Knowledge Spillovers through Networks of Scientists

Paolo Zacchia*

CERGE-EI† & IMT School Lucca‡

May 15, 2019

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

JEL Classification Codes: C31, D85, O31, O33

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

*I am grateful to the editor Áureo de Paula, as well as four anonymous referees, for their careful review of earlier manuscripts and their remarks, which all contributed towards vast improvements of this article. I am also thankful to Enrico Moretti, Bryan Graham, Patrick Kline and Bronwyn Hall for their insightful advice throughout the development of this study. I also extend the due thanks to Nicholas Bloom, Daniel Gross, Elena Manresa, Santiago Pereda Fernández, John Van Reenen, Jose Vasquez, Brian Wright and all participants to seminars at U.C. Berkeley, the 2014 MCIC and IWCEE conferences, Banca d’Italia, Stockholm School of Economics, IMT School, New Economic School, Sant’Anna School, EIEF, K.U. Leuven, L.M.U. Munich, London School of Economics and Hungarian Academy of Sciences for their helpful comments and suggestions. All outstanding mistakes are mine.

†The affiliation to CERGE-EI is effective starting in September 2019.

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

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

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

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

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

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

Without applying this strategy, I find substantial effects of connected firms’ R&D on productivity. However, when instrumenting the knowledge of peers with the R&D choices of indirect links located at distance two or three, I obtain larger point estimates of spillover effects. This difference, though, only becomes apparent when employing the distance three instrument in isolation, which is remarkably consistent with the proposed framework. I interpret these findings as evidence that R&D is in fact driven by common unobservables across connected firms, and I advance several hypotheses to explain why this translates into a negative bias of OLS estimates. In light of the results, I estimate the social returns of R&D to be about 112% of the private returns among connected firms. Extending this strategy to models of firm market value and patent output results in similar patterns, as detailed in one of the paper’s appendices.

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. The quest for R&D spillovers in particular, initiated with the original intuitions by Griliches (1964, 1979, 1992), has developed into its current empirical framework with the cited contributions by Jaffe. Successive research has experimented with metrics of spillovers, based for example on cross-industry transactions or flows of patent citations, that are alternative to Jaffe’s concept of technological proximity; for a review of these studies see e.g. Hall, Mairesse, and Mohnen (2010). Other authors have assessed more specific mechanisms of knowledge diffusion. For example, Branstetter and Sakakibara (2002), as well as König, Liu, and Zenou (2018) study the effect of R&D joint ventures. Griffith, Harrison, and Van Reenen (2006) instead examine the consequences of UK firms’ technological outsourcing in the US.

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

---

1 For a general survey see Syverson (2011).
2 König et al. (2018) also take a network-based approach to their analysis of joint ventures, which bears similarities with the one in this paper. Their identification strategy is, however, different, and it does not address potential problems of correlated confounders.
first time a measure of cross-firms spillovers grounded on the observation of an actual social network of inventors: the patent collaboration network. This measure aims at capturing all kinds of individual interactions between inventors that eventually result in joint projects. While spillovers certainly also occur through less solid, harder to observe types of interactions, the proposed measure has the advantage of generality. In fact, collaborative cross-firm projects are common to other mechanisms examined in the literature, such as R&D joint ventures or technological outsourcing.

This work provides empirical evidence to support the hypothesis that spillovers are caused by the exchange of ideas between individuals. Therefore, it is related to the research about the micro-level determinants of performance in the workplace. Moretti (2004) argues that productivity is related to how well-educated the workforce is in the environment in which a plant is located, suggesting that knowledge spillovers have a local scope. Mas and Moretti (2009) demonstrate how “peer effects” apply at work, as coworkers intensify their efforts when they observe others doing increasingly so. Serafinelli (2018) 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.

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 are different. In the cited

---

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

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

5The model developed in this paper is inspired by those in Calvo-Armengol et al. (2009), Conley and Udry (2010), Bramoullé et al. (2014), Blume et al. (2015), while differing from all of them. The empirical model that I estimate is a production function with R&D spillovers; it corresponds to a Spatial Durbin Error Model (SDEM) from the classification of spatial econometric models by Elhorst (2014), as it includes both an analogue of Mankiw’s “exogenous effect” and spatially correlated errors.
articles, in fact, IVs are meant to solve Manski’s “reflection” problem, by extending methods originally devised for spatially autoregressive models (Kelejian and Prucha, 1998) to the case of networks. Here instead, the aim of the proposed methodology is to disentangle spillover effects from spatially distributed unobservables or “correlated effects.” The latter are, according to some authors (Angrist, 2014), the main cause for concern regarding studies on peer effects. Given that, as I illustrate, correlated effects and network formation are closely related, under the maintained hypotheses the proposed strategy is also robust to endogeneity of firms’ connections. To the best of my knowledge this approach is new in the spatial econometrics literature, and it can be viewed as a spatial extension of familiar GMM methods for dynamic panels.

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

1 Analytical Framework

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

The main feature of my model is that both are interdependent, inducing endogeneity. Note that a complete information version of this model can be manipulated so to return a spatially autoregressive (SAR) model with either R&D or firm output as the dependent variable. In such a case, Manski’s “endogenous effect” would be a function of the spillover parameter of the production function.

Bramoullé et al. (2009), De Giorgi et al. (2010), and some subsequent studies deal with common shocks by introducing fixed effects that are constant within the same network (for data constituted by a set of separate networks) or within distinct subdivisions of larger networks; under assumptions of this kind, appropriate transformations the data would address the problem. When the unobservables are – like in this paper – heterogeneous, and their spatial correlation is is a more complex function of network topology, such an approach would still result in inconsistent estimates.
1.1 Model Setup

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

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

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

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

Knowledge capital $\tilde{S}_i$ enters as an additional input into the general production

\(^7\)External knowledge can be either a complement or a substitute of in-house R&D. Which of the two aspects dominates empirically is a controversial matter, and a difficult one to solve with the information that is typically available to econometricians (see e.g. the discussion by Griliches, 1998, with regard to the complement vs. substitute nature of federal R&D).
function of firm $i$, which is also Cobb-Douglas:

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

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

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

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

$$w(S_i, \omega_i) = e^{\omega_i} S_i$$

that is, the cost borne for an additional unit of R&D $S_i$ increases with larger values of $\omega_i$. Cross-correlation of R&D costs in the network can be due to the fact that firms with similar technological characteristics $\omega_i$, which as argued may relate to the process of network formation, also face similar costs $\omega_i$. In practice, the spatial dependence in R&D costs may reflect, for example, common developments in the supply of labor endowed with specific technological skills, or in financing opportunities. Note that I make no restriction on the sign of the covariance $\text{Cov}(\omega_i, \omega_j)$ between the productivity and cost shocks of a firm, but I expect it to be positive in high-tech industries.
To allow for dependence between the network, the technological characteristics of different firms and their cost factors, I do not explicitly model network formation. Instead, I keep the model general by treating $G$, the vector of technological shocks $\omega = (\omega_1, \ldots, \omega_N)$ and the vector of cost shocks $\varpi = (\varpi_1, \ldots, \varpi_N)$ as random draws from some joint distribution $F(\omega, \varpi, G)$. I impose no restriction on this distribution, except that the process of network formation or other determinants reflected in $F$ are unlikely to set firms with similar characteristics too far apart in network space. To formalize this idea, it is useful to introduce a notion of distance between any two firms in the network. Specifically, let $d_{ij} \in \mathbb{N}$ be the minimum path length between firms $i$ and $j$: the lowest number of firms linked together as a sequence (path) indirectly connecting $i$ to $j$.\(^8\) Note that $d_{ij}$ is a function of $G$, and is itself a random variable. Armed with this concept, it is easier to express the following assumption.

