i: Sales (Millions 1996$)	530	1086	1402	2198	10736
	(1206)	(2396)	(2509)	(4530)	(20440)
V_i/W_i: Tobin's Q	1.812	1.882	2.521	2.726	3.420
	(1.870)	(1.757)	(2.939)	(3.259)	(4.084)
P_i: Cit.-weighted patents	4.099	15.60	22.76	70.35	647.8
	(12.45)	(43.28)	(44.74)	(136.2)	(1328.6)
L_i: Employees (Thousands)	3.428	6.862	9.407	12.35	57.22
	(6.560)	(15.64)	(16.90)	(22.57)	(98.06)
Y_i/L_i: Labor Productivity	138.6	138.9	163.8	165.5	207.7
	(84.5)	(109.9)	(94.6)	(122.2)	(165.1)
Y_i/K_i: Capital Productivity	7.156	5.670	5.618	5.421	4.781
	(6.319)	(3.561)	(4.368)	(3.777)	(3.902)
$Y_i/R\&D_i$	43.67	20.27	53.73	11.53	4.435
	(144.3)	(72.1)	(494.5)	(35.4)	(3.936)
$Y_i/$ Jaffe Measure (i)	56.5	109.7	142.2	213.9	986.2
	(128.3)	(242.2)	(267.2)	(434.5)	(1819.9)
$Y_i/\prod_j R\&D_j^{g(ij)t}$		980.5	899.0	587.8	197.6
		(2299.2)	(1858.3)	(1816.3)	(1167.1)
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 $\bar{\bar{g}}_i$. All descriptive statistics are pooled over years. $R\&D_i$ denotes the R&D *stock* of firm i. Standard deviations are in parentheses.

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 6 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 some variable of interest for some given level of distance (Kelejian and Prucha, 2001). The calculation is performed by pooling together all pairs of firms at the same level of distance throughout all the years. Figure 6 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.

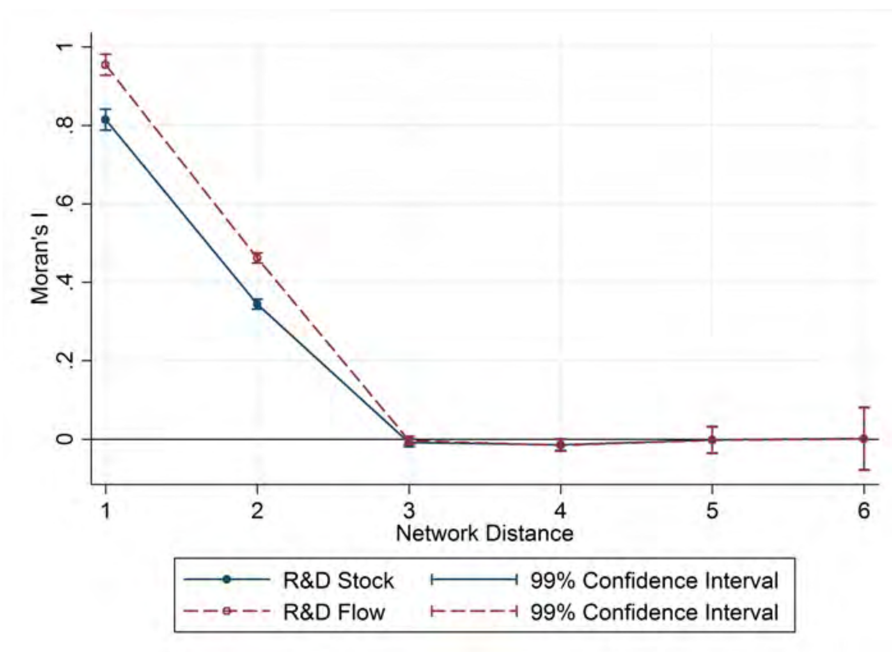


Figure 6: Spatial Correlogram of R&D Measures, 1981-2001

According to the analytical framework of the paper, the spatial cross-correlation of R&D reflects either the exogenous cross-correlation of firm-specific characteristics (ω_i, ϖ_i), or the endogenous strategic dependence between R&D choices, or both. In light of Proposition 2, the evidence in Figure 6 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-type 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 of third degree 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 degrees of distance. Instruments based on higher distances present no correlation with the R&D of direct connections, as evidenced by Figure 6.

3 Econometric Model

In this section I outline the econometric methodology that I employ to estimate R&D spillovers induced through personal connections. This section is divided in three parts. In the first one, I introduce the workhorse model I use to evaluate the productivity effects of connections' R&D: an augmented production function. In the second part, I characterize the Instrumental Variable strategy that I adopt to control for correlated effects. In particular, I detail the construction of the instruments. In the third part I describe the two models for the estimation spillovers on both the market value and the innovation rate of firms.

3.1 Production Function

The workhorse empirical model of the empirical analysis is an augmented production function. Specifically, 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 Total Factor Productivity A_{it} is allowed to vary across firms and over time, with $\log A_{it} + \omega_{it} = \alpha_i + \tau_t + \upsilon_{it}$ being decomposed into 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 $t-1$. 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 one firm's 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 $\bar{g}_{it} = 4$ receives a 4δ percentage increase in 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 private (in-house) R&D stock.

In most specifications, I estimate the model using the same set of controls $\{X_{itq}\}_{q=1}^{Q}$ as in BSV. These include measures of the capital and labor inputs elaborated from accounting data, as well as synthetic controls for industry-level sales and price indicators. In addition, I include the main spillover variables employed in the study by BSV. The first one of them corresponds to the classical "Jaffe" measure of spillovers, which is based on the similarity in the technological classification of any two firms' set of patents. This is meant to capture the positive effect of knowledge spillovers. The second measure accounts for the negative "business stealing" effect of competitors' R&D in downstream product markets. It weighs R&D on the basis of the overlap of two firms' sales across industries.[20] In order to mitigate concerns of endogeneity, in their study BSV substitute several variables in $\{X_{itq}\}_{q=1}^{Q}$ – including conventional inputs and measures of spillovers – with their first lags. I conform to their choices so to facilitate the comparison and interpretation of the respective results.

In most regressions I additionally include a measure that accounts for the relative intensity of R&D performed in the metropolitan areas where a firm's inventors are predominantly concentrated. This way, I attempt to control for the possibility that cross-firm connections as defined in (11) simply capture their spatial proximity as well as other parallel endogenous factors.[21] I call this measure "Geospills;" Appendix

[20]These are respectively called "Spilltech" and "Spillsic" by BSV. In their analysis, only "Spilltech" has a significant effect in the production function. The "Spillsic" measure has, though, a significant effect on other outcome measures in the analysis by BSV. Because of its importance, the construction of the Jaffe measure ("Spilltech") is described in greater detail in Appendix C.

[21]I construct this measure in close analogy with how I construct the measure of connection, by weighing R&D of other firms with appropriate pair-specific and time-varying metrics. For the sake of simplicity, I call such metrics "measures of proximity". They are based on the relative number of

C provides additional details on its construction. In general, I simultaneously include different measures of spillovers in the same estimation models. Hence, I am able to more convincingly restrict the interpretation of the estimates of δ to the sole effect of the R&D performed by firms that are linked through the coauthor-induced network.

3.2 Instrumental Variables

The estimation of δ in equation (12) suffers from two potential endogeneity problems. The first one is the possible presence of common confounders that drive *both* the choice of R&D and productivity for connected firms. If such confounders are not observed, the OLS estimate of δ incorporates their effect on the outcome, to the degree that they are correlated to R&D of connected firms. They correspond to the *correlated effects* as per the analysis by Manski (1993) of spillovers in the classroom, and are the empirical counterpart of the correlated introduced in the analytical framework. The second problem is endogeneity of connections. I address them both.

Consider the case where the real population regression function reads as

$$\log Y_{it} = \alpha_i + \sum_{q=1}^{Q} \beta_q \log X_{itq} + \gamma \log S_{it} + \delta \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tau_t + \omega_{it} + \varepsilon_{it} \tag{13}$$

with $\upsilon_{it} \equiv \omega_{it} + \varepsilon_{it}$, $\mathbb{E}\left[\omega_{it}\right] = \mathbb{E}\left[\upsilon_{it}\right] = 0$ and ε_{it} is a pure white nose error term. As discussed in the analytical framework, if both $\mathbb{E}\left[\omega_{it}\omega_{jt}\right] \neq 0$ and $\mathbb{E}\left[\omega_{it} \log S_{it}\right] \neq 0$ hold simultaneously, it follows that:

$$\mathbb{E}\left[\omega_{it} \log S_{jt}\right] \neq 0$$

in light of the discussion of Corollary 1, the expression above can be either positive or negative. Weighting by $g_{(ij)t}$ and summing over j results in:

$$\mathbb{E}\left[\omega_{it} \sum_{j=1}^{N} g_{(ij)t} \log S_{jt}\right] \neq 0 \tag{14}$$

an inequality that may go in either direction. In fact, (14) indicates both the sign and the size of the bias in the OLS estimation of δ.

inventors who, for every pair of firms, are observed to be resident in the same set of metropolitan areas, that are defined at the CBSA level.

The analytical framework suggests a strategy to address the problem of correlated confounders: to predict the R&D of one firm's direct connections with the R&D of other firms that are "sufficiently" distant in the network. According to the theoretical analysis, this is always possible as long as the spatial cross-correlation of R&D has "finite memory." Figure 6 evidences that this is indeed the case and that appropriate instruments are based on the R&D of firms either located at both distances 2 or 3, or just the latter. In the example given in Graph 1, the R&D of both firms k and ℓ could serve as an instrument for the R&D of firm j. Formally:

$$\mathbb{E}\left[\omega_{it}\log S_{kt}\middle|\left\{\log X_{itq}\right\}_{q=1}^{Q},\log S_{it}\right]=0 \tag{15}$$

$$\mathbb{C}\mathrm{ov}\left[\log S_{kt}\log S_{jt}\middle|\left\{\log X_{itq}\right\}_{q=1}^{Q},\log S_{it}\right]\neq 0 \tag{16}$$

where (15) can be thought of as the component of a more general moment condition and (16) motivates the power of the instrument. Since a firm generally has more than one connection, and each of them corresponds with more than one indirect linkage, in theory one can combine the entire resulting set of moments in several ways.[22]

Here I propose a straightforward way to aggregate the R&D stocks of all indirect connections of given distance into a single instrument. I start with the simplest case of "indirect connections" of second degree, which are indexed by k. Define the indicator $\tilde{h}^{i}_{(jk)s}=g_{(jk)s}\mathbb{I}\left[g_{(ik)s}=0\right]$ for $s\leq t$: thus $\tilde{h}^{i}_{(jk)s}$ selects the "friends of my friends who are not my friends themselves." The "indirect spillovers" instrument reads as

$$\begin{aligned}\sum_{k\neq i}h_{(ik)s}\log S_{kt}&=\sum_{j\neq i}g_{(ij)s}\sum_{k\neq i,j}\tilde{h}^{i}_{(jk)s}\log S_{kt}\\&=\sum_{k\neq i,j}\left(\sum_{j\neq i}\left(g_{(ij)s}g_{(jk)s}\right)\mathbb{I}\left[g_{(ik)s}=0\right]\right)\log S_{kt}\end{aligned}$$

where the weights $h_{(ik)s}$ on the left hand side are implicitly defined by the expression on the right hand side. It is a direct consequence of (15) that

$$\mathbb{E}\left[\left(\sum_{k\neq i}h_{(ik)s}\log S_{kt}\right)\omega_{it}\middle|\left\{\log X_{itq}\right\}_{q=1}^{Q},\log S_{it}\right]=0$$

[22] A possibility would be to aggregate all the appropriate moment conditions such as (15) in a larger GMM problem, and analyze how the efficient weighting matrix varies as a function of network topology. This is an intriguing topic for further research.

holds, and that at the same time the instrument should retain some predictive power for the endogenous regressor $\sum_{j=1}^{N} g_{(ij)t} \log S_{jt}$ as long as (16) is true. Furthermore, Corollary 2 from the analytical framework ensures that the instrument is also uncorrelated with the set of estimated inputs, resulting in consistent estimates of δ.

