R&D Networks: Theory, Empirics and Policy Implications☆

Michael D. König[a], Xiaodong Liu[b], Yves Zenou[c]

[a]*Department of Economics, University of Zurich, Schönberggasse 1, CH-8001 Zurich, Switzerland.*
[b]*Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309-0256, United States.*
[c]*Department of Economics, Monash University, Caulfield VIC 3145, Australia, and IFN.*

Abstract

We analyze a model of R&D alliance networks where firms are engaged in R&D collaborations that lower their production costs while competing on the product market. We provide a complete characterization of the Nash equilibrium and determine the optimal R&D subsidy program that maximizes total welfare. We then structurally estimate this model using a unique panel of R&D collaborations and annual company reports. We use our estimates to study the impact of targeted vs. non-discriminatory R&D subsidy policies and empirically rank firms according to the welfare-maximizing subsidies they should receive.

Key words: R&D networks, innovation, spillovers, optimal subsidies, industrial policy
JEL: D85, L24, O33

☆We would like to thank the editor, two anonymous referees, Philippe Aghion, Ufuk Akcigit, Coralio Ballester, Francis Bloch, Nick Bloom, Stefan Bühler, Guido Cozzi, Greg Crawford, Andrew F. Daughety, Marcel Fafchamps, Alfonso Gambardella, Christian Helmers, Han Hong, Matt O. Jackson, Chad Jones, Art Owen, Jennifer Reiganum, Michelle Sovinsky, Adam Szeidl, Nikolas Tsakas, Bastian Westbrock, Fabrizio Zilibotti, and seminar participants at Cornell University, University of Zurich, University of St.Gallen, Utrecht University, Stanford University, MIT, University College London, University of Washington, the NBER Summer Institute's Productivity/Innovation Meeting, the PEPA/cemmap workshop on Microeconomic Applications of Social Networks Analysis, the Public Economic Theory Conference, the IZA Workshop on Social Networks in Bonn and the CEPR Workshop on Moving to the Innovation Frontier in Vienna for their helpful comments. We further thank Nick Bloom, Christian Helmers and Lalvani Peter for data sharing, Enghin Atalay and Ali Hortacsu for sharing their name matching algorithm with us, and Sebastian Ottinger for the excellent research assistance. Michael D. König acknowledges financial support from Swiss National Science Foundation through research grants PBEZP1–131169 and 100018_140266, and thanks SIEPR and the Department of Economics at Stanford University for their hospitality during 2010-2012. Yves Zenou acknowledges financial support from the Swedish Research Council (Vetenskaprådet) through research grant 421–2010–1310.

Email addresses: michael.koenig@econ.uzh.ch (Michael D. König), xiaodong.liu@colorado.edu (Xiaodong Liu), yves.zenou@monash.edu (Yves Zenou)

1. Introduction

R&D collaborations have become a widespread phenomenon especially in industries with a rapid technological development such as the pharmaceutical, chemical and computer industries [cf. Hagedoorn, 2002; Roijakkers and Hagedoorn, 2006]. Through such collaborations firms generate R&D spillovers not only to their direct collaboration partners but also indirectly to other firms that are connected to them within a complex network of R&D collaborations. At the same time an increasing number of countries have resorted to various financial policies to stimulate R&D investments by private firms [cf. e.g. Cohen, 1994; Czarnitzki et al., 2007]. In particular, OECD countries spend more than 50 billion dollars per year on such R&D policies [cf. Takalo et al., 2017], including direct R&D subsidies and R&D tax credits.[1] The aim of this paper is to develop and structurally estimate an R&D network model and to empirically evaluate different R&D subsidy policies that take spillovers in R&D networks into account.

We consider a general model of competition à la Cournot where firms choose both their R&D expenditures and output levels. Firms can reduce their costs of production by exerting R&D efforts. We characterize the Nash equilibrium of this game for any type of R&D collaboration network as well as for any type of competition structure between firms (Proposition 1). We show that there exists a key trade-off faced by firms between the *technology (or knowledge) spillover effect* of R&D collaborations and the *product rivalry effect* of competition. The former effect captures the *positive* impact of R&D collaborations on output and profit while the latter captures the *negative* impact of competition and market stealing effects.

Due to the existence of externalities through technology spillovers and competition effects that are not internalized in the R&D decisions of firms, the social benefits of R&D differ from the private returns of R&D. This creates an environment where government funding programs that aim at fostering firms' R&D activities can be welfare improving. We analyze the optimal design of such R&D subsidy programs (where a planner can subsidize a firm's R&D effort) that take into account the network externalities in our model. We derive an exact formula for any type of network and competition structure that determines the optimal amount of subsidies per unit of R&D effort that should be given to each firm. We discriminate between homogeneous

[1]Different papers have evaluated how effective these policies are. See e.g. Zunica-Vicente et al. [2014] for an overview of this literature.

subsidies (Proposition 2), where each firm obtains the same amount of subsidy per unit of R&D effort and targeted subsidies (Proposition 3), where subsidies are firm specific.