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

$$
\text{Cov} (\omega_i, \omega_j | d_{ij} > C) = 0
$$

$$
\text{Cov} (\varpi_i, \varpi_j | d_{ij} > C) = 0
$$

that is, if the minimum path length between $i$ and $j$ is higher than $C$, their productivity shocks $\omega_i$ and $\omega_j$ are mutually independent, and so are their cost shocks $\varpi_i$ and $\varpi_j$.

In the context of this article Assumption 1 is interpreted as follows: if two firms have similar technologies, say one mainly operates in semiconductors and the other in Information and Communication Technology, they are not to be found very far away from one another in the network. This implies that if any two firms are sufficiently distant in a given observed network, their technological and cost characteristics are expected to be unrelated. In Appendix B I explore some models of network formation that involve multi-technology firms,\(^9\) and I illustrate how they implicate Assumption 1 either exactly or approximately, with low associated values of $C$.

---

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

\(^9\)These relate to the beta model of network formation (Chatterjee et al., 2011; Yan and Xu, 2013) and to results and insights by Charbonneau (2017), de Paula (2013, 2017) and Graham (2015, 2017).
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

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

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

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

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

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

1. Nature draws \((\omega, \varpi, \mathcal{G})\) from the common knowledge distribution \(\mathcal{F}(\omega, \varpi, \mathcal{G})\).

   Every firm \(i\) observes the network \(\mathcal{G}\) as well as its own information set \(\Omega_i\).

2. Firms simultaneously make their R&D and conventional input choices.

3. Payoffs (profits) are paid out.

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

1.2 Equilibrium Predictions

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

\[
E\left[\pi_i (S^*_i, X^*_i; S^{-i}_i) \mid \Omega_i\right] \geq E\left[\pi_i (S_i, X_i; S^{-i}_i) \mid \Omega_i\right] \quad \forall (S_i, X_i) \neq (S^*_i, X^*_i)
\]

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

**Proposition 1.** If \((\omega, \varpi)\) has bounded support, \(\gamma + \sum_{q=1}^{Q} \beta_q < 1\), and

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

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

\[
\log S^*_i = \log \gamma + \sum_{q=1}^{Q} \beta_q \left(\log \beta_q - \log \xi_q - \log \gamma\right) b^*_i (G; \vartheta) + s^*_i (\Omega_i, G) \quad (4)
\]

\[
\log X^*_{iq} = \log S^*_i + \log \beta_q - \log \xi_q - \log \gamma + \varpi_i \quad \text{for } q = 1, \ldots, Q \quad (5)
\]

where \(b^*_i (G; \vartheta)\) is the Katz-Bonacich network measure of centrality with attenuation factor \(\vartheta\) for \(i = 1, \ldots, N\), while \(s^*_i (\Omega_i, G)\) is a non-linear, spatially recursive function of firm \(i\)’s information set \(\Omega_i\) and network topology:

\[
s^*_i (\Omega_i, G) = \frac{1}{1 - \gamma - \sum_{q=1}^{Q} \beta_q} \left\{\bar{\omega}_i + \log E \left[\prod_{j=1}^{N} \exp \left(\sum_{j=1}^{Q} g_{ij} \delta \cdot s^*_j (\Omega_j, G)\right) \mid \Omega_i\right]\right\}
\]

where \(\bar{\omega}_i \equiv \omega_i \left(1 - \sum_{q=1}^{Q} \beta_q\right) \varpi_i\) for \(i = 1, \ldots, N\).
This result is easily interpreted. First, consider that (5) is simply a set of constant relative input shares conditions, which is typical of the maximization of Cobb-Douglas functions. By contrast, equilibrium R&D given in (4) can be decomposed in two parts. The first one represents the deterministic component, for firm \( i \), of the marginal return of R&D. It accounts for the complementarity of private R&D with both conventional inputs and the “certain” component of peers’ R&D, itself a function \( b_i^* (G; \vartheta) \) of firm \( i \)’s position in the network. The second part represents the best equilibrium prediction that firm \( i \) can make, on the basis of private information, on how random shocks to both productivity and R&D costs of all firms in the network (including itself) would alter its own net marginal return of R&D. In equilibrium, in fact, all the shocks may affect the R&D investment of connections, which is complementary to private R&D.

The Bayes-Nash equilibrium expressed in Proposition 1 is unique so long as the spillover parameter \( \delta \) is sufficiently small relative to the overall spillover weights of all other firms, and the technological and cost shocks \( (\omega_i; \varpi_i) \) of all firms are finite. Both conditions are necessary to rule out “explosive” scenarios in which some firms make infinite investments in R&D. In theory, an explosive scenario could be catalyzed by a single, highly connected firm in the network: because of strategic complementarities from the many connections, the optimal R&D investment of that one firm would be infinite; the strategic response of the connected firms would be analogous; and so on. Similar reactions might occur in the unrealistic cases where some firms are extremely technologically advanced \( (\omega_i = +\infty) \), or have zero R&D costs \( (\varpi_i = -\infty) \). Explosive scenarios are not encountered in the real world, and the empirical results of this paper about the \( \delta \) parameter are consistent with the conditions for uniqueness.

The next result underpins the empirical strategy of this paper.

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

\[
\text{Cov} \left( \omega_i, \log S_j^* \mid d_{ij} > C + L \right) = 0 \tag{6}
\]
\[
\text{Cov} \left( \log S_i^*, \log S_j^* \mid d_{ij} > C + 2L \right) = 0 \tag{7}
\]
\[
\text{Cov} \left( \log S_i^*, \log S_j^* \mid d_{ij} \leq C + 2L \right) \geq 0 \tag{8}
\]

that is, the unobserved shock of one firm and the equilibrium R&D strategy of another one are independent as long as the two are distanced by a minimal path length higher than \( C + L \); similarly, the equilibrium strategies of any two firms at distance higher than \( C + 2L \) are also independent, but may be correlated at distance \( C + 2L \) or lower.
Proposition 2 places a bound, in terms of “degrees of separation”, on the equilibrium correlation across R&D choices and unobserved shocks in the network. The intuition is the following: even if in equilibrium firms endogenously internalize the shocks of other organizations that are “sufficiently close” (up to distance $L$), and this in turn amplifies the exogenous cross-correlation (up to distance $C$), if both mechanisms are bounded their combined effect also is. In other words the shocks of other firms that are “very distant” in the network, whose R&D investment is of little relevance, are never internalized by individual firms. An implication of this result is that, for any firm $i$, the R&D choices of firms that are “sufficiently distant” in the network might 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 may depend on technological and cost factors that also affect the R&D of firm $i$’s connections, but not the R&D of firm $i$ itself. Given (6), (7) and (8), the candidate predicting firms are located at any distance $D$ between $C + L + 1$ and $C + 2L + 1$. Observe, though, that nothing in the hypotheses made so far ensures that the covariance (8) between the R&D of two sufficiently close firms is nonzero: whether any relevant instruments can be obtained is ultimately an empirical matter to be tested in the data.

A representation of this result is provided in Graph 1, which displays a network of four firms ($i, j, k, \ell$): a tetrad. In fact, this graph is composed by two open triads\(^{10}\) that partially overlap, as they share two nodes and an edge (the link between nodes $j$ and $k$). Consider first the simple situation where the shocks of any two firms are correlated if these are connected, but not otherwise (that is, $C = 1$). Furthermore, all

\(^{10}\)An open triad is a network (or a subset of a network) composed by three nodes; two of these three nodes are not connected to one another, but are both connected to the third one. In Graph 1 the two semi-overlapping open triads are represented by a solid and a dashed line, respectively.
firms only observe their own shocks \((L = 0)\). Since R&D only reflects private shocks, in equilibrium it is correlated only across pairs of firms that are connected to one another. Consequently, from the point of view of firm \(i\), the R&D of firm \(k\) \((S^*_k)\) can serve as an exogenous predictor of firm \(j\)'s R&D \((S^*_j)\), because the two are correlated but the former is independent from the R&D of firm \(i\) \((S^*_i)\). However, firm \(\ell\)'s R&D \((S^*_\ell)\) cannot act as a valid predictor, as it is necessarily uncorrelated with that of firm \(j\). Similarly, from the point of view of firm \(k\), \(S^*_i\) exogenously predicts \(S^*_j\). The same properties symmetrically apply to the dashed triad made of nodes \((j, k, \ell)\).