I now describe how to aggregate the indirect connections located at distance three. These can be referred with colorful terminology as "indirectly indirect connections:" they are the nodes in the network, indexed by ℓ, that are three degrees of separation far, and that lack direct linkages with both direct "friends" and indirect connections of second degree. In analogy with the distance two case, let $\tilde{q}^i_{(k\ell)s} = h_{(k\ell)s}\mathbb{I}\left[g_{(i\ell)s} = 0\right]$ and aggregate over ℓ:

$$\begin{aligned}\sum_{\ell \neq i} q_{(i\ell)s} \log S_{\ell t} &= \sum_{k \neq i} h_{(ik)s} \sum_{\ell \neq i,j,k} \tilde{q}^i_{(k\ell)s} \log S_{\ell t} \\ &= \sum_{\ell \neq i,j,k} \sum_{k \neq i,j} \sum_{j \neq i} \left(g_{(ij)s} g_{(jk)s} g_{(k\ell)s}\right) \mathbb{I}\left[g_{(ik)s} = 0\right] \mathbb{I}\left[g_{(i\ell)s} = 0\right] \log S_{\ell t}\end{aligned}$$

where the weights $q_{(i\ell)s}$, implicitly defined by the expression on the right-hand side, represent the strength of all indirect connection paths between i and ℓ. These weights are equal to zero if any firm ℓ has a direct connection with firm i, or with any of the first-degree direct or second-degree indirect connections of firm i.

While this strategy might address the problem of correlated confounders, another endogeneity problem could still affect the estimates: connections themselves could be correlated to firms' unobserved factors. For example, it is likely that highly prolific, well connected inventors are more inclined to move towards more productive firms.[23] I integrate a proposed solution to the problem in the construction of the instruments above, by exploiting time variation in the networks. In particular, the two instrumental variables actually employed in the estimation, for both "indirect" and "indirectly indirect" connections, are defined for $s = t - 1$. This amounts to implicitly assume that

$$\mathbb{E}\left[\omega_{it} \middle| \left(g_{(11)t-1}, \ldots, g_{(1N)t-1}\right), \ldots, \left(g_{(N1)t-1}, \ldots, g_{(NN)t-1}\right)\right] = 0$$

that is, productivity shocks are orthogonal to the first lag of the network topology – and thus to all possible linear combinations of connection weights. However, unobserved factors are still allowed to correlate with *changes* in the network topology at

[23] In this case it is unclear whether the resulting bias would be positive, as moving inventors link together both high- and low-productivity firms.

time t. In the empirical analysis I only show results based on instruments defined at $s = 1$, but I have also experimented with further lags of the network topology. They result in diminished precision, while point estimates are not affected considerably.

Finally, it is important to consider that the cross-correlation across unobserved shocks invalidates standard asymptotic properties of any GMM/2SLS 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 any partition of a network into different clusters would result in some form of cross-cluster dependence. Bester et al. advocate using as few and 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 which is used to identify hierarchies of "communities" or clusters. At each level of the hierarchy, connections are dense within groups and sparse between groups. The algorithm can be fine-tuned by varying the "resolution parameter" ρ which selects different levels of the hierarchy.[26] In order to strike a balance between the the CCE approach by Bester et al. and standard practices of clustering standard errors, I set $\rho = 0.6$ so to obtain 20 clusters. Because of serial correlation, all observations of the same firm in the panel enter the same cluster. Appendix E provides further details

[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 coordinates (e.g. a set of locations on the map). However, networks are inherently multidimensional, and there exist many competing notions "distance" in networks. 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.

[26]A value of $\rho = 1$ defines a partition of few large communities; smaller values of ρ break down these clusters and select groups by moving down the hierarchy.

and a visualization of cluster assignment. Inferences are not substantially altered by the definition of clusters or the choice of ρ – which determines their number.

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 of inventions. The effects on productivity can be indirect (thanks to new or better patents/products) or direct (because of the immediate applicability of spilled knowledge in 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, largely following the empirical specifications by BSV. I begin from a market value specification: I regress the Tobin's q on a model of this sort:

$$\log\left(\frac{V_{it}}{W_{it}}\right) = \tilde{\alpha}_i + \sum_{q=1}^{\tilde{Q}} \tilde{\beta}_q C_{itq} + \tilde{\delta} \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tilde{\tau}_t + \tilde{\upsilon}_{it} \tag{17}$$

where V_{it} is the market value of a firm measured at time t and W_{it} is the replacement value of its assets. Notice here that the set of controls is different than in model (12). In particular, $\{C_{itq}\}_{q=1}^{\tilde{Q}}$ includes the Jaffe measure of spillovers as well as a polynomial of sixth degree of the ratio S_{it}/K_{it}, to control for differences in R&D intensity.[27]

The estimation of spillovers effects on the innovation rate is based on a count model for citation-weighted patents P_{it}:

$$P_{it} = \exp\left(\breve{\alpha}_s + \sum_{q=1}^{\breve{Q}} \breve{\beta}_q R_{itq} + \breve{\gamma} \log S_{it} + \breve{\delta} \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \breve{\tau}_t + \breve{\upsilon}_{it}\right) \tag{18}$$

[27] A polynomial of S_{it}/K_{it} is derived from a Taylor series approximation of the right-hand side term $\log(1 + \theta S_{it}/K_{it})$, which is customary in the specification of Tobin's q type of models. BSV show that the empirical results are similar whether one uses a six-degree polynomial or non-linear estimation methods instead.

which, in order to account for values of $P_{it} = 0$, is typically estimated via maximum likelihood[28] (Hausman et al., 1984; Blundell et al., 1995). The set of controls $\{R_{itq}\}_{q=1}^{\check{Q}}$ includes a term for the lag the dependent variable ($\log P_{i(t-1)}$) as well as the Jaffe measure. In order to control for endogeneity, I adapt my IV strategy by employing a control function approach. Specifically, I take the residuals from first stage linear regressions of the endogenous outcome on the excluded instruments and the other controls, and I include them in the estimation of the non-linear model.

4 Empirical Results

In this section I present the empirical results of the paper; it is divided in five parts. In the first part I present the baseline (OLS) results for the production function model. In the second part, I address the endogeneity concerns of correlated confounders by applying the proposed IV strategy. 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 and last part I offer some additional considerations, in particular about 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 the standard errors at by the communities defined with the Louvain algorithm with resolution $\rho = 0.6$. Along with the estimate of γ and δ I report those for Capital and Labor. One can interpret the estimate of $\delta = 0.017$ from column (1) in light of different thought experiments. For example, 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 0.0014. 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 an average row-sum connection of $\bar{g}_{(ij)t} = 0.44$, this corresponds to a 0.0073% increase. The elasticity of privately undertaken R&D, $\hat{\gamma} = 0.044$ appears in comparison to be one order of magnitude larger.

[28]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.

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

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0322	0.0282	0.0396	0.0384	0.0389
	(0.0084)	(0.0080)	(0.0089)	(0.0106)	(0.0115)
Spillovers (δ)	0.0165	0.0140	0.0120	0.0133	0.0111
	(0.0036)	(0.0029)	(0.0030)	(0.0034)	(0.0032)
Geospills		0.0003	0.0003	0.0003	0.0003
		(0.0002)	(0.0001)	(0.0001)	(0.0001)
Capital	0.1640	0.1625	0.1741	0.1712	0.1676
	(0.0127)	(0.0133)	(0.0214)	(0.0229)	(0.0226)
Labor	0.6414	0.6465	0.6356	0.6352	0.6323
	(0.0179)	(0.0183)	(0.0282)	(0.0307)	(0.0295)
Jaffe Tech. Proximity		0.2324	0.0837	0.1068	0.0693
		(0.0843)	(0.0770)	(0.0902)	(0.1015)
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	NO	YES	YES	YES
No. of Communities (Community $\times$ Year Effects)	0	0	0	10	20
No. of Observations	12009	12009	7336	7336	7336

Notes: The table reports OLS estimates of model (12). Columns 1 and 2 are estimated on the entire original sample of 707 firms in the time interval 1981-2001. Columns 3, 4 and 5 restrict the analysis to only those firms with a nonzero connection ($g_{(ij)t} = 0$) in any year t; firms meeting this requirement are included also in years when connections are absent. All the 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 $\rho = 0.8$ (10 communities) in column 4 and $\rho = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\rho = 0.6$ (small sample corrections are applied). All the observations of an individual firm in multiple years enter the same cluster.

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 ("Geospills"). Their inclusion does not dramatically impact the point estimate $\hat{\delta}$, which falls to 0.014 while remaining statistically significant. The geographic control, on the other hand, seems to have very little economic significance. In column (3) I restrict the sample only to those firms that enter the network at any point in time, even absent any connection in a specific year. This is an attempt 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. This exercise has an interesting implication: while the estimate $\hat{\delta}$ is again not largely affected (if decreases slightly to 0.012), the coefficient for the Jaffe measure of spillovers falls sharply and becomes not statistically significant. Since 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.[29]

In columns (4) and (5) I also include an additional set of dummy variables, in a first attempt to control for the fact that connected firms may be subjected 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 ($\rho = 0.8$); while in column (5) the additional dummy variables are based on the same 20 communities also used for clustering standard errors ($\rho = 0.6$). Increasing the number of clusters does not result in a dramatic variation of the point estimate $\hat{\delta}$ (in column (4), it actually increases). 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

I now illustrate the empirical results from the application of the IV strategy that addresses the problem of correlated confounders. I instrument the R&D stock of one firm's direct connections by aggregating the R&D of its indirect "friends" of second and third degree. In light of the analytical framework and the spatial autocorrelation of R&D in the network evidenced by Figure 6, both instruments could in principle be valid. However, the further instrument – the one based on third degree indirect connections – is the one more that is likely to be uncorrelated with both unobserved factors and the other input variables of firm i.

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

[29]This fact may also be interpreted as a censoring problem. COMPUSTAT only reports data for public firms. Small firms that go 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.

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

	(1)	(2)	(3)	(4)	(5)
2-degree Instrument	1.0645	1.1283			
	(0.0354)	(0.0457)			
3-degree Instrument		-0.5964	2.2044	2.0832	2.0310
		(0.1709)	(0.1456)	(0.1609)	(0.1649)
Private R&D	0.0797	0.0952	0.3789	0.3096	0.2794
	(0.0488)	(0.0476)	(0.1293)	(0.0989)	(0.1053)
Capital	0.0974	0.1092	0.5275	0.5187	0.4374
	(0.1098)	(0.1023)	(0.3038)	(0.2977)	(0.3125)
Labor	-0.1342	-0.1096	-0.7098	-0.6416	-0.6279
	(0.1309)	(0.1236)	(0.2644)	(0.2656)	(0.2854)
Jaffe Tech. Proximity	0.7567	0.7424	3.0520	3.1767	2.8117
	(0.3489)	(0.3300)	(1.3446)	(1.3204)	(1.2022)
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	1044	683	125	76	59
No. of Observations	7336	7336	7336	7336	7336

Notes: The table reports OLS regressions of the spillover variable $\sum_{j\neq i} g_{(ij)t} \log S_{jt}$ on appropriate instruments and all the controls included in the regressions from Table 2 (first stage regressions). The sample is restricted to firms with a nonzero connection ($g_{(ij)t} = 0$) in any year t. Columns 1 and 2 include the "second degree" instrument on the right hand side, while columns 2 through 5 include the "third degree" instrument. All the 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 $\rho = 0.8$ (10 communities) in column 4 and $\rho = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\rho = 0.6$ (small sample corrections are applied). All the observations of an individual firm in multiple years enter the same cluster.

ever enter the network. I regress the connections-induced spillovers variable on the aggregated log R&D stock of indirect connections of either second degree (column 1), second *and* third degree (column 2), third degree only (column 3). The estimates from columns (4) and (5) are analogous to those in column, but they additionally include two different sets of community-by-year fixed effects (respectively based on 10 and 20 communities, in analogy with Table 2). I include all the controls from equation (12) in all specifications. As expected, both instruments are strongly, positively correlated with the endogenous spillover variable.[30] The F-statistic from all first stage estimates is reassuringly strong: the lowest measured F-statistic – the one from column (5) – equals 59, and it is associated to a t-Statistic of the third-degree instrument which is larger than 12. Conventional inputs seem to have little residual predictive power relative to the spillover variable, unlike private R&D and the Jaffe measure.