Consider now some slightly more complex cases. Suppose that firms still cannot observe the shocks of others \((L = 0)\), but the cross-correlation of R&D extends up to two degrees of distance as it reflects some primitive cross-correlation of the shocks \((C = 2)\). Hence, from the point of view of firm \(i\), \(S^*_\ell\) can act as a valid predictor of \(S^*_j\); symmetrically \(S^*_i\) would predict \(S^*_k\) for firm \(\ell\). In the case where \(C = 0\) and shocks are always observed between any two connected firms, but not beyond \((L = 1)\), the endogenous reflection of shocks is the only mechanism that drives the cross-correlation of R&D, which would still extend up to distance two. It follows that, for example, both \(S^*_i\) and \(S^*_k\) depend on \((\omega_j, \omega_j)\); yet, \(S^*_k\) is still a valid predictor of \(S^*_j\) for firm \(i\), as it is uncorrelated with \(\omega_i\), and vice versa. Note that \(S^*_\ell\) also correlates with \(S^*_j\); as they are both a function of \((\omega_k, \omega_k)\): thus, also firm \(\ell\)'s R&D is a valid predictor of firm \(i\)'s spillovers. The same logic applies when inverting the order of the nodes. Finally consider the case in which \(C = 1\) and \(L = 1\). Observe how R&D is correlated across the entire tetrad, but the R&D of firms at a distance of at least three degrees of separation are still valid predictors as per (6).

In Appendix A I discuss some other implications of Proposition 2, presented in the form of a corollary, that are relevant for the interpretation of the empirical estimates.

## 2 Networks and Data

This section is divided in three parts. In the first one, I formally introduce cross-firm measures of connection based on the underlying professional connections between the inventor teams of any two firms. In the second part, I describe the dynamic network of R&D spillovers obtained by measuring connections over time for each pair of firms from a panel of US companies. In the third part I discuss some descriptive statistics, including the important spatial cross-correlation of R&D in the network.
2.1 The Measures of Connection

Scientists typically maintain relationships with each other even beyond the borders of their respective organizations. Denote as $\mathcal{M}_{it}$ the set of inventors of firm $i$ at time $t$, with $\mathcal{M}_t \equiv \mathcal{M}_{it} \cup \cdots \cup \mathcal{M}_{Nt}$. I represent a co-patenting relationship between any two elements of $\mathcal{M}_t$, be they $m$ and $n$, with the notation $p_{(mn)t} = 1$. This indicates that two individuals, at time $t$, share some professional collaboration on any past research project that has resulted in a patent application featuring both their names. If such a relationship is absent, it is $p_{(mn)t} = 0$. One could visualize the resulting network as a graph where the elements of $\mathcal{M}_t$ are nodes that are linked by edges when $p_{(mn)t} = 1$. Graph 2 displays a stylized example about a hypothetical co-patenting network that is first observed at some point in time $t = 0$. In the network, inventors are represented as nodes, and they are displayed with different colors – one for each of three different firms (red for firm $i$, blue for $j$, green for $k$). Many pairs of inventors are linked to one another through co-patenting relationships $p_{(mn)t}$, but only one of these connections involves inventors from two different firms (specifically, firms $i$ and $k$).

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

$$c_{(ij)t} = f \left( \frac{\# \text{ inv.s of } i \text{ connected to } j \text{ at } t + \# \text{ inv.s of } j \text{ connected to } i \text{ at } t}{\# \text{ inv.s of } i \text{ at } t + \# \text{ inv.s of } j \text{ at } t} \right)$$ (9)
where $f : [0, 1] \to [0, 1]$, $f(0) = 0$ and $f(1) = 1$. These restrictions on $f$ ensure that measures of connection take values between 0 (no connection) and 1 (full connection) as per the standard normalization of the strength of edges in a weighted network. For the three firms in the example of Graph 2, $c^{f}_{(ij)t} = c^{f}_{(jk)t} = 0$, while $c^{f}_{(ik)t} = f(1/3)$.

The facts that $c^{f}_{(ij)t} \in [0, 1]$, and that any measure of connection is symmetric ($c^{f}_{(ij)t} = c^{f}_{(ji)t}$) bear important implications. The former means that an extra unit of external R&D cannot be more valuable for a firm than internally performed R&D, which is a reasonable hypothesis because in-house R&D is under direct control of the firm’s management. The latter implicitly assumes that the spillover relationship is symmetric between the two firms $i$ and $j$, regardless of the relative size of their R&D departments.\footnote{This is apparent from the example in Figure 2 where the two connected firms have different size. This assumption can have advantages: for example, it conveniently handles measurement errors in the assignment of individual inventors to firms. It may not be the most appropriate description of reality, however. For example, it might be that few “insiders” are enough to grasp much of another firm’s knowledge. In such a case, the symmetric ratio in (9) would downplay spillovers received by the smaller firm in a pair, and an “asymmetric” measure would be better suited.}

In addition, it must be stressed that such a connection measure essentially captures the relative number of personal relationships established in the past; it is silent about the relative importance of single linkages between inventors.\footnote{Alternatively, one can assume that relationships between inventors that are prolonged over the years, or collaborations resulting in many joint patents, are more relevant than others. Similarly, connections involving superstar inventors who issue many patents, of which some have been extremely well cited, can be more valuable for a firm than linkages to “ordinary” inventors.}

Connection measures between two firms can change over time. Their dynamics are the result of different types of events that are in principle observable, although I am not able to do so with the available data. Some of these events are: \(i\). cross-firm R&D collaborations, such as for instance joint ventures that result in joint patents, and \(ii\). 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, but their net effect is ambiguous and depends on the specific circumstances of the inventors who are involved.\footnote{New entrants increase the denominator of (9), but can also generate new cross-firm linkages, tightening connections. Similarly, the exit of scientists can decrease the denominator of (9), as well as the numerator if the leaving inventors used to connect some firms to each other.}

To illustrate, consider Graph 3, which extends the hypothetical network from Graph 2 by advancing one time period to $t = 1$. Some inventors have moved between firms or have left the network, while new ones have entered it. Graph 3 also shows new co-patenting relationships, represented
by dashed lines, that are due to patents appeared at \( t = 0 \). As a consequence of all
the various changes,\(^{14}\) it is \( c^{f}_{(ij)1} = f \left( \frac{1}{4} \right) \), \( c^{f}_{(jk)1} = f \left( \frac{1}{2} \right) \) and \( c^{f}_{(ik)1} = 0 \).

\[
\text{Graph 3: Inventors Network Example, } t = 1
\]

In the applied analysis I employ connection measures \( c^{f}_{(ij)1} = g_{(ij)t} \) that are based
on the square root function.

\[
g_{(ij)t} = \sqrt{\frac{\text{# inv.s of } i \text{ connected to } j \text{ at } t + \text{# inv.s of } j \text{ connected to } i \text{ at } t}{\text{# inv.s of } i \text{ at } t + \text{# inv.s of } j \text{ at } t}} \quad (10)
\]

This choice responds to a precise economic assumption. The typical anecdotal narra-
tive on technological spillovers usually involves some solitary individual who transfers,
perhaps by mistake, much of the knowledge internally developed by one firm to some
of its partners or competitors. The very expression “spillovers” is verbally associated
in such anecdotes to the “leakage” of few accumulating “drops” of knowledge. By ap-
plying 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 term “connection” so to indicate the squared root metric \( g_{(ij)t} \).