Table 4 shows 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 estimate of parameter δ, since all other variables included in the model are estimated similarly as in the OLS baseline. By instrumenting the spillover variable with the R&D of indirect friends of second degree (column 1), δ is estimated around 0.0116, a figure substantially identical to the one obtained from OLS estimates. When including both instruments (column 2) the result is similar: $\hat{\delta} = 0.0114$.[31] By only instrumenting for the third-degree indirect connections instead (column 3), the result is different: the point estimate of δ is substantially higher, hovering around 0.0155. Interestingly, the inclusion of community-by-year effects results in even larger values: $\hat{\delta} = 0.0192$ with 10 communities (column 4) and $\hat{\delta} = 0.0172$ with 20 communities (column 5). All estimates of δ are statistically significant at the 5% level.[32]

These results are telling in two respects. First, they evidence a negative bias in simple OLS estimates. In light of the discussion of Corollary 1, this can be due to the circumstance whereby the prevalent factors driving the cross-correlation of R&D in the network are related to R&D costs. This fact would generate a mechanical negative correlation between one firm's productivity shock ω_{it} and the spillovers variable. Second, that a change in the point estimate of δ in only apparent when instrumenting spillovers only with third-degree indirect connections suggests that the second-degree instrument instrument might be itself correlated with ω_{it}. Given Proposition 2, this is consistent with the hypothesis that the spatial correlation of R&D is driven by exogenous factors: $(C, L) = (2, 0)$ – as opposed to the endogenous reflection of shocks.

[30] When the two instruments are included together, the coefficient for the third-degree instrument is negative and statistically different from zero. Nevertheless, in magnitude it is remarkably closer to zero, as it should be expected from conditioning on the R&D of "second-degree" firms.

[31] The Hansen J overidentification test has a p-value of about 0.45, indicating that two instruments effectively capture different sources of variation. This is consistent with the hypotheses on the network structure of common dependence that have been outlined in Section 1.

[32] The three estimates of δ in columns (1), (2) and (4) are all significant at the 1% level. The p-value for the estimate in column 3 is 2.2%, the one for the estimate in column 5 is 1.16%. As expected, the standard errors are substantially larger – about double in magnitude – when instrumenting spillovers only with the R&D of third-degree indirect connections.

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

	(1)	(2)	(3)	(4)	(5)
Private R&D (γ)	0.0398	0.0399	0.0377	0.0358	0.0364
	(0.0087)	(0.0087)	(0.0098)	(0.0111)	(0.0122)
Spillovers (δ)	0.0116	0.0114	0.0155	0.0192	0.0172
	(0.0033)	(0.0032)	(0.0062)	(0.0065)	(0.0065)
Geospills	0.0003	0.0003	0.0002	0.0002	0.0002
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
Capital	0.1744	0.1746	0.1716	0.1672	0.1641
	(0.0214)	(0.0214)	(0.0210)	(0.0228)	(0.0222)
Labor	0.6353	0.6351	0.6382	0.6391	0.6362
	(0.0282)	(0.0283)	(0.0281)	(0.0307)	(0.0293)
Jaffe Tech. Proximity	0.0852	0.0859	0.0708	0.0845	0.0482
	(0.0800)	(0.0796)	(0.0888)	(0.1028)	(0.1142)
2nd degree IV	YES	YES	NO	NO	NO
3rd degree IV	NO	YES	YES	YES	YES
Hansen J-statistic		0.556			
(p-value)		(0.456)			
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	7336	7336	7336	7336	7336

Notes: The table reports 2SLS estimates of model (12). The sample is restricted to firms with a nonzero connection ($g_{(ij)t} = 0$) in any year t. Models in columns 1 and 2 employ the second degree instrument; models in columns 2 through 5 employ the third degree one. All the 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 $\rho = 0.8$ (10 communities) in column 4 and $\rho = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\rho = 0.6$ (small sample corrections are applied). All the observations of an individual firm in multiple years enter the same cluster.

4.3 Market Value

The results for the market value model (17) are displayed in table 5. In column (1) and (2) I estimate the model via OLS, respectively on the whole sample and on the network subsample. In columns (3), (4) and (5) I show results from 2SLS estimates performed on the subsample, employing the 2nd degree instrument, both instruments and just the 3rd degree instrument, respectively. The estimates for the spillover parameter lie in an interval around 0.03. Unlike the case of the production function, whether OLS estimates are biased is less clear for the market value. In fact, 2SLS estimates that include the second degree instrument point to a higher value of $\tilde{\delta}$, while using only the third degree instrument yields an estimate that is closer to the OLS baseline and not statistically significant (with p-value 11.5%).

Table 5: Market Value, 1981-2001

	(1)	(2)	(3)	(4)	(5)
R&D Stock / Capital $(t-1)$	0.3681	0.4407	0.4347	0.4344	0.4400
	(0.2746)	(0.3944)	(0.3933)	(0.3935)	(0.3896)
Spillovers ($\tilde{\delta}$)	0.0349	0.0268	0.0368	0.0372	0.0280
	(0.0085)	(0.0072)	(0.0097)	(0.0097)	(0.0170)
Geospills	-0.0001	0.0000	-0.0001	-0.0001	0.0000
	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)
Jaffe Tech. Proximity	0.0720	-0.1638	-0.2065	-0.2083	-0.1687
	(0.1541)	(0.2373)	(0.2245)	(0.2244)	(0.2353)
BSV Business Stealing	-0.0218	0.1347	0.1324	0.1323	0.1344
	(0.0641)	(0.1054)	(0.1059)	(0.1059)	(0.1050)
Industry Level Sales	0.1887	0.1886	0.1888	0.1888	0.1886
	(0.0467)	(0.0606)	(0.0589)	(0.0588)	(0.0604)
Instrument(s)	OLS	OLS	2nd deg.	Both	3rd deg.
Fixed Effects	YES	YES	YES	YES	YES
Only Network	NO	YES	YES	YES	YES
No. of Observations	11816	7226	7226	7226	7226

Notes: The table reports various estimates of model (17). Except for the the results in column 1, the sample is restricted to to firms with a nonzero connection ($g_{(ij)t} = 0$) in any year t for all other estimates. Columns 1 and 2 report OLS estimates, while columns 3, 4 and 5 report 2SLS estimates using various combinations of exogenous instruments: the second degree instrument (3), both the second and the third degree instruments (4), and only the third degree instrument (5). All the estimates include firm and year fixed effects. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\rho = 0.6$ (small sample corrections are applied). All the observations of an individual firm in multiple years enter the same cluster.

4.4 Patent Count

The results for the patent count model (18) are reported in table 6, which is organized along the lines of table 5. Specifically, column (1) reports the results from the entire sample; column (2) those restricted to the network subsample, while the results from the control function approach, that are based on the usual sequence of instrument combinations, are given in columns (3), (4) and (5). The coefficient for the spillover-connections parameter is estimated in an interval around 0.03 in columns (1) through (4). The estimates from column (5) however, obtained via a control function approach that only exploits the third-degree instrument, register a much larger point estimate for $\breve{\delta}$, this time equal to 0.0865. This is again interpreted as the elasticity of patent output relative to an increase of all connections' R&D, which is approximately equal to 0.038 for a firm with average row-sum of connections. It is worth noticing how the Jaffe measure loses again all its economic and statistical significance once the analysis is restricted to only firms in the network.

Table 6: Patent Count, 1981-2001

	(1)	(2)	(3)	(4)	(5)
Private R&D ($\breve{\gamma}$)	0.0831	0.0793	0.0774	0.0803	0.0563
	(0.0313)	(0.0350)	(0.0357)	(0.0355)	(0.0385)
Spillovers ($\breve{\delta}$)	0.0295	0.0267	0.0372	0.0310	0.0865
	(0.0088)	(0.0064)	(0.0087)	(0.0082)	(0.0163)
Geospills	0.0008	0.0012	0.0011	0.0012	-0.0001
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0006)
Patents ($t-1$)	0.3956	0.4164	0.4150	0.4172	0.4108
	(0.0197)	(0.0178)	(0.0167)	(0.0169)	(0.0169)
Jaffe Tech. Proximity	0.3044	0.0309	0.0281	0.0282	0.0260
	(0.0820)	(0.0672)	(0.0660)	(0.0661)	(0.0663)
Industry Dummies	YES	YES	YES	YES	YES
Only Network	NO	YES	YES	YES	YES
Control Function	NO	NO	YES	YES	YES
			2nd deg.	Both	3rd deg.
No. of Observations	11444	6704	6704	6704	6704

Notes: The table reports maximum likelihood estimates of model (18). Except for the results in column 1, the sample is restricted to firms with a nonzero connection ($g_{(ij)t} = 0$) in any year t for all other estimates. Columns 3, 4 and 5 include additional regressors corresponding to the predicted residuals of "control function" regressions. Specifically, $\sum_{j\neq i} g_{(ij)t} \log S_{jt}$ is regressed on both the other controls R_{iq} and: the second degree instrument (column 3), both the second and the third degree instruments (column 4), and only the third degree instrument (column 5). All the estimates include 4-digits industry and year fixed effects. Standard errors are clustered by the 20 "communities" obtained via the Louvain algorithm with $\rho = 0.6$ (small sample corrections are applied). All the observations of an individual firm in multiple years enter the same cluster.

The economic interpretation of the results from column (5) is analogous to the case of the production function estimates. The sizable increase in the point estimate of $\breve{\delta}$ is due to the correction of the simultaneity problem induced by correlated factors driving both the spatial correlation of R&D and firms' propensity to patent. If these confounders are predominantly cost factors, in presence of complementarities they would negatively correlate with the effort made by individual firms towards the realization of new patents. An intriguing, alternative hypothesis is that simultaneous increases in R&D spending correlate to worse innovation on the quality margin (recall that the patent outcome measure weighs patents by citations). A typical narrative about some industries, like the pharmaceutical sector, associates lower quality patents to an increase in the total number of USPTO registered inventions.