In Appendix E, I present the empirical results that are obtained from different choices
of function \( f \), as well as from alternative definitions of connection that depart from
some of the assumptions and requirements outlined in this discussion.

2.2 Firm-level Network

In the empirical analysis I combine different data sources. The firm-level network is
constructed from the data assembled by Bloom et al. (henceforth BSV) for their cited

\(^{14}\)Specifically, some incumbent inventors of firm \( j \) have applied for patents jointly with researchers
from firm \( k \), including an entrant inventor from that company; a new entrant in firm \( i \), not connected
to anyone elsewhere, has appeared; instead, among firm \( i \)'s incumbents one inventor has now moved
to firm \( j \), while the one who used to maintain the connection with firm \( k \) has left the network.
study. This is an unbalanced panel\textsuperscript{15} consisting of 736 mostly manufacturing, R&D-intensive firms listed on the US stock market, observed over the years 1976-2001. The BSV dataset combines accounting data from COMPUSTAT, firm-level patent counts, as well as Jaffe-type measures used by BSV to disentangle different types of spillover effects. Via firm and patent identifiers (see Hall et al., 2001), I match the BSV data to the “disambiguated” patent dataset by Li et al. (2014). The latter provides unique identifiers for individual inventors across different patents, thanks to a disambiguation algorithm that exploits information available in the USPTO database. Ultimately, this results in the selection of 1,315,060 patents granted to 565,019 inventors.

To calculate the connection measures, I need to associate inventors to each other as well as to firms. The first task is accomplished by looking at jointly filed patents. Specifically, for two inventors \( m \) and \( n \), I assign \( p_{(mn)t} = 1 \) if at time \( t+1 \) the USPTO has received at least one patent application (to be eventually granted) filed at any \textit{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.\textsuperscript{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 co-filed by inventor \( m \) and assigned to firm \( i \) in order to define a time interval in which one can reasonably presume that the individual was crucial for the R&D activity of that organization. The details of the assignment rule are provided in Appendix C.

I calculate measure (10) for each pair of firms and for every year from 1981 to 2001, using patents granted since 1976. As calculating connections requires the observation of enough antecedent patents, I abstain from doing it for 1976-1980; since the BSV panel is thin in those years this is a small loss. In total 460 firms display at least one positive connection with another firm in any year of the time interval under analysis. The number of firms that are actually connected in any year varies over time: some of the initially unconnected firms would eventually develop bonds. Conversely, firms that are already connected in 1981 may experience variations in the number of their connections (possibly resulting in the loss of all of them), or leave the sample. Thus, one never observes all the 460 firms of the dynamic network in each cross section.

\textsuperscript{15}The panel is unbalanced because of the entry and exit of firms, as they go public or are subject to mergers and acquisitions (see BSV’s appendix). The average length of the panel is 17.8 years.

\textsuperscript{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.
Figure 1 portrays, for every year from 1981 to 2001, the overall number of firms in the original panel (blue dashed line), the number of those displaying any nonzero connection (blue continuous line), as well as the average number of connections per firm, both in the whole sample and for the subset of connected firms (red dashed and continuous lines, respectively). Figure 1 displays a steady increase in the number of connected firms between 1981 and 1998, followed by a drop from 1998 to 2001 – partly because of attrition in the original panel, which is particularly severe in later years. Among connected firms, the average number of connections increases steadily over the entire time frame.\footnote{This increase can be attributed to several factors, for example: the mechanical effect due to the overall increase in the number of patents over the two decades, the diffusion of R&D joint ventures, and the emergence of collaborations between universities and firms, linking inventors from various firms together. More informative patent data are necessary to distinguish between these mechanisms.} By construction, this is also reflected by the unconditional average represented by the dashed red line. Another way to appreciate the temporal evolution of the network is to visualize it in the form of graphs; selected graphs for the years 1985, 1990, 1995 and 2000 are reported in Appendix D.

Figure 2 summarizes the yearly distributions in the number of connections per firm (called “degree”) of connected firms. Like in most networks, the degree distribution is
very asymmetric; moreover it tends to widen over time. The most connected firms go from less than 10 links in the early 80s to several dozens of them by the year 2000, but
the median number of connections only increases from 1 to 5. As it is shown in Figure 3, the empirical distribution of connections is quite asymmetric, but stable over time. The average of $g_{(ij)t}$ is 0.083, with 0.066 standard deviation.\footnote{Recall that this refers to the squared-root connection measure as defined in (10). The average for baseline linear measure is 0.012, with 0.028 standard deviation.} A measure useful for interpreting the empirical results is the row sum of connections, that is the sum of all of one firm’s connections in one year: $\bar{g}_{it} = \sum_{j \neq i} g_{(ij)t}$. Among connected firms, the mean and standard deviation of $\bar{g}_{it}$ are respectively 0.50 and 0.57. The yearly empirical distribution of $\bar{g}_{it}$, not shown for brevity, spreads out over time mirroring the dynamics of the degree distribution.

The average of $g_{(ij)t}$ is 0.083, with 0.066 standard deviation. A measure useful for interpreting the empirical results is the row sum of connections, that is the sum of all of one firm’s connections in one year: $\bar{g}_{it} = \sum_{j \neq i} g_{(ij)t}$. Among connected firms, the mean and standard deviation of $\bar{g}_{it}$ are respectively 0.50 and 0.57. The yearly empirical distribution of $\bar{g}_{it}$, not shown for brevity, spreads out over time mirroring the dynamics of the degree distribution.

Figure 4: Triad Census

Figure 4 displays the temporal evolution of the Triad Census, namely the total count of open and closed triads.\footnote{In analogy with the definition of open triad, a closed triad is a (sub)network whose three nodes are all connected to one another.} Between years 1981 and 2001, one counts in total 160,365 open triads and 15,623 closed triads; the latter are about 10% of the former. The number of both types of triads grows over time in analogy with average degree. The preponderance of open triads indicates that the network does not feature an excessive degree of clustering for the purpose of this article’s empirical strategy. If most
triads were closed, in fact, it would not be possible to identify “indirect connections” (other firms located at distance 2 or higher) for many firms in the network.

### 2.3 Summary Statistics and Spatial Correlation

In Table 1 I report some firm-level summary statistics. To this end I split the sample into five groups: one for the firms that never enter the network, and four groups for those that do. Specifically, I calculate the overall sum of connections for each firm as \( g_i = \sum_t g_{it} \), and I assign each firm to a group on the basis of its classification within quartiles of \( g_i \). Quartile 1 contains the least connected firms in the network over the whole period; quartile 4 contains the most connected ones. Note that the four quartile groups contain different numbers of unique observations because of panel attrition.

**Table 1:** Summary Statistics, 1981-2001

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

**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 \( g_i \). Standard deviations are in parentheses. \( S_{it} \) is the R&D stock, see Section 3. For a description of the Jaffe measure, see Appendix C.
For each of these groups, I pool all the observations of the relevant firms and calculate the mean and the standard deviation of specific variables. In addition to real sales $Y_{it}$, Tobin’s $q$ – defined as the ratio between a firm’s market value $V_{it}$ and the replacement value of its assets $A_{it}$ – citation-weighted patents $P_{it}$, and the number of employees $E_{it}$, I also report (for easier interpretation) statistics for the ratio of $Y_{it}$ to several inputs or spillover measures, which can be deduced from the table. Finally, the last row is about the ratio of $Y_{it}$ to the main independent variable of the empirical analysis: the connections-based spillover component of knowledge capital. The table illustrates that the firms inside the network – especially the most connected ones – are larger, more R&D intensive, more innovative and more productive than those outside of it.