4.5 Discussion

A way to quantify the economic relevance of the estimated spillover effects is to calculate the average Marginal Private Returns (MPR) and Marginal Social Returns (MSR) of R&D (see e.g. BSV). I define the MPR as the average increase in output relative to an increase in the R&D stock of the individual firm (MPR $= \frac{1}{N}\sum_{i=1}^{N} dY_i/dS_i$), while the MSR is the average increase in output relative to the average increase in the R&D stock of all the other firms (MPR $= \frac{1}{N}\sum_{i=1}^{N} dY_i/\overline{dS}$). These are easily calculated under the hypothesis of an homogeneous percentage increase in R&D by all firms ($dS_i/S_i = dS/S$ for all i). In this case, the average response of output dY/dS can be derived from (12) and decomposed as follows.

$$\frac{1}{N}\sum_{i=1}^{N}\frac{dY_i}{dS} = \underbrace{\hat{\gamma}\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{S_i}}_{=\text{MPR}} + \underbrace{\hat{\delta}\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N} g_{ij}\frac{Y_i}{S_i}}_{=\text{MSR}} \tag{19}$$

To evaluate the MPR and the MSR, I use the estimates for $\hat{\gamma}$ and $\hat{\delta}$ from column (5) of Table 4, as well as the values $\frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{S_i} = 11.36$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}^{N} g_{ij}\frac{Y_i}{S_i} = 5.66$ calculated on the pooled panel. As a result, the MPR is approximately equal to 41.3% ($= 11.36 \times 3.64\%$) and the MSR to about 9.7% ($= 5.66 \times 1.72\%$):[33] the latter

[33]Notice 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. Using the typical assumption of a 0.20 steady-state ratio, one gets an approximate 8.26% private return

is approximately equal to 24% of the former. While not as large as the evaluations from other studies, these are realistic and economically significant values. Notice that these calculations do not take into account the "amplification effect" due to strategic response and complementarities, something that should be taken into account when evaluating, say, the effect of an R&D-stimulating policy.

The empirical strategy adopted in this paper seeks to solved the problem of common confounders. There are, however, some additional empirical concerns that remain only partially addressed. The first one is ultimately an issue of measurement error. Ignore for the moment the problem of measuring the innovation effort and the knowledge stock of connected firms with cumulated R&D expenditures, as well as the additional difficulty that the connection metrics may not fully capture the degree of interactions between R&D teams. Even in absence of these issues, there is a problem of *network sampling* that ultimately depends on the inability to observe nodes and links, even if their intensity were perfectly measurable (Chandrasekhar and Lewis, 2011). In the context of this paper, given the type of sample selection implied by COMPUSTAT data (small and private firms are excluded), this would result in an underestimation of the actual effect of connections. A proper assessment of this problem requires the application of proper sampling strategies to high quality firm-level data that can be matched to patents: a non-trivial set of requirements.

The second problem is the endogeneity of connections. Specifically, factors that affect firm-level outcomes might be correlated with the network topology or with its changes over time. For instance – as I have briefly discussed in Section 3 – inventor flows between firms might be correlated with the unobserved productivity shocks, causing a bias of undetermined sign. A similar issue may also affect the estimates of the market value and of the patent outcome equation. To deal with this problem, I have already taken two measures. First, I have narrowed down the analysis to only those firms that enter the network at any point in time. Second, I have constructed both instrumental variables by taking first or further lags of the network structure, by exploiting its dynamic properties. Both approaches do not greatly impact the point estimates associated with the spillovers variable. However, these procedures might not account for more complex patterns of serial and spatial dependence between the network topology and the unobserved errors. The analysis of this problem is tightly connected to theories of network formation, and it deserves a separate study.

---

and an approximate 1.94% social return from yearly R&D expenditures.

5 Conclusion

In this paper I propose a new method of evaluating R&D spillovers. By aggregating information on patent coauthorship relationships between individuals that work for different organizations, I construct a network of firms that are reciprocally connected through their R&D teams. I evaluate the dependence of firm productivity, market value and innovation rate from the R&D performed by firms connected in the network, weighted by the intensity of mutual links. Concerned by the possibility of common confounders that simultaneously drive R&D choices and firm-level outcomes, I employ an 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 links. Under conditions specified by a formal model 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 sizeable 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 striking conformity with the prediction of the analytical model, the application of the identification strategy that I propose shows that when instrumenting peers' R&D with the R&D of sufficiently distant firms, point estimates of spillover effects on both productivity and patent output increase substantially. This suggests that common factors driving both R&D and firm outcomes might do so in opposite directions. In particular, this finding may reflect the stylized fact that over the industry life cycle faster innovation is typically associated with higher R&D costs. I use the estimates of spillovers obtained from the proposed methodology in order to evaluate the relative importance of the marginal social returns to R&D relative to the private returns, finding that the former are about 24% of the latter.

References

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

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., 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.

Chandrasekhar, A. and R. Lewis (2011). Econometrics of sampled networks. Mimeo.

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.

Gort, M. and S. Klepper (1982). Time Paths in the Diffusion of Product Innovations. *The Economic Journal 92*(367), 630–653.

Graham, B. (2014). An empirical model of network formation: detecting homophily when agents are heterogeneous. Mimeo.

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*, 29–47. Supplement.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Paper.

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.

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.

Klepper, S. (1996). Entry, Exit, Growth, and Innovation over the Product Life Cycle. *The American Economic Review 86*(3), 562–583.

König, M., X. Liu, and Y. Zenou (2016). R&d networks: Theory, empirics and policy implications. Mimeo.

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

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 (2010). Spillovers in space: does geography matter? Forthcoming in *The Journal of Industrial Economics*.

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 ed.).

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.

Serafinelli, M. (2013). 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.

Appendix A Analytical Model: Proofs

Lemma 1. *Suppose that given a mapping* $g_{ij} : \mathbb{I} \times \mathbb{I} \to [0,1]$ *with* $\dim |\mathbb{I}| = N$, $i, j = 1, \ldots, N$, *and for* $0 < \delta < 1$, *it holds*

$$0 \leq \delta \sum_{j=1}^{N} g_{ij} < 1$$

for every $i = 1, \ldots, N$. *Then, there always exists a vector* $\mathbf{p}_i = (p_{i1}, \ldots, p_{iN}) \geq \iota$ *for every* $i = 1, \ldots, N$ *such that*

$$\sum_{j=1}^{N} \frac{1}{p_{ij}} = 1$$

and

$$g_{ij} p_{ij} \delta \leq 1$$

for every $j = 1, \ldots, N$ *and every* $i = 1, \ldots, N$.

Proof. The proof is constructive. Consider first the case when $0 < g_{ij}\delta < 1$ for every $j = 1, \ldots, N$ and every $i = 1, \ldots, N$. For any i-th element of $\mathbb{I}$ construct the vector $\mathbf{p}'_i = (p'_{i1}, \ldots, p'_{iN}) > \iota$ whose elements are defined as

$$p'_{ij} = (g_{ij}\delta)^{-1}$$

for every $j = 1, \ldots, N$. Inverting and summing over j one obtains

$$0 < \sum_{j=1}^{N} \frac{1}{p'_{ij}} = \delta \sum_{j=1}^{N} g_{ij} < 1$$

hence, a vector $\mathbf{p}_i$ with the desired properties can be obtained by appropriately *decreasing* any combination of the elements of $\mathbf{p}'_i$. Now consider the circumstance in which for some i-th element of $\mathbb{I}$, some values of the g mapping are equal to 0. In this case, modify the rule defining $\mathbf{p}'_i$ as

$$p'_{ij} = \begin{cases} (g_{ij}\delta)^{-1} & \text{if } 0 < g_{ij}\delta < 1 \\ P_{ij} & \text{if } g_{ij}\delta = 0 \end{cases}$$

where values $P_{ij} > 1$ can always be chosen to be sufficiently large for

$$\sum_{j=1}^{N} \frac{1}{p'_{ij}} \leq 1$$

to hold. Then adjust the values of $\mathbf{p}'_i$ in order to obtain vector $\mathbf{p}_i$, as above.

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

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

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 predict them accurately. Consider that the $Q+1$ First Order Conditions relative to S_i and $(X_{i1}, \ldots, X_{iQ})$ are sufficient to characterize a maximum, since the problem is concave in all its choice variables. The FOCs are, respectively:

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

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

with (A.2) taken for $c = 1, \ldots, Q$. 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 $q = 1, \ldots, Q$. 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:

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

where

$$\mu \equiv \log A + \log \gamma + \sum_{q=1}^{Q} \beta_q \left(\log \beta_q - \log \xi_q - \log \gamma \right)$$

and $\vartheta \equiv \left(1 - \gamma - \sum_{q=1}^{Q} \beta_q\right)^{-1} > 1$. Notice that (A.4) is a mapping from the space

$\mathbf{\Omega}$ of all information sets available to players onto the set of positive real numbers, which I define as $S_i^*: \mathbf{\Omega} \rightarrow \mathbb{R}_{++}$. Clearly, a fixed point of the vector-valued function $\mathrm{S}^* = (S_1^*, \ldots, S_N^*)$ is a Bayes-Nash equilibrium of the game. Notice that there is also a one-to-one relationship between $\mathrm{S}^* = (S_1^*, \ldots, S_N^*)$ and the associated 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 $\mathfrak{L}$, and endow it of the max-norm $\|\log \mathrm{S}^*\| = \max_i \|\log S_i^*\|_\infty$. Define the operator $T: \mathfrak{L} \rightarrow \mathfrak{L}$ as:

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

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. Now consider any vector $\mathbf{p}_i$ satisfying the conditions expressed in Lemma 1:

$$\begin{aligned}
\log \mathbb{E}\left[\left(\prod_{j=1}^{N} S_j^{*g_{ij}} \right)^{\delta} \middle| \Omega_i \right] &\leq \log \prod_{j=1}^{N} \mathbb{E}\left[\left| S_j^{*(g_{ij}\delta)} \right|^{p_{ij}} \middle| \Omega_i \right]^{\frac{1}{p_{ij}}} \\
&= \sum_{j=1}^{N} \frac{1}{p_{ij}} \log \mathbb{E}\left[S_j^{*(g_{ij}p_{ij}\delta)} \middle| \Omega_i \right] \\
&\leq \sum_{j=1}^{N} \frac{1}{p_{ij}} \log \mathbb{E}\left[S_j^* \middle| \Omega_i \right]^{g_{ij}p_{ij}\delta} \\
&= \delta \sum_{j=1}^{N} g_{ij} \log \mathbb{E}\left[S_j^* \middle| \Omega_i \right]
\end{aligned}$$

where the first line is an application of Hölder's inequality, the second exploits the fact that function S_j^* only takes positive values, and finally the third one is an application of Jensen's inequality. It is easy to show that the operator defined as

$$D\left(S_i^*\right) = \delta \sum_{j=1}^{N} g_{ij} \log \mathbb{E}\left[S_j^* \middle| \Omega_i \right] - \log \mathbb{E}\left[\left(\prod_{j=1}^{N} S_j^{*g_{ij}} \right)^{\delta} \middle| \Omega_i \right] \geq 0$$

is monotone, hence for any two $(\log \mathrm{S}^*, \log \mathrm{Z}^*) \in \mathfrak{L}^2$ it holds that

$$\left| \log \mathbb{E}\left[\left(\prod_{j=1}^{N} S_j^{*g_{ij}} \right)^{\delta} \middle| \Omega_i \right] - \log \mathbb{E}\left[\left(\prod_{j=1}^{N} Z_j^{*g_{ij}} \right)^{\delta} \middle| \Omega_i \right] \right| \leq \left| \delta \sum_{j=1}^{N} g_{ij} \log \frac{\mathbb{E}\left[S_j^* \middle| \Omega_i \right]}{\mathbb{E}\left[Z_j^* \middle| \Omega_i \right]} \right|$$

implying that

$$\begin{aligned}
\|T\left(\log \mathrm{S}^*\right), T(\log \mathrm{Z}^*)\| &= \max_i \left|T_i\left(\log \mathrm{S}^*\right) - T_i(\log \mathrm{Z}^*)\right| \\
&\leq \max_i \left|\vartheta\delta \sum_{j=1}^{N} g_{ij} \log \frac{\mathbb{E}\left[\left. S_j^* \right| \Omega_i\right]}{\mathbb{E}\left[\left. Z_j^* \right| \Omega_i\right]}\right| \\
&\leq \vartheta\delta \max_i \left(\sum_{j=1}^{N} g_{ij}\right) \max_i \left|\log S_i^* - \log Z_i^*\right| \\
&= \left\|\vartheta\delta \max_i \left(\sum_{j=1}^{N} g_{ij}\right)\right\| \left\|\log S_i^* - \log Z_i^*\right\|
\end{aligned}$$

hence, T is a contraction with Lipschitz constant

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

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 shown that the equilibrium R&D S_i^* can be expressed as in (4). To this end, rewrite the latter equation as $\log S_i^* = \mu\vartheta b_i + g_i^*\left(\Omega_i, \mathcal{G}\right)$ for some generic $b_i > 0$, and substitute it into (A.4) for all $j \neq i$, thus obtaining:

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

taking logarithms and rearranging terms this becomes

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

it easy to see that this expression conforms to the definition of the contraction operator T, and that $g_i^*\left(\Omega_i, \mathcal{G}\right)$ has the form given in the text (to derive its bound, it is sufficient to apply Lemma 1 and Hölder's inequality). For (A.5) to be consistent also with (4), it must be shown that

$$1 + \theta \sum_{j=1}^{N} g_{ij} b_j = b_i = b_i\left(\mathcal{G}\right)$$

for every firm $i = 1, \ldots, N$ (recall that $\theta \equiv \delta\vartheta$). Rewrite the first equality above in matrix form:

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

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

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

where:

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

notice that the series converges for $\theta < 1$, as stated in the text. The solution is exactly the vector of Bonacich-Katz centrality measures with attenuation parameter θ.

Proof of Corollary 1. Corollary 1 admits the possibility that the set of inequalities (6) might hold for some pair of connected firms i and j under some specific restrictions of the model; it does not state that these inequalities must hold for all pairs of firms under all circumstances. Therefore, in order to "prove" this result it is sufficient to show that (6) hold in a particular example. To this end, consider the case where $\mathcal{F}(\cdot)$ is a multivariate normal distribution such that

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

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 $i = 1, \ldots, N$. In this specific case it particularly easy to see – although the result is more general – that equilibrium R&D is a linear function of private shocks: $g_i^*\left(\Omega_i, \mathcal{G}\right) = f_i^* + g_i^*\left(\omega_i - \lambda \varpi_i\right)$ for $f_i^*, g_i^* > 0$ and $\lambda \equiv 1 - \sum_{q=1}^{Q} \beta_q \in (0, 1)$. Now, denote $V_\ell \equiv \mathbb{V}\mathrm{ar}\left(\omega_\ell\right)$ and $\psi_\ell \equiv \mathbb{C}\mathrm{ov}\left(\omega_\ell, \varpi_\ell\right)$ for $\ell = i, j$, and assume further that

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left(\varpi_i\right) &= \psi_i^2 V_i \\ \mathbb{V}\mathrm{ar}\left(\varpi_j\right) &= \psi_j^2 V_j \\ \mathbb{C}\mathrm{orr}\left(\omega_i, \varpi_j\right) &= g_{ij} \\ \mathbb{C}\mathrm{orr}\left(\omega_j, \varpi_i\right) &= g_{ij} \end{aligned}$$

where

$$\mathbb{V}\mathrm{ar}\begin{pmatrix} \omega_i \\ \omega_j \\ \varpi_i \\ \varpi_j \end{pmatrix} = \begin{pmatrix} V_i & \cdots & \cdots & \cdots \\ g_{ij}\sqrt{V_i V_j} & V_j & \cdots & \cdots \\ \psi_i & g_{ij}\psi_i\sqrt{V_i V_j} & \psi_i^2 V_i & \cdots \\ g_{ij}\psi_j\sqrt{V_i V_j} & \psi_j & g_{ij}\psi_i\psi_j\sqrt{V_i V_j} & \psi_j^2 V_j \end{pmatrix}$$

is a positive semidefinite matrix for $V_\ell \geq 1$ and $0 \leq g_{ij} \leq \sqrt{\frac{1}{2}\left(1 + V_\ell^{-1}\right)}$ with $\ell = i, j$

(notice that these are always realistic values of g_{ij} given the descriptives in Table 2), hence this is a legitimate characterization of the random vector $(\omega_i, \omega_j, \varpi_i, \varpi_j)$. Under all these hypotheses, the inequalities in (6) can be expressed after some calculation as:

$$\begin{gathered}\mathbb{C}\text{ov}\left(g_i^*\left(\omega_j - \lambda\varpi_j\right), g_j^*\left(\omega_j - \lambda\varpi_j\right)\right) \propto \left(\lambda\psi_i - 1\right)\left(\lambda\psi_j - 1\right) > 0 \\ \mathbb{C}\text{ov}\left(\omega_i, g_j^*\left(\omega_j - \lambda\varpi_j\right)\right) \propto 1 - \lambda\psi_j < 0 \\ \mathbb{C}\text{ov}\left(\omega_j, g_i^*\left(\omega_i - \lambda\varpi_i\right)\right) \propto 1 - \lambda\psi_i < 0\end{gathered}$$

which hold simultaneously as long as $\psi_i, \psi_j > \lambda^{-1}$. While this example is stylized, it illustrates quite well that in order for (6) to hold both $\mathbb{V}\text{ar}\left(\varpi_i\right)$ and $\mathbb{C}\text{ov}\left(\omega_i, \varpi_j\right)$ must be sufficiently large relative to $\mathbb{V}\text{ar}\left(\omega_i\right)$, and vice versa.

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. Order the intermediate connections along the chosen path: $\ell = 0, \dots, D$ where (without loss of generality) $i = 0$ and $j = D$. By Assumption 3 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 $Q \in \Omega_j$ has length $D - L$. Since $D - L > C$, $\{\omega_i, \varpi_i\}$ and all the elements of Ω_j are orthogonal by Assumptions 1-2, implying $\mathbb{C}\text{ov}\left(\omega_i, \log S_j^*\right) = 0$ because of equation (4). If this is true for the shortest path connecting i and j, so it is for any other path, thereby establishing (7). 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 $P \in \Omega_i$ with another element $Q \in \Omega_j$ has length $D - 2L > C$. Consequently, firms i and j are in equilibrium functions of mutually independent sets of random variables, which implies (8) and completes the proof.

Proof of Corollary 2. 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 $q = 1, \dots, Q$, which can be rewritten as

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

for some firm-specific function of the network topology $x_i^*\left(\mathcal{G}\right)$. The stochastic properties of equilibrium inputs are driven by the term $g_i^*\left(\Omega_i; \mathcal{G}\right) - \varpi_i$; but since ϖ_i is always listed in Ω_i by Assumption 3, $\log X_{iq}^*$ 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 Dynamic Model (Sketched)

In this appendix I sketch a two-period version of the model where firms accumulate R&D stocks by making yearly investments in R&D (flows). The objective is to show 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 might be conditional on long-term projects. The main reason is probably related to the very nature of R&D activity: highly risky, characterized by large fixed costs and requiring many years to yield (potentially high) rewards. Hence, it may be optimal for firms to commit in advance to long-term plans.

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

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

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 $Z_{it}\in\mathbb{R}_{++}$ is the R&D investment (flow) in period t, $\zeta\in[0,1]$ is the depreciation parameter, while $\phi\in[0,1]$ is the discount factor. The R&D stock for $t=1$ is identical to the flow: $S_{i1}=Z_{i1}$. For $t=2$ instead, it is given by the current investment plus the past depreciated flow: $S_{i2}=\zeta Z_{i1}+Z_{i2}$. Notice 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 Z_{i1} and Z_{i2} for both periods. Now the Bayes-Nash equilibrium is technically expressed as fixed point of $(\mathrm{Z}_1,\mathrm{Z}_2)=(Z_{11},\dots,Z_{N1},Z_{21},\dots,Z_{N2})$. 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:

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

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

therefore, 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 Corollary 2, 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.

Notice 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 Z_{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:

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

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

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:

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

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 work, 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 6. 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 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 Data

The main panel of firms has been reconstructed by Bloom et al. (2013) (BSV) by selecting firms from COMPUSTAT with at least one entry in the "Segment" complementary dataset. The latter breaks down sales by line of business for specific firms. The main variables employed in the estimation of the production function are constructed according to standard methodologies. In particular, monetary values are deflated using appropriate price indices. The stock of R&D is calculated from flows using the perpetual inventory method with a 15% depreciation parameter. I refer to the online appendix from BSV for the details.

Firm-level identifiers are matched to patents as per the NBER patent dataset developed until 2006; see Hall et al. (2001) for the details. 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. Notice 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, $Spilltech_{it} = \sum_j TECH_{ij} S_{jt}$. It enters logarithmically in the estimation of the Cobb-Douglas production function. To facilitate comparisons, I employ the same variables in my estimates.

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 coauthorship relationships; iii) the time interval in which each inventor is associated to a firm. I obtain information on i) and ii) thanks to the match of the patent identifiers from the USPTO across the NBER and the HPND datasets. I rely on the work performed by the authors of the HPND dataset for the quality of their disambiguation algorithm, see Li et al. (2014) for details. However, I have no direct information about iii). In order to associate individuals to firms, I use indirect information extrapolated from the patent data.

In particular, I can establish to which firm are assigned patents, that are signed by specific individual inventors. By defining the time interval in which every individual is observed to collaborate on patents for a specific firm, I can provide an approximate time interval that defines their mutual association. Define $\underline{p}_{im}$ as the first year when inventor i is observed patenting (application year) for firm m. Similarly, $\overline{p}_{im}$ is the last year. The assignment rule between the inventor and the firm in year t is

$$f_{(mi)t} = \begin{cases} 1 & \text{if } t \in \left[\underline{p}_{im} - 1, \overline{p}_{im} + 1\right] \\ 0 & \text{otherwise} \end{cases}$$

which is extended one year in the past relative to $\underline{p}_{im}$ and one year in the future relative to $\overline{p}_{im}$. This choice is based 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 may miss years in which inventors, while not producing patents, are still part of an organization. This would be relevant (and generate problems of measurement error) mostly if these idle inventors were connected to individuals in other firms. Furthermore, it is also arguable that idle inventors are not very active in the process of knowledge creation and exchange. Such a restricted time window essentially captures the size of the R&D-performing team of a firm, whether it is made of regular employees or, say, academic collaborators. It is reassuring that the results are very robust to perturbations in this assignment rule (such additional results are available upon request).

One can collect all the binary indicators $f_{(mi)t}$ in a matrix $\mathbf{F}_t$ which has N rows (number of firms in the data) and M_t columns (the number of inventors at time t). To calculate the connection measures, one should first obtain the binary and symmetric adjacency matrix $\mathbf{P}_t$ of coauthors at time t. It is a matrix of dimension $M_t \times M_t$ where $p_{(ij)t} = p_{(ji)t} = 1$ if the two inventors i and j have at least one joint patent at $t+1$. Define $\mathcal{B}(\cdot)$ as a boolean operator that applied to matrices, returns other matrices whose entries are equal to 1 for positive corresponding entries in the argument 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:

$$\mathbf{K}_t = \mathbf{F}_t \cdot \mathcal{B}\left(\mathbf{P}_t \mathbf{F}_t'\right) = \mathcal{B}\left(\mathbf{F}_t \mathbf{P}_t\right) \cdot \mathbf{F}_t'$$

and obtain the numerator of the expression within parentheses in (10) for every pair of firms as $k_{(ij)t} + k_{(ji)t}$. Notice that the diagonal elements of $\mathbf{K}_t$ denote 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.