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

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

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

3 Econometric Model

This section concerns the empirical methodology of this article, and it is divided in three parts. In the first one, I discuss the workhorse model for the evaluation of R&D spillovers on productivity. In the second part I describe the empirical strategy aimed at addressing endogeneity. In the third part I make some miscellaneous considerations about models for other outcome variables, network communities and standard errors.

3.1 Production Function Specification

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

$$ \log Y_{it} = \alpha_i + \sum_{q=1}^{Q} \beta_q \log X_{itq} + \gamma \log S_{it} + \delta \sum_{j=1}^{N} g_{(ij)t} \log S_{jt} + \tau_t + \nu_{it} \quad (11) $$
where the unobserved shock $\omega_{it}$, which is now allowed to vary across firms and over time, is decomposed as $\omega_{it} = \alpha_i + \tau_t + \upsilon_{it}$, that is by a firm-invariant effect ($\alpha_i$), a year effect ($\tau_t$), and finally a residual error term ($\upsilon_{it}$). Here $S_{jt}$ denotes the R&D stock of firm $j$ at time $t$, and $g_{(ij)t}$ is the connection measure between firms $i$ and $j$ at time $t$, with $g_{(ii)t} = 0$ for all $i$ and for all $t$. The R&D stock $S_{it}$ is constructed, following a customary approach in the literature, as the depreciated sum of past expenditures on R&D up to year $t - 1$ (Griliches, 1998).20 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 $\delta$ represents the overall strength of the R&D spillovers in the network. It is interpreted as the elasticity of a connection-weighted neighbor’s R&D on firm productivity. It is useful for different kinds of thought experiments: for example, a firm $i$ connected to a neighbor $j$ with connection $g_{ij} = 0.4$ receives a $0.4\delta$ percentage increase in productivity following a 1% increase in the R&D stock of firm $j$. Similarly, a firm with row sum of connections $\bar{g}_{it} = 4$ benefits from a $4\delta$ percentage increase in productivity following a 1% rise in the research effort of all its neighbors. By contrast, parameter $\gamma$ measures the elasticity of firm productivity with respect to changes in the private (in-house) R&D stock.

Since actual physical quantity $Y_{it}$ is not observed in the BSV dataset, I proxy it by the deflated sales of firm $i$ in year $t$, as it is customary in studies dealing with the estimation of production functions. Deflated sales, however, conflate both supply and demand factors. To control for the latter, like BSV I include in the specification of (11) some industry-level market outcomes, namely current and lagged industry sales, and current industry prices. In addition to demand side controls and conventional inputs (capital and labor in year $t$), I also include other R&D spillover variables on the right-hand side of (11). Their purpose is to more convincingly restrict the interpretation of $\delta$ to the sole effect of other firms’ R&D induced through the collaboration network. For simplicity, in expression (11) I do not explicitly distinguish between conventional inputs and other demand or spillover controls; with some abuse of notation I treat all of these as different elements of the set of covariates $\{X_{itq}\}_{q=1}^Q$.

The additional R&D spillover variables deserve further elaboration. Two of them

---

20In the BSV dataset the R&D stock is constructed, like in other studies, by applying the perpetual inventory method with a 15% depreciation rate. The R&D stock measure employed in this analysis reads as $S_{it} = R_{i(t-1)} + .85S_{i(t-1)}$, where $R_{it}$ is the flow of R&D investment at time $t$. 
are the key regressors of the BSV study: Jaffe’s classical measure of (beneficial) R&D spillovers, and BSV’s original measure of “business stealing,” which aims at capturing the negative externality in terms of residual demand that is due to market rivals’ R&D. Both are constructed weighing external R&D $S_j$ by some metric of similarity between firm $j$ and a reference firm $i$: the correlation of patent technological classes in the case of Jaffe’s measure, and that of sales’ allocation across sectors for the BSV business stealing measure.\(^\text{21}\) In addition, I include one spillover variable which accounts for the geographic co-location of firms’ R&D, so to control for the possibility that connections $g_{ij}t$ simply capture firms’ spatial proximity and other spatially correlated factors. To construct this measure, I weigh external R&D by a measure of proximity analogous to (10): the square root of the relative proportion of two firms’ inventors who reside in the same statistical metropolitan area (CBSA). Appendix C provides additional details and descriptive statistics about this “Geographic Spillovers” measure.

### 3.2 Endogeneity and Instrumental Variables

The estimation of $\delta$ in equation (11) suffers from two potential endogeneity problems. The first is the possible correlation between the error term of one firm $i$ and the R&D of its connections $j$, that is $\mathbb{E} [\log S_{jt} \cdot v_{it}] \neq 0$. According to the analytical framework of this paper, this follows either if the R&D of firms endogenously reflects the shocks of connections (in addition to own shocks), or if the shocks themselves are correlated ($\mathbb{E} [v_{it} v_{jt}] \neq 0$). Both cases relate to the issue of correlated effects from the analysis of spillovers in the classroom by Manski (1993); in the production function context, these are spatial generalizations of the classical “transmission bias” due to unobserved shocks. The second problem is the endogeneity of connections: $\mathbb{E} [g_{ij}t v_{it}] \neq 0$. This is caused by differences in firms’ propensity to establish connections as a function of their unobserved shocks; for example, highly prolific, well connected inventors may be more inclined to move towards more productive firms.\(^\text{22}\)

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

\[
E \left[ \log S_{kt} \cdot v_{it} \mid h_{(ik)t}^D = 1 \right] = 0
\]

(12)

where \(h_{(ik)t}^D = 1 \left[ d_{(ik)t} = D \right] \) is a dummy variable indicating that firms \(i\) and \(k\) are located, in year \(t\), at some given distance \(C + L + 1 \leq D \leq C + 2L + 1\). Holding (12), the following moment condition is then easily derived by iterating expectations.

\[
E \left[ h_{(ik)t}^D \log S_{kt} \cdot v_{it} \right] = 0
\]

(13)

Within this framework the implicit instrument \(h_{(ik)t}^D \log S_{kt}\) is always valid; whether it is also relevant is – as argued earlier – an open question to be evaluated empirically.

The econometrics of moment condition (13) deserves more discussion. Intuitively, the R&D of indirect connections located at distance \(D\) is generated, under the maintained hypotheses, by some variation that is independent of firm unobservables \(v_{it}\) (it is econometrically exogenous). Also observe how the model explicitly allows, within the confines of Assumption 1 and 2, for the network to be endogeneous; hence, condition (13) allows to address both endogeneity problems outlined above. It is useful to draw a parallel between this approach and GMM methods that are typical of the dynamic panel literature (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). In the latter, following a transformation of the model aimed at removing fixed effects, endogenous lagged variables of interest are instrumented by “sufficiently past” further lags. Similarly, here I instrument the possibly endogenous first spatial lag of R&D by “sufficiently distant” farther spatial lags of R&D.

In terms of practical implementation, it must be noted that it is generally possible to find – for each firm \(i\) – multiple other firms \(k\) located at distance \(D\). Each of them corresponds to a moment condition expressible as (13); all of these moments can be empirically operationalized in different ways, for example as independent conditions in a GMM problem.\(^{24}\) However, in this analysis I opt for a simple approach: I aggregate

\(^{23}\) Firms accumulate R&D stock (knowledge capital) over time. In Appendix B I discuss conditions under which Proposition 2 can be extended to a dynamic model. In particular, it is necessary that the network tends to become denser instead of sparser (intuitively, if distances are lengthened (12) might not hold, because both \(\log S_{kt}\) and \(v_{it}\) would preserve the memory of past linkages). Moreover, if errors are serially correlated and firms do not pre-commit to long term R&D plans, firms must be unable to observe past R&D choices of sufficiently distant firms (“finite spatial memory”).

\(^{24}\) In this case, it would be interesting to analyze how the optimal GMM weighting matrix varies as a function of network topology. This is an intriguing topic for further research.
all those moment conditions linearly, taking the summation of (13) over all firms in year $t$. This results in a single moment condition:

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

which amounts to selecting the log-R&D of all “indirectly connected” firms $k$ located at distance $D$, giving each an equal weight in the summation. As I illustrate in the next section, in order to identify $\delta$ I employ moment condition (14) and variations of it in both simple IV-2SLS and System GMM estimates of the production function.

This still leaves the question about the appropriate value of $D$ to be chosen for estimation open. The previous discussion of Figure 5 suggests selecting indirect links located at both distances 2 or 3, or just distance 3. It is important to discuss how the choice of $D$ relates to the data employed in the analysis. For moment condition (14) to be credible, it is necessary that such values of $D$ be large enough to separate out firms with different unobservables. In fact, firms from the BSV sample are large multi-business, multi-technology companies, presenting only weak correlation between their patent and product portfolios (see Figure D.9 from the Appendix). While any two connected firms are likely to share some characteristics, they are also expected to be different in other respects. Since any spatial cross-correlation that is stronger at low distances must mechanically decay at higher distances, cross-firm similarities are at best negligible at distance 3. This argument provides an empirical rationale for the indirect evidence given in Figure 5 and the choice of $D = 3$.

### 3.3 Miscellaneous Considerations

**Additional Outcomes.** From the seminal study by Jaffe (1986) to BSV, it has been customary to estimate R&D spillover effects on indicators of firm performance and innovation rate other than productivity. In Appendix E, I describe extensions of the BSV specifications for models of firm market value and patent output which accommodate the connection-based measure of spillovers.\(^{25}\) The identification strategy that I propose is easily adapted to additional outcomes; in the case of non-linear models estimated via maximum likelihood, it is morphed into a control function approach.

\(^{25}\)Specifically, the regression equation for market value extends the semi-logarithmic specification by Jaffe (1986) and Hirsch and Seaks (1993); the effects on patent output are estimated via a negative binomial model (Hausman et al., 1984; Blundell et al., 1995) that is typical of these studies.
Network Communities. For the sake of improving the credibility of the empirical estimates and for calculating standard errors, 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 adding up connection weights $g_{ij,t}$ over the time series. The Louvain algorithm is a popular tool in network analysis that is used to identify hierarchies of communities or clusters. At every level of one hierarchy, connections are dense within groups and sparse between groups. The algorithm can be fine-tuned by varying the “resolution parameter” $\varphi$ which selects different levels of the hierarchy. Appendix D provides a visualization of cluster assignment for different values of $\varphi$.

Standard errors. The hypothesized cross-correlation across unobserved shocks invalidates standard asymptotic properties of any GMM/2SLS or MLE estimator. To perform credible statistical inferences, I adopt a transparent clustering approach based on network communities that is consistent across both linear and non-linear models. 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). This is particularly important with large networks, because if the structure of cross-node dependence is unknown, any partition of a network into different communities would result in some cross-cluster dependence. Bester et al. advocate using as few and as large clusters as possible. To strike a balance between the CCE approach by Bester et al. and standard practices of clustering standard errors, I define 20 communities by running the Louvain algorithm with resolution $\varphi = 0.6$. Because of serial correlation, all observations of the same firm in the panel enter the same cluster; for estimates that include firms outside the network, each of these constitutes an additional cluster. Statistical inferences are not substantially altered by the average size or number of clusters.

---

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

27 A possible alternative is an HAC estimator adapted to spatial correlation, like the one proposed by Conley (1999) for cross-sectional data characterized by well-defined notions of “distance” between observations, which might be extended to networked data. In this context, though, 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.

28 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 the CCE approach they advance. This difference is particularly pronounced in the case of spatial dependence.
4 Empirical Results

In this section I present the empirical results of the paper. This section is divided in four parts. In the first part I present the OLS estimates of the production function. In the second and third parts I illustrate, respectively, IV/2SLS and System GMM estimates based on the proposed identification strategy. Finally, in the fourth part I recapitulate and interpret the results, and I discuss their economic significance.

4.1 OLS Estimates

Table 2 displays the results from the estimation of equation (11). Across all estimates I take both firm and year fixed effects; and I cluster standard errors according to the approach based on network “communities” outlined above. Along with the estimate of $\gamma$ and $\delta$ I report those for Capital and Labor. The estimate of spillovers $\hat{\delta} = 0.016$ from column (1) can be interpreted in light of different thought experiments. For example, the quantity $\hat{\delta}g_{ij,t}$ represents the elasticity of output with respect to a 1% increase in the R&D stock of another firm with connection $g_{ij,t}$. In the case of an average connection $g_{ij,t} = 0.083$, the implied elasticity is 0.0013. Hypothesizing instead a 1% increase in the R&D stock of all of one firm’s neighbors, the implied effect on firm $i$’s output is a $\hat{\delta}g_{it}$% rise. For a firm connected to the network with average row-sum $\bar{g}_{it} = 0.50$, this corresponds to a 0.008% increase. The estimated elasticity of private R&D, $\hat{\gamma} = 0.045$, appears in comparison to be one order of magnitude larger.

Relative to column (1), in (2) I show the effect of controlling for the Jaffe measure of knowledge spillovers based on technological proximity, as well as for the geographic R&D intensity measure. The inclusion of both does not dramatically impact the point estimate $\hat{\delta}$, which falls to about 0.015 while remaining statistically significant. The geographic control, on the other hand, seems to have very little economic significance. To control for the possibility that the estimate $\hat{\delta}$ is driven by persistent productivity differences between firms that belong to the network and those that do not, in column (3) I report estimates restricted to the subsample of firms that enter the network at any point in time. This exercise has an interesting implication: while parameter $\hat{\delta}$ is estimated substantially smaller (down to 0.011) yet still statistically significant, the coefficient for the Jaffe measure of spillovers falls more sharply becoming no longer statistically significant. As firms that do not belong to the network are the smallest and least R&D-intensive ones, this result implies that the positive correlation between
### Table 2: Production Function, Ordinary Least Squares Estimates, 1981-2001

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

Notes: The table reports OLS estimates of model (11). Columns 1 and 2 are estimated over the entire original sample of 736 firms in the time interval 1981-2001. Estimates in columns 3, 4 and 5 restrict the sample to firms with at least one nonzero connection ($g_{ijt} \neq 0$) in any year $t$; all observations of these firms are also included for years with no connections. All estimates include firm and year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with $\varphi = 0.8$ (10 communities) in column 4 and $\varphi = 0.6$ (20 communities) in column 5. Standard errors are clustered by the 20 “communities” obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. For estimates not restricted to the network, firms outside the network constitute single clusters. Asterisks denote conventional significance levels of $t$-tests (* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$).

Real sales and the Jaffe measure is largely driven by small firms patenting in the most R&D-intensive technological fields.\(^{29}\)

In estimates (4) and (5) I also include an additional set of dummy variables, in an initial attempt to control for the fact that connected firms may be subject to similar

---

\(^{29}\)This finding can be interpreted as a sample selection bias. COMPUSTAT only reports data for public firms; small firms going public are usually successful firms, and those that “make it into the news” are typically from fast developing high-tech sectors (and being in the news is itself endogenous). If a correlation exists between the Jaffe measure and the probability that small firms go public, this would be reflected into 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.
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 ($\phi = 0.8$); while in column (5) the additional dummy variables are based on the same 20 communities also used for clustering standard errors ($\phi = 0.6$). Increasing the number of clusters does not dramatically affect the point estimate $\hat{\delta}$. This suggests that the correlation between the connections-induced measure of spillovers and one firm’s output is in fact driven by the variation in the R&D stock of that firm’s linkages.