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

Finally, it is worth mentioning how I compute the matrix $\mathbf{H}$ that collects the weights $h_{(ik)t}$ for the second-degree instrument (the case of the third-degree instrument is analogous):

$$\mathbf{H}_t = \mathbf{G}_t^2 \circ (\mathbf{1} - \mathbb{I}[\mathbf{I} + \mathbf{G}_t])$$

where the symbol $\circ$ denotes the Hadamard (pointwise) multiplication and the indicator function $\mathbb{I}[\cdot]$ is also taken pointwise on its matrix argument. To better see why, consider that for $i \neq k$, $h_{(ik)t} = \sum_{j\neq i} \left(g_{(ij)t} g_{(jk)t}\right) \mathbb{I}\left[g_{(ik)t} = 0\right]$. Therefore the expression for the instrument, that is the i-th entry of the column vector $\mathbf{H}_t \mathbf{s}_t$ (where $\mathbf{s}_t$ is the vector of log R&D) reads as

$$\begin{aligned}\sum_{k\neq i} h_{(ik)t} \log S_{kt} &= \sum_{k\neq i} \left(\sum_{j\neq i} \left(g_{(ij)t} g_{(jk)t}\right) \mathbb{I}\left[g_{(ik)t} = 0\right] \right) \log S_{kt} \\ &= \sum_{j\neq i} g_{(ij)t} \sum_{k\neq i,j} \underbrace{g_{(jk)t} \mathbb{I}\left[g_{(ik)t} = 0\right]}_{=\tilde{h}^i_{(jk)t}} \log S_{kt}\end{aligned}$$

which corresponds to the definition given in the text, as expected.

C.3 Geographic Control and Measures of Proximity

An empirical concern of the analysis is that patent coauthorship relationships may simply capture the fact that inventors live close to one another. Therefore, measures of connection might reflect the fact that firms have their R&D labs in the most innovative areas – something that might have a direct impact on innovation and productivity. To clear this concern I calculate a measure of R&D spillovers that is weighted against the relative spatial proximity of two R&D teams. These weights are called measures of *proximity* and they are conceptually similar to the connection measures. In lieu of a patent coauthorship relationship, however, two inventors are identified as being "linked" if they are "neighbors" in spatial terms, that is they are observed to patent from the same Core Based Statistical Area (CBSA) in a given year. I obtain this information from patent data, that report the ZIP code of the address of residence of each signing inventor. Proximity measures read as

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

and they are calculated with a procedure that is analogous to the one of connection measures. The actual control employed in the regressions is also analogous to the variable of connection-induced spillovers, and it is defined as $Geospills_{it} = \sum_j b_{(ij)t} \log S_{jt}$.

Appendix D Alternative Measures of Connection

In this appendix I discuss alternative connection measures $g^{alt}_{(ij)t}$ and the estimates obtained from their application to model (12). I focus in particular on four alternatives.

1. **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.

2. **Second Degree Connections**. I define "connected" inventors as not simply those individuals who are patent-coauthors of someone in the other firm, but also coauthors of coauthors of someone in the other firm (second-degree coauthors). I take the square root of the corresponding measure, which I call $g^{2dg}_{(ij)t}$. Relative to the baseline, this alternative measure downplays those connected scientists who do not develop many bonds *within* the firm they are assigned to (occasional inventors).

3. **Asymmetric "Receiving" Connections**. I abandon the framework of directed networks and I consider the possibility that spillover relationships are asymmetric between firms. In particular, I suppose that the degree of a firm's access to the knowledge of another depends only by its own share of connected inventors: $asr_{(ij)t} = k_{(ij)t}/k_{(ii)t}$. In the estimation, however, I use its square root $g^{asr}_{(ij)t} = \sqrt{asr_{(ij)t}}$. This measure gives more importance to smaller, well connected firms in the process of ideas exchange.

4. **Asymmetric "Spilling" Connections**. An alternative economic assumption is that spillovers do not depend on active acquisition of knowledge by well-connected firms, but rather by their passive access to naturally leaked information. In this case it would be more advantageous to have access to as many inventors as possible in the "spilling" firm. The connection measure is in this case defined as $g^{ass}_{(ij)t} = \sqrt{ass_{(ij)t}}$ with $ass_{(ij)t} = k_{(ji)t}/k_{(jj)t}$. This measure gives more relevance to firms that are well connected to larger ones.

For any of these measures $g^{alt}_{(ij)t}$ the spillover variable is constructed as $\sum_j g^{alt}_{(ij)t} \log S_{jt}$.

Table A.1 shows the results from the estimation of model (12) using these alternative connection metrics. For each measure, both the OLS and the 2SLS estimates are reported. All estimates are restricted to only the firms included in the network, they include the full set of fixed effects (including community-by-year fixed effects on 20 communities), and the 2SLS estimates are based on the third degree instrument. Point estimates vary in magnitude because of the rescaling implied by different measures. Noticeably, for the "Second Degree" and the "Asymmetric Spilling" measures, 2SLS estimates of δ are larger than OLS and significant at the 1% level. For the "Linear" and the "Asymmetric Receiving" measures instead the 2SLS estimates are smaller, and only in the Linear case is $\hat{\delta}_{2SLS}$ statistically significant (at the 10% level).

Table A.1: Alternative Connection Measures, Production Function, 1981-2001

	(Linear)		(2nd Degree)	
	OLS	2SLS	OLS	2SLS
Private R&D (γ)	0.0387	0.0396	0.0392	0.0360
	(0.0120)	(0.0117)	(0.0114)	(0.0123)
Spillovers (δ)	0.0805	0.0648	0.0041	0.0073
	(0.0163)	(0.0352)	(0.0013)	(0.0027)
Capital	0.1683	0.1694	0.1669	0.1616
	(0.0225)	(0.0223)	(0.0229)	(0.0224)
Labor	0.6304	0.6294	0.6343	0.6412
	(0.0293)	(0.0285)	(0.0301)	(0.0296)
Jaffe Tech. Proximity	0.0694	0.0768	0.0734	0.0474
	(0.1000)	(0.1064)	(0.1021)	(0.1133)
Fixed Effects	YES	YES	YES	YES
Only Network	YES	YES	YES	YES
Instrument	NO	3rd deg.	NO	3rd deg.
No. of Communities ($\times$ Year)	20	20	20	20
No. of Observations	7336	7336	7336	7336

	(As. Receiving)		(As. Spilling)	
	OLS	2SLS	OLS	2SLS
Private R&D (γ)	0.0406	0.0424	0.0406	0.0379
	(0.0120)	(0.0132)	(0.0120)	(0.0119)
Spillovers (δ)	0.0087	0.0032	0.0058	0.0114
	(0.0020)	(0.0062)	(0.0023)	(0.0031)
Capital	0.1695	0.1722	0.1698	0.1658
	(0.0221)	(0.0223)	(0.0224)	(0.0222)
Labor	0.6287	0.6265	0.6308	0.6361
	(0.0287)	(0.0278)	(0.0290)	(0.0291)
Jaffe Tech. Proximity	0.0754	0.0955	0.0784	0.0505
	(0.1042)	(0.1187)	(0.1034)	(0.1115)
Fixed Effects	YES	YES	YES	YES
Only Network	YES	YES	YES	YES
Instrument	NO	3rd deg.	NO	3rd deg.
No. of Communities ($\times$ Year)	20	20	20	20
No. of Observations	7336	7336	7336	7336

Table A.2: MPR and MSR: Comparative Prospect

	Baseline Measure	Alternative Measures Linear	2nd Degree	As. Spilling
Mean of the Row Sum $\bar{g}_{it}$	0.440	0.065	1.049	0.675
(Standard Deviation)	(0.178)	(0.084)	(1.422)	(1.127)
Marginal Private Return	41.3%	49.7%	45.2%	47.9%
Marginal Social Return	9.7%	9.8%	7.8%	2.5%

Table A.2 offers a comparative prospectus about the means and standard deviations of the row sum of connections for the "linear," "second degree" and "asymmetric spilling" measures, contrasted with the baseline case. In addition, I report the MPRs and MSRs calculated with the respective 2SLS estimates of $\hat{\gamma}$ and $\hat{\delta}$ from Table A.1. As expected, the calculated MPRs are similar across the four connection measures. The MSRs, by contrast, are similar only across the baseline, the Linear and the Second Degree measures. The MSR calculated from the Linear measure is actually slightly larger than in the baseline case, but it is obtained from a much less precise estimate of δ. By contrast, the MSR calculated for the Asymmetric Spilling measure is much smaller in magnitude, essentially because larger values of $g^{ass}_{(ij)t}$ are by construction associated with lower values of S_j as per (19).

Appendix E Graphical Description of the Network

This appendix collects, in the next few pages, 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 a positive function of the total strength of links that are summed over the years. In Figures from E.1 to E.5 nodes are also distinguished by various shades of the same orange color; in particular, darker colors indicate larger measures of network centrality associated with a node. The names of selected firms are apposed to some of the largest, most central nodes. A brief introduction or commentary for each of the following graphs is given in the list below.

- Figure E.1 displays the network in 1985.
- Figure E.2 displays the network in 1990.
- Figure E.3 displays the network in 1995.
- Figure E.4 displays the network in 2000.
- Figure E.5 displays the "pooled" network, which results from aggregating all edges (connections) over time for all nodes.
- Figure E.6 displays the communities obtained by applying the Louvain algorithm on the "pooled" network with maximum (1) resolution. The top hierarchy of communities is composed of six groups. The semiconductor/ICT, mechanical, biotech/pharmaceutical and chemical industries are clearly identifiable as separate communities; in addition there are two smaller, mixed groups whose nodes are dispersed across the graph.
- Figure E.7 displays the communities obtained by applying the Louvain algorithm with the $\rho = 0.6$ resolution. The resulting partition is the one used to cluster standard errors in all all empirical estimates featured in this work.

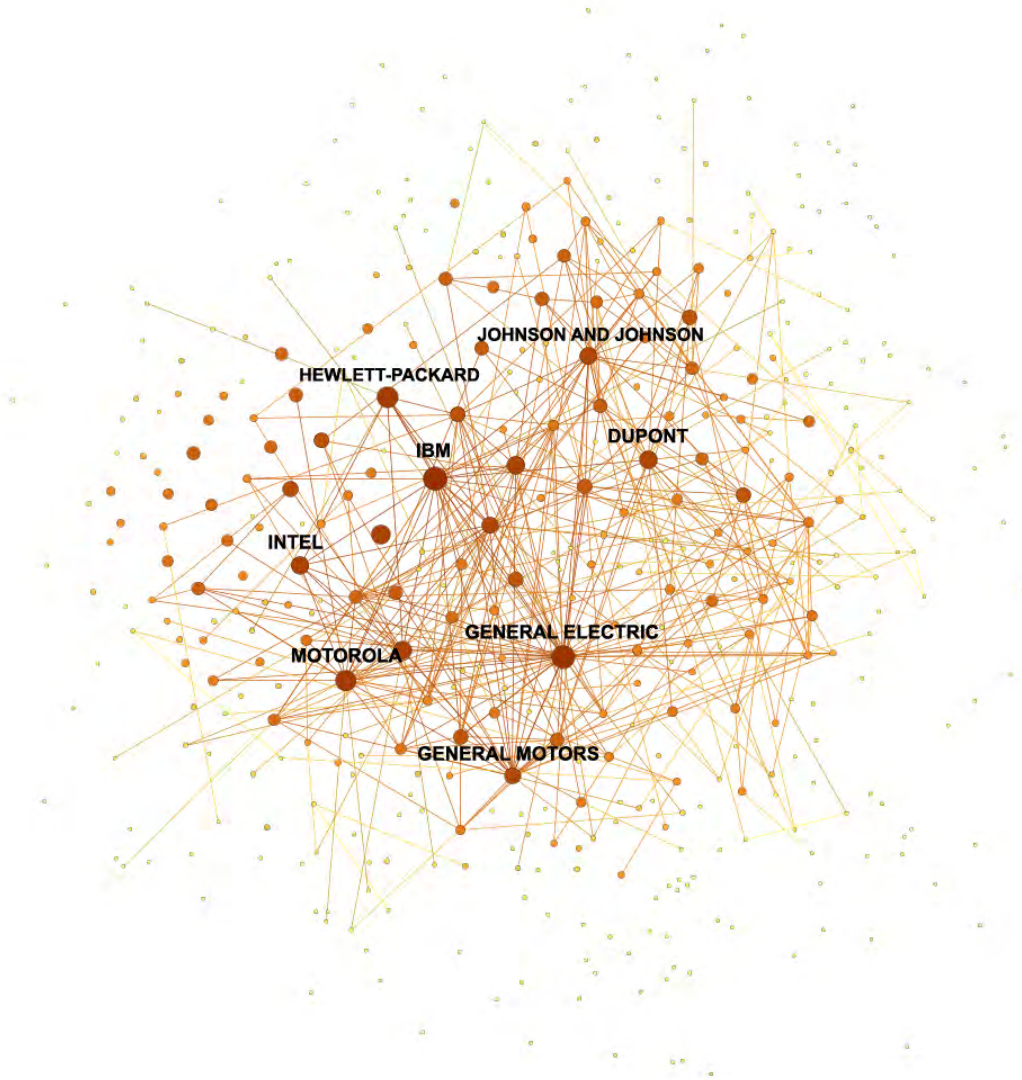


Figure E.1: The Network in 1985

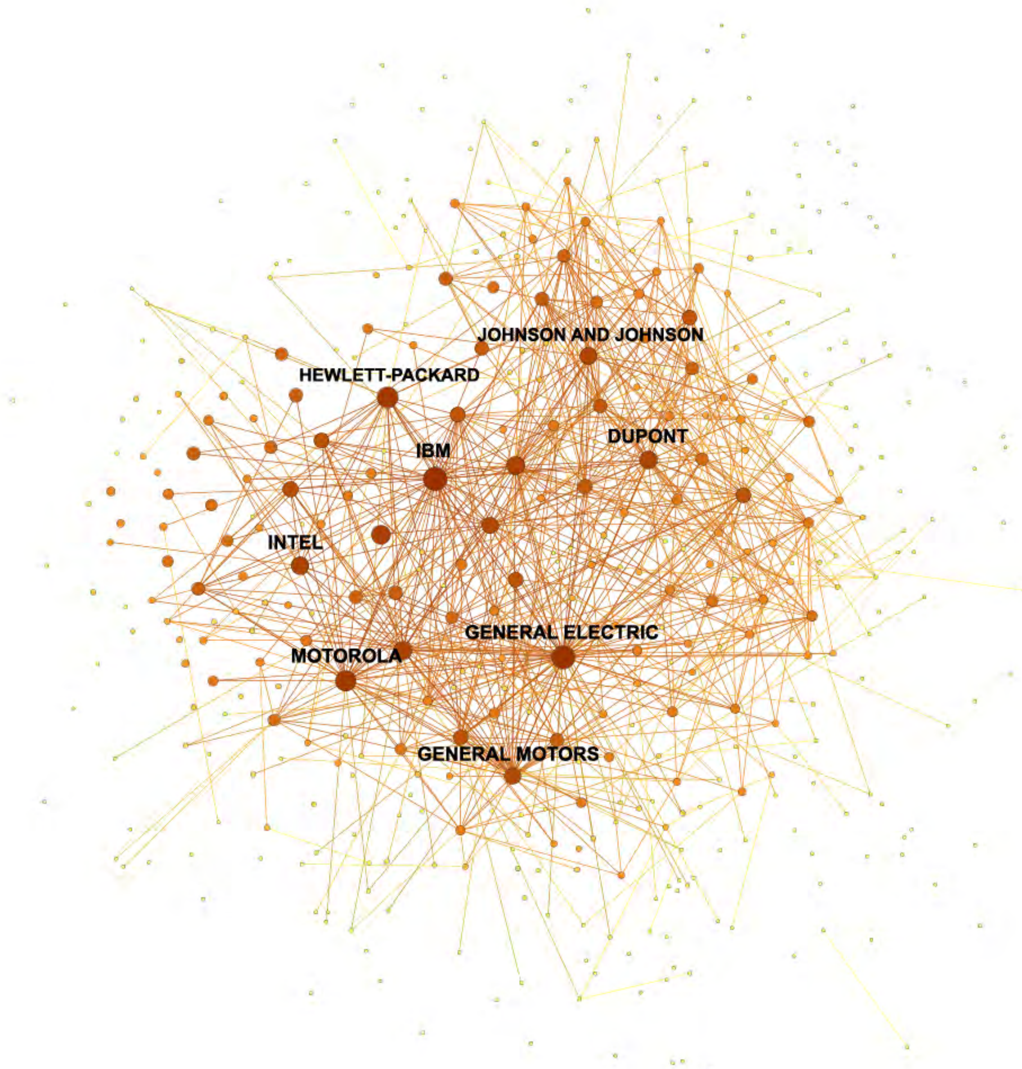


Figure E.2: The Network in 1990

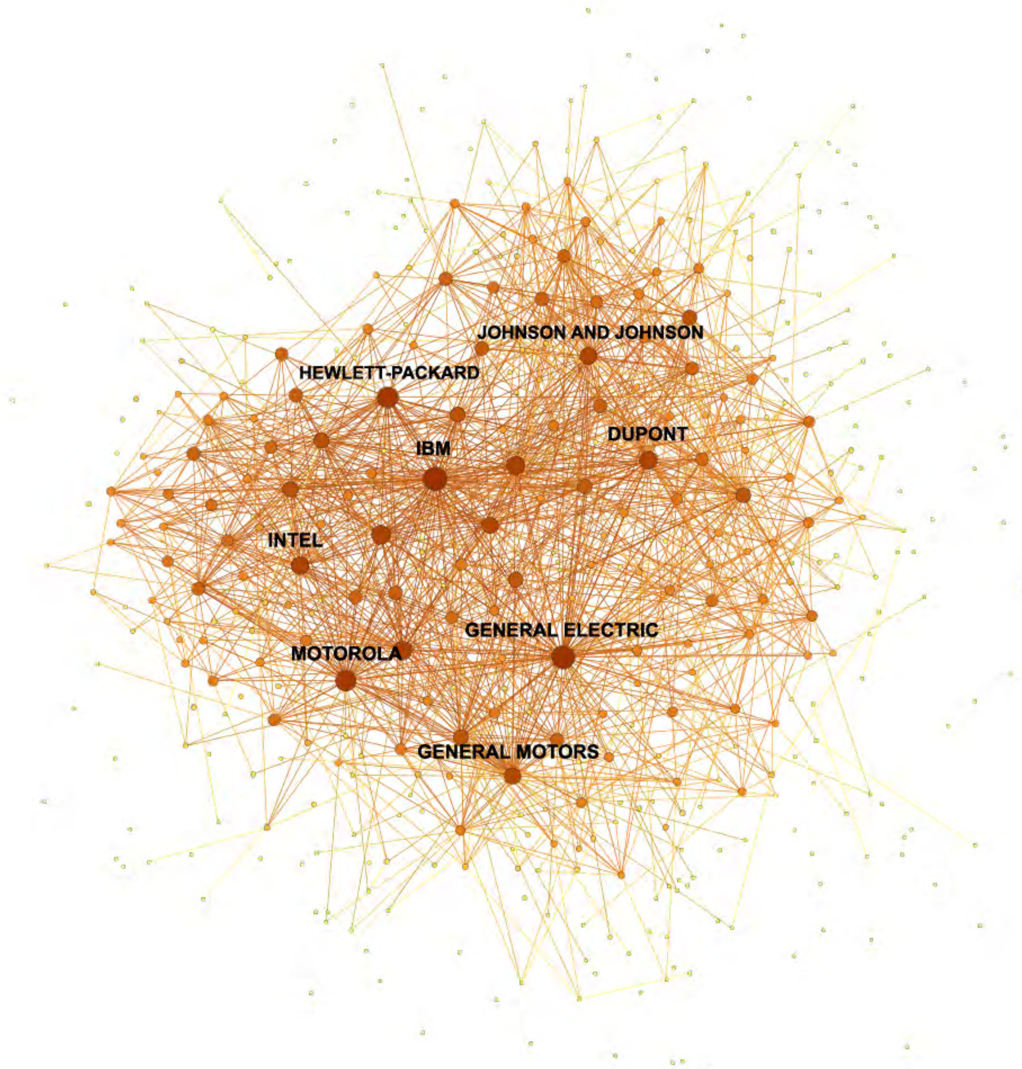


Figure E.3: The Network in 1995

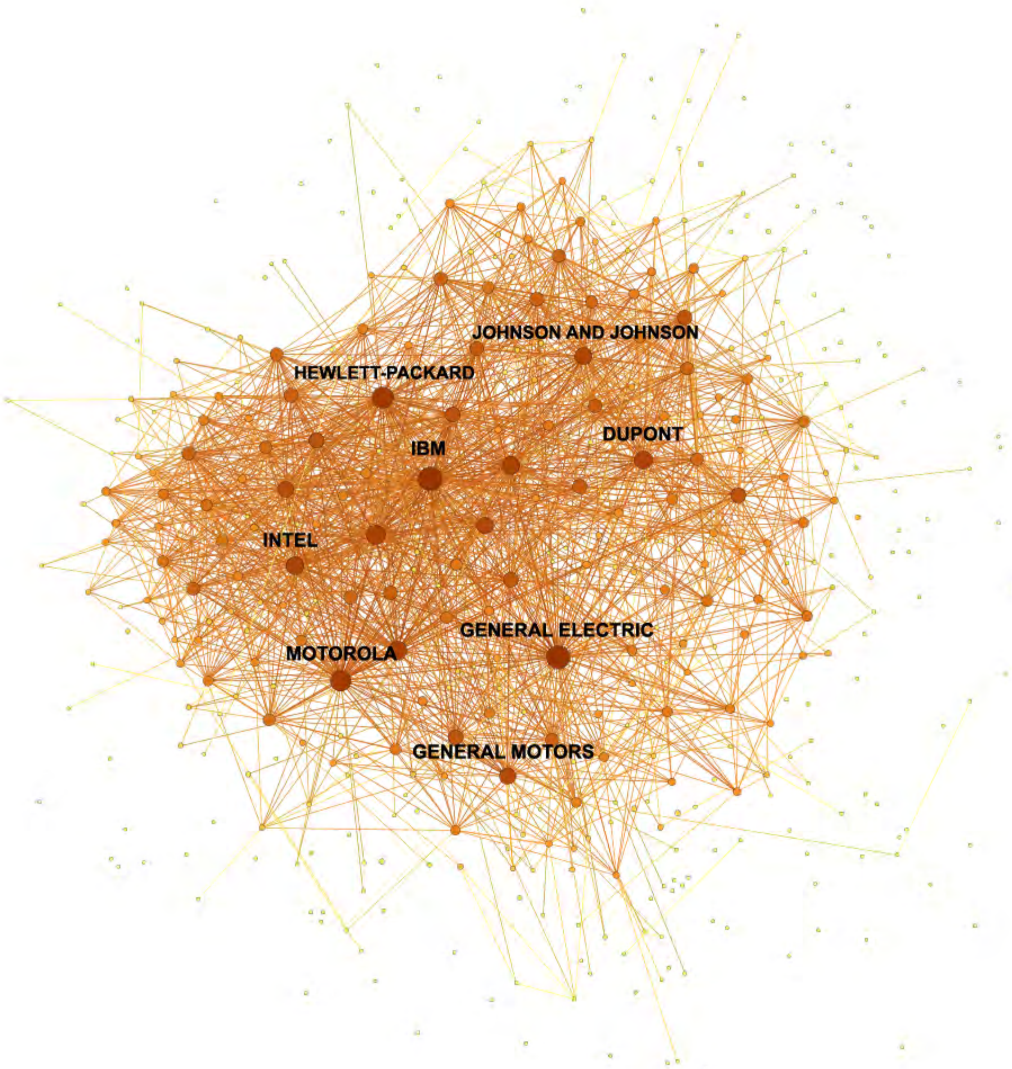


Figure E.4: The Network in 2000