### 4.2 IV/2SLS Estimates

I now illustrate the empirical results from the application of the IV strategy aimed at addressing correlated confounders and network endogeneity. I instrument the R&D stock of one firm’s direct connections by aggregating the R&D of its “indirect connections” located at distance 2 and 3. In light of the theoretical analysis and of the evidence on the spatial autocorrelation of R&D in the network presented in Figure 5, both instruments could be valid in principle. However, the farther instrument constructed at distance $D = 3$ is more likely to be uncorrelated with both unobserved factors and the other input variables of firm $i$.

In Table 3 I report the results of various first stage regressions associated with model (11). All estimates are restricted to the subsample formed by those firms that ever enter the network. I regress the network-based spillovers variable on the aggregated log R&D stock of indirect connections located at either distance 2 (column 1), distance 2 and distance 3 (column 2), distance 3 only (column 3). The estimates from columns (4) and (5) are analogous to those in column (3), but they additionally include community-by-year fixed effects (respectively based on 10 and 20 communities, in analogy with Table 2). Noticeably, both instruments are strongly, positively correlated with the endogenous spillover variable. The measured $F$-statistics across all first stage estimates are reassuringly high: the lowest $F$-statistic is larger than 19. As expected, the correlation between the R&D of direct connections and the R&D of distance 2 indirect links appears much larger (by about one order of magnitude) than the analogous correlation with the R&D of distance 3 indirect connections.\(^{30}\)

\(^{30}\)These first stage regressions are appropriate linear projections for the sake of 2SLS estimation, but do not provide consistent estimates of the patterns of spatial correlation of R&D. In fact, OLS is by construction an inconsistent estimator of any spatially autoregressive model. This may explain
Table 3: Production Function, First Stage Estimates, 1981-2001

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance 2 Instrument</td>
<td>0.0043***</td>
<td>0.0044***</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td></td>
</tr>
<tr>
<td>Distance 3 Instrument</td>
<td>-0.0001*</td>
<td>0.0006***</td>
<td>0.0005***</td>
<td>0.0005***</td>
<td></td>
</tr>
<tr>
<td>Private R&amp;D</td>
<td>0.1598***</td>
<td>0.1774***</td>
<td>0.5022***</td>
<td>0.4612***</td>
<td>0.4357***</td>
</tr>
<tr>
<td>Capital</td>
<td>0.1827</td>
<td>0.1895</td>
<td>0.6169**</td>
<td>0.5903**</td>
<td>0.5923**</td>
</tr>
<tr>
<td>Labor</td>
<td>-0.0761</td>
<td>-0.0530</td>
<td>-0.6482**</td>
<td>-0.6021**</td>
<td>-0.6541**</td>
</tr>
<tr>
<td>Jaffe Tech. Proximity</td>
<td>1.8439***</td>
<td>1.8224***</td>
<td>3.3663**</td>
<td>3.2740**</td>
<td>3.3672**</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Only Network</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>No. of Communities</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>(Community × Year Effects)</td>
<td>7607</td>
<td>7607</td>
<td>7607</td>
<td>7607</td>
<td>7607</td>
</tr>
<tr>
<td>F-statistic</td>
<td>255.17</td>
<td>219.47</td>
<td>24.92</td>
<td>32.06</td>
<td>19.18</td>
</tr>
<tr>
<td>No. of Observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Table 4 displays the results from the 2SLS estimates which correspond, column-by-column, to the first stage regressions reported in Table 3. I focus the discussion on the parameter of interest $\delta$. By instrumenting the spillover variable with the R&D of indirect connections located at distance 2 (column 1), $\delta$ is estimated at around 0.013, why the coefficient associated with the distance 3 instrument is estimated negative and statistically significant at the 10% level in column (2), while it is not significantly different from zero in the corresponding reduced form regression reported in Appendix E. Moran’s I statistic is the appropriate means for consistent estimation of spatial correlation patterns.
<table>
<thead>
<tr>
<th>Table 4: Production Function, Two Stages Least Squares Estimates, 1981-2001</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Private R&amp;D (γ)</strong></td>
</tr>
<tr>
<td>Private R&amp;D (γ)</td>
</tr>
<tr>
<td>R&amp;D Spillovers (δ)</td>
</tr>
<tr>
<td>Geographic Spillovers</td>
</tr>
<tr>
<td>Capital</td>
</tr>
<tr>
<td>Labor</td>
</tr>
<tr>
<td>Jaffe Tech. Proximity</td>
</tr>
<tr>
<td>Spillovers IV(s)</td>
</tr>
<tr>
<td>Fixed Effects</td>
</tr>
<tr>
<td>Only Network</td>
</tr>
<tr>
<td>No. of Communities</td>
</tr>
</tbody>
</table>

Notes: The table reports IV-2SLS estimates of model (11). All estimates are restricted to firms with at least one nonzero connection \( (g_{ij})_{t} \neq 0 \) in any year \( t \); all observations of these firms are also included for years with no connections. Models in columns 1 and 2 employ the distance 2 instrument; models in columns 2 through 5 employ the distance 3 instrument. All estimates include firm and year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with \( \varphi = 0.8 \) (10 communities) in column 4 and \( \varphi = 0.6 \) (20 communities) in column 5. Standard errors are clustered by the 20 “communities” obtained via the Louvain algorithm with \( \varphi = 0.6 \) (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. Asterisks denote conventional significance levels of \( t \)-tests (* \( p < 0.1 \); ** \( p < 0.05 \); *** \( p < 0.01 \)).

A figure slightly larger than the corresponding OLS estimate. When including both instruments (column 2) the result is again similar. By instrumenting only for the R&D of indirect connections located at distance 3 (column 3), the result is instead different: the point estimate of \( \delta \) is substantially higher, hovering around 0.020. Interestingly, the inclusion of community-by-year effects results in even larger estimates: \( \hat{\delta} \approx 0.023 \) with 10 communities (column 4) and \( \hat{\delta} \approx 0.021 \) with 20 communities (column 5). All estimates of \( \delta \) are statistically significant at the 5% level. The parameters associated with the other variables are estimated similarly to the OLS case.
4.3 System GMM Estimates

A possible concern about these results is that since the private input variables (capital, labor, private R&D) are also endogenous, the corresponding elasticities may be estimated incorrectly, possibly affecting the IV/2SLS estimates of \( \delta \). An implication of Proposition 2 from the analytical framework is that, for \( D = C + 2L + 1 \), the instrument would be uncorrelated with all the firm-level inputs, which would result in consistent estimates of \( \delta \) even if said inputs are endogenous.\(^{31}\) However, this may not hold in practice for all eligible distances \( D \). To overcome the problem, I assume that the residual shock \( v_{it} \) presents an AR(1) time dependence structure with “innovation” \( \varepsilon_{it} \):

\[
v_{it} = \rho v_{i(t-1)} + \varepsilon_{it}, \quad \rho \in (0, 1)
\]

and I estimate the model by performing System GMM on the \( \rho \)-differenced version of (11), following Blundell and Bond (1998, 2000).\(^{32}\) Specifically, I let conventional inputs, internal R&D and external R&D to be correlated with \( \varepsilon_{it} \); I employ third and higher lags of both their levels and first differences as instruments (in the differences and level equations, respectively). Conversely, the other spillover and demand side controls are either lagged quantities or factors outside the control of firms, thus I treat them as predetermined and independent of \( \varepsilon_{it} \). Following a customary approach, I recover the structural parameters of interest from the steady-state representation of the \( \rho \)-differenced equation; these estimates are displayed in Table 5. In Appendix E I illustrate the estimation procedure in more detail; in addition I report and discuss the System GMM estimates of the \( \rho \)-differenced model.