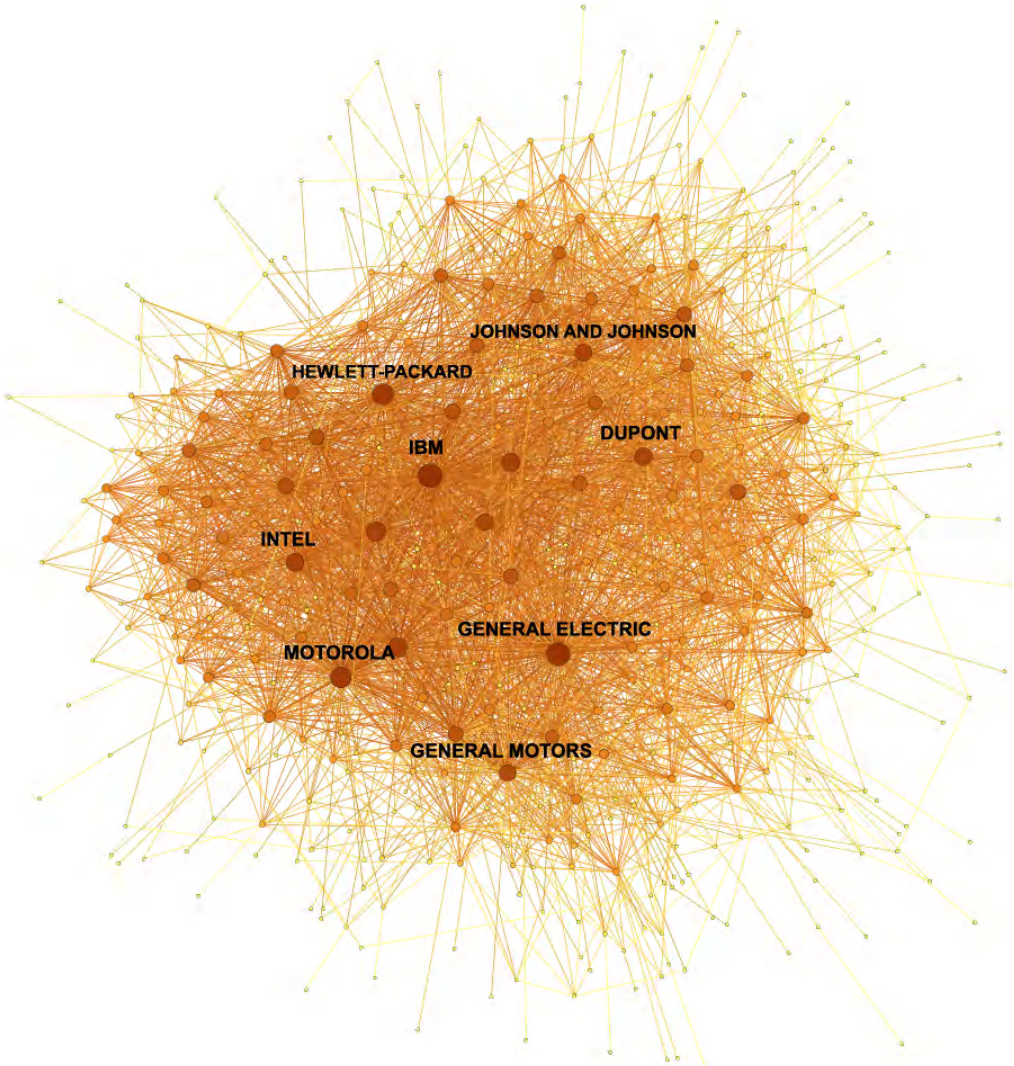


Figure E.5: The "Pooled" Network

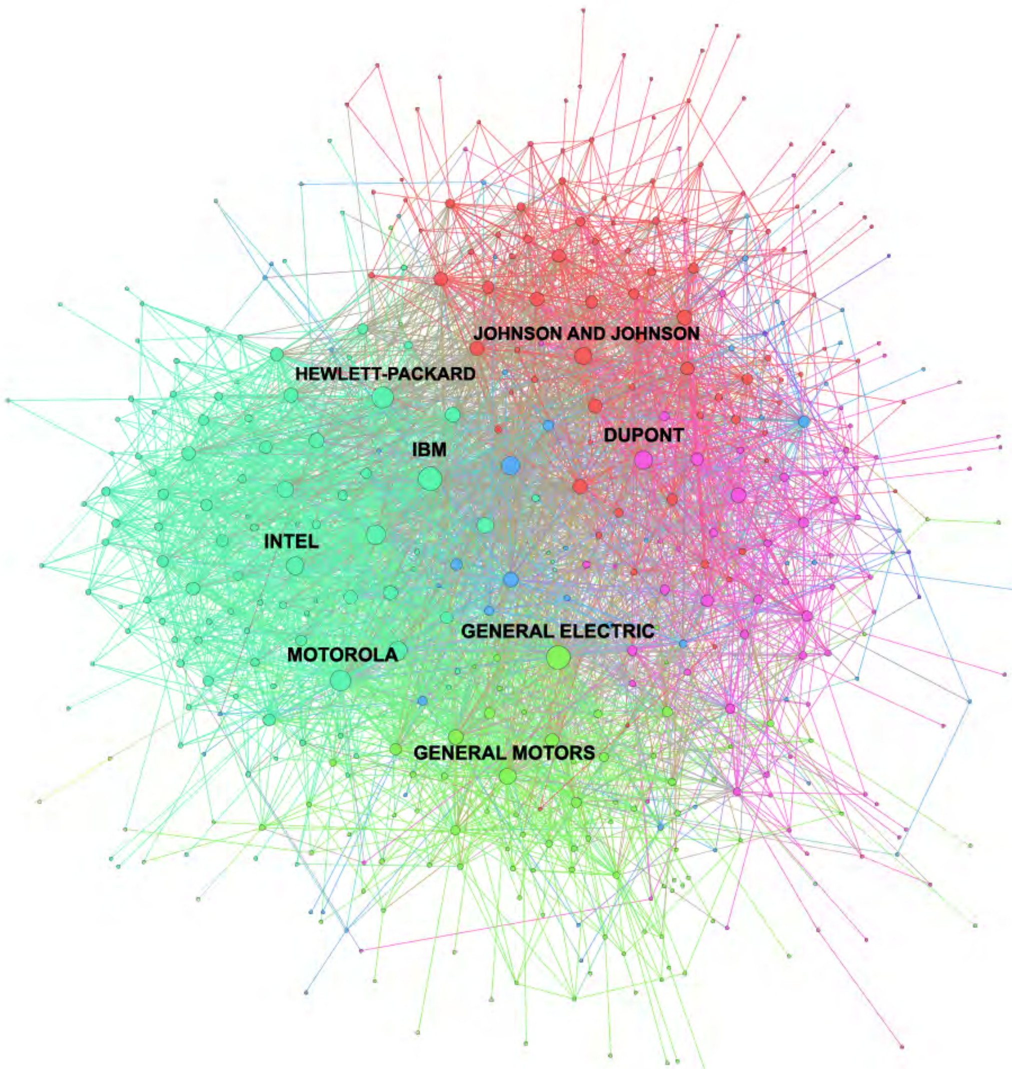


Figure E.6: Network Communities, Resolution $\rho = 1$

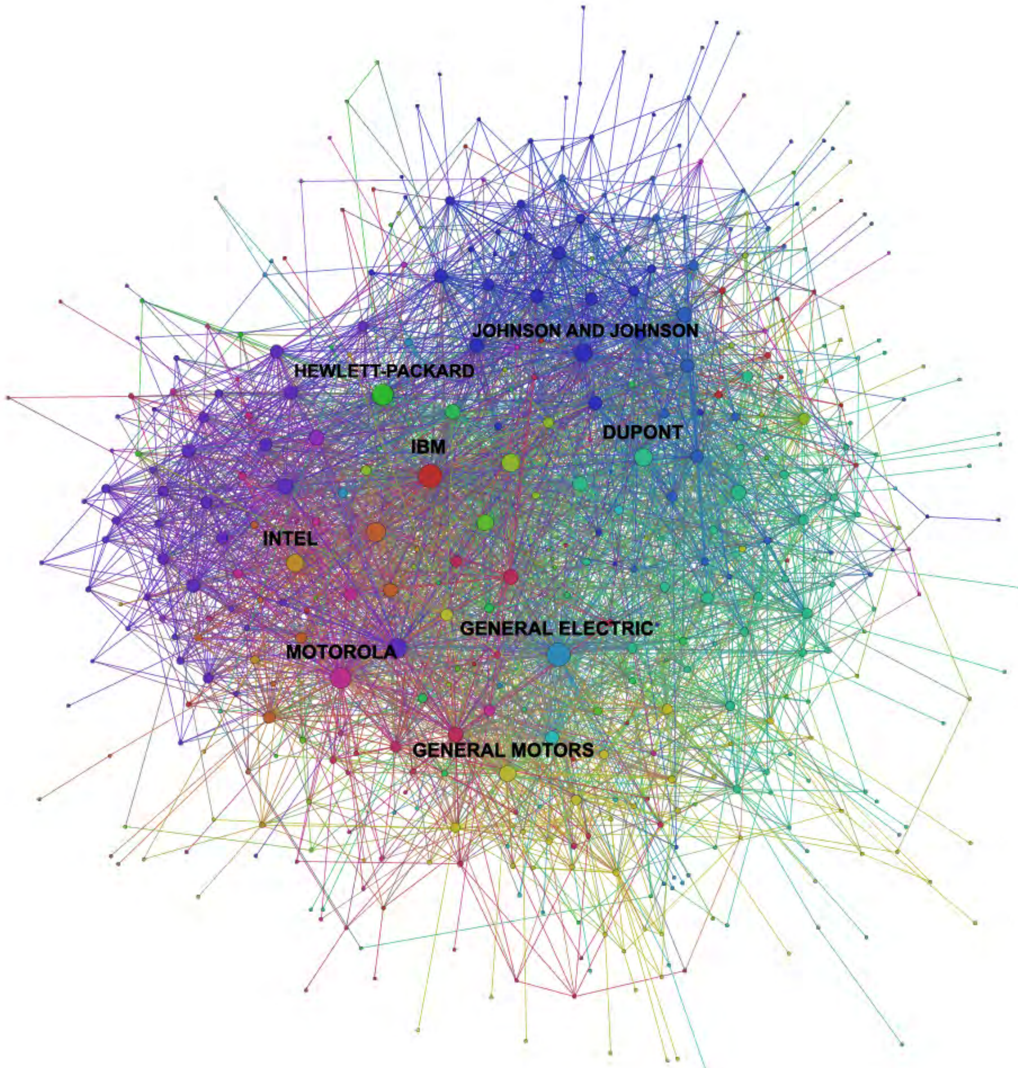


Figure E.7: Network Communities, Resolution $\rho = 0.6$