Column (1) of Table 5 displays the baseline results from System GMM estimation of the model, obtained by instrumenting the spillover variable with appropriate “GMM style” lags. The elasticities of the private inputs are estimated differently with respect to their OLS or 2SLS counterparts: as expected, the labor elasticity shrinks while the capital and private R&D elasticities are larger. By contrast, the estimate of \( \delta \) (around

\(^{31}\)The intuition is that if the standard inputs are endogenous, their bias is transmitted to \( \delta \) only if they are correlated with the instrument. Note that the elasticities of endogenous inputs would still be estimated inconsistently. These considerations are elaborated in more depth in Appendix A, as part of the discussion about the corollary to Proposition 2 that is mentioned earlier.

\(^{32}\)Building on some earlier observations by Wooldridge (2009), Ackerberg, Caves, and Frazer (2015) note that the \( \rho \)-differenced “dynamic panel” approach to the estimation of production functions bears important analogies with their semi-parametric control function method à la Olley and Pakes (1996) and Levinsohn and Petrin (2003). The two approaches make different assumptions about the unobserved shock \( \omega_{it} \), but nevertheless result in analogous moment conditions.
Table 5: Production Function, System GMM Estimates, 1981-2001

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

Notes: The table reports System GMM estimates of model (11). All estimates are restricted to firms with at least one nonzero connection ($g_{ij,t} \neq 0$) in any year $t$; all observations of these firms are also included for years with no connections. The $\rho$-differenced production function is estimated by one step System GMM: lags $s$ of the endogenous variables in levels are employed as instruments in the differenced equation; lags $s$ of the same variables in first differences are used as instruments in the level equation. In columns 2 through 5 the standard GMM instruments of the spillover variables for the level equation are substituted by moments (16) for the given value of $D$. All estimates include year fixed effects. Columns 4 and 5 include additional community-by-year fixed effects, where communities are obtained via the Louvain algorithm with $\varphi = 0.8$ (10 communities) in column 4 and $\varphi = 0.6$ (20 communities) in column 5. Structural parameters are recovered from the steady-state representation of the $\rho$-differenced model; standard errors are calculated by the Delta Method (see Appendix E for details). Standard errors are clustered by the 20 “communities” obtained via the Louvain algorithm with $\varphi = 0.6$ (small sample corrections are applied). All observations of the same individual firm in different years enter the same cluster. Asterisks denote conventional significance levels of $t$-tests ($^* p < 0.1$; $^{**} p < 0.05$; $^{***} p < 0.01$).

0.010) is not statistically significant from zero. Note that under the standard System GMM assumptions, $\delta$ is identified. Yet, the network’s characteristics may exacerbate problems of weak instruments type that are typical of this procedure: since one firm’s
connections may be very different after three years, standard GMM instruments may have low predictive power. To address this I substitute the spillover instruments in the level equation with those implied by the following moments:

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

which extend (14). These instruments have the effect of improving the statistical efficiency of the estimates. Column (2) reports the results from the inclusion of moments (16) with \( D = 2 \): \( \hat{\delta} \) increases to 0.014, but it is still not statistically significant. Moving to \( D = 3 \) (column 3) results in \( \hat{\delta} \simeq 0.017 \), now statistically significant at the 10% level. With the inclusion of community-by-year effects (columns 4-5) \( \delta \) is estimated at around 0.021 – about the same value as the corresponding 2SLS estimates – and is significant at the 5% level. Observe that the estimates of the capital, labor and private R&D elasticities from columns 4 and 5 take realistic and conventional values.

### 4.4 Discussion

In summary, both the 2SLS and System GMM estimates of the production function obtained with distance 3 instruments point to a value of \( \delta \) in a neighborhood of 0.021. For a connected pair of firms with a link of average strength, this implies a 0.0013% increase in firm sales following a 1% increase in external R&D. For a connected firm with an average row sum of connections, the implied elasticity is about 0.0105 given a 1% increase in the R&D stock of all its linkages. This estimate of \( \delta \) is about twice as large as the corresponding value obtained by performing simple OLS on the same subsample. The difference is only apparent when employing the distance 3 instrument in isolation, suggesting that the distance 2 instrument might be itself correlated with the unobserved characteristics of firms. In light of Proposition 2, this fact is consistent with the hypothesis that the spatial correlation of R&D is driven by exogenous factors rather than by the endogenous reflection of shocks: \((C, L) = (2, 0)\).

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

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

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

(17)

this illustrates how, in general, the MSR exceeds the MPR by an amount that depends on the strength of connections. To calculate aggregate values for the MPR and MSR, I take the average of both expressions in (17) by pooling firms with nonzero connections ($\bar{g}_{it} > 0$) over all years, and using estimates $\hat{\gamma} \simeq 6.68\%$ and $\hat{\delta} \simeq 2.07\%$ from column (5) of Table 5. This exercise results in a network-wide MPR evaluated at around 102%, and in a corresponding aggregate MSR approximately equal to 114%: about 112% of the MPR.33 These are realistic and economically significant values, comparable with evaluations from other studies (see Hall et al., 2010).34

33Note that these are the calculated returns from the R&D stock. To estimate the returns from annual R&D expenditures, one should divide these figures by the steady-state flow/stock ratio. Under the typical assumption of a 0.20 steady-state ratio, one obtains approximately a 20% private return and a 23% social return from yearly R&D expenditures. Within the network subsample pooled over years, the average flow/stock ratio is 0.183 with 0.118 standard deviation.

34Note that excluding firms without connections from the calculation results in a higher MSR by construction, because those firms do not receive spillovers. However, connected firms are larger and more R&D intensive, and they represent a disproportionally larger share of the U.S. economy.
I conclude this section with a brief mention of the additional results reported in Appendix E. First, the estimation of the model on selected cuts of the data suggests that connection-mediated spillovers appear to be not only prevalent among smaller firms, but also heterogeneous across industries (although by no means restricted to high-tech sectors). Second, employing different measures of connections yields quite similar results, unless the alternative metrics depart too much from the square-root measure \(\delta\) – in which case \(\delta\) can be estimated very imprecisely. Finally, the extension of the proposed identification strategy to models of firm market value and patent output that depend on the connection-based measure of spillovers results in patterns similar to that of production function estimates. In particular, the typical elasticity of Tobin’s \(q\) from a 1% increase in the R&D of all connections is about 0.017; this is obtained from IV estimates with instruments based on distances two or three, and is about 46% higher than the one derived from OLS estimates. In the citation-weighted patents model, a control function approach estimates the analogous elasticity more than tripled relative to the baseline estimates, but only when using the distance three instrument in isolation; in this case, the elasticity is measured as about 0.057.

5 Conclusion

In this paper I propose a new method for evaluating R&D spillovers. By aggregating information on patent collaboration relationships between individuals that work for different organizations, I construct a network of firms that are reciprocally connected through their respective R&D teams. I evaluate the dependence of firm productivity as well as of other firm-level outcomes from the R&D performed by firms connected in the network, weighted by the intensity of mutual links. Concerned by the possible presence of common confounders that simultaneously drive R&D choices, the intensity of cross-firm connections as well as firm-level outcomes, I employ a novel identification strategy based on the network topology. In particular, I instrument the R&D choices of one firm’s direct connections with those of sufficiently distant indirect links. Under conditions specified by a formal model of firms’ interaction, appropriately constructed instrumental variables predict the intensity of spillovers received by one firm, but are otherwise unrelated to its performance and innovation outcomes.

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

References