We then bring the model to the data by using a unique panel of R&D collaborations and annual company reports over different sectors, regions and years. We adopt an instrumental variable (IV) strategy to estimate the best-response function implied by the theoretical model to identify the *technology (or knowledge) spillover effect* of R&D collaborations and the *product rivalry effect* of competition in a panel data model with both firm and time fixed effects. In particular, following Bloom et al. [2013], we use changes in the firm-specific tax price of R&D to construct IVs for R&D expenditures. Furthermore, to address the potential endogeneity of R&D networks, we use predicted R&D networks based on predetermined dyadic characteristics to construct IVs to identify the casual effect of R&D spillovers. As predicted by the theoretical model, we find that the spillover effect has a *positive* and significant impact on output and profit while the competition effect has a *negative* and significant impact.

Using our estimates and following our theoretical results, we then empirically determine the optimal subsidy policy, both for the homogenous case where all firms receive the same subsidy per unit of R&D effort, and for the targeted case, where the subsidy per unit of R&D effort may vary across firms. The targeted subsidy program turns out to have a much higher impact on total welfare as it can improve welfare by up to 80%, while the homogeneous subsidies can improve total welfare only by up to 4%. We then empirically rank firms according to the welfare-maximizing subsidies that they receive by the planner. We find that the firms that should be subsidized the most are not necessarily the ones that have the highest market share, the largest number of patents or the most central position in the R&D network. Indeed, these measures can only partially explain the ranking of firms that we find, as the market share is more related to the product market rivalry effect, while the R&D network and the patent stocks are more related to the technology spillover effect, and both effects are incorporated in the design of the optimal subsidy program.

The rest of the paper is organized as follows. In Section 2, we compare our contribution to the existing literature. In Section 3, we develop our theoretical model, characterize the Nash equilibrium of this game, and define the total welfare. Section 4 discusses optimal R&D subsidies. Section 5 describes the data. Section 6 is divided into four parts. In Section 6.1, we define the econometric specification of our model while, in Section 6.2, we highlight our identification strategy. The estimation results are given in Section 6.3. Section 6.4 provides a robustness check. The policy results of our empirical analysis are given in Section 7. We discuss our main assumptions in Section 8. Finally, Section 9 concludes. In the Online Appendix, we provide the proofs of the propositions (Appendix A), introduce the network definitions and characterizations used throughout the paper (Appendix B), highlight the contribution of our model with respect to the literature on games on networks (Appendix C), discuss the Herfindahl concentration index (Appendix D), perform an analysis in terms of Bertrand competition instead of Cournot competition (Appendix E), provide a theoretical model of direct and indirect technology spillovers (Appendix F), determine market failures due to technological externalities that are not internalized by the firms and investigate the optimal network structure of R&D collaborations (Appendix G), give a detailed description of how we construct and combine our different datasets for the empirical analysis (Appendix H), provide a numerical algorithm for computing optimal subsidies (Appendix I) and, finally, provide some additional robustness checks for the empirical analysis (Appendix J).

2. Related Literature

Our theoretical model analyzes a game with strategic complementarities where firms decide about production and R&D effort by treating the network as exogenously given. Thus, it belongs to a particular class of games known as *games on networks* [cf. Jackson and Zenou, 2015].[2,3]

Compared to this literature, we develop an R&D network model where competition between firms is explicitly modeled, not only within the same product market but also across different product markets (see Proposition 1). This yields very general results that can encompass any possible network of collaborations and any possible market interaction structure of competition between firms. We also provide an explicit welfare characterization and determine which network maximizes total welfare in certain parameter ranges (see Proposition 4 in the Online

[2]The economics of networks is a growing field. For recent surveys of the literature, see Jackson [2008] and Jackson et al. [2017].

[3]Two prominent papers in this literature are that of Ballester et al. [2006] and Bramoullé et al. [2014]. In the Online Appendix C, we discuss in detail the differences between our model and theirs.

Appendix G). To the best of our knowledge, this is one of the first papers that provides such an analysis.[4]

We also perform a policy analysis of R&D subsidies that consists in subsidizing firms' R&D costs. We are able to determine the optimal subsidy levels both, when it is homogenous across firms (Proposition 2) and when it is targeted to specific firms (Proposition 3). We are not aware of any other studies of subsidy policies in the context of R&D collaboration networks.[5]

In the industrial organization literature, there is a long tradition of models that analyze product and price competition with R&D collaborations (see, e.g. D'Aspremont and Jacquemin [1988] and Suzumura [1992]). One of their main insights is that the incentives to invest in R&D are reduced by the presence of such technology spillovers. In this literature, however, there is no explicit network of R&D collaborations. The first paper that provides an explicit analysis of R&D networks is that by Goyal and Moraga-Gonzalez [2001]. The authors introduce a strategic Cournot oligopoly game in the presence of externalities induced by a network of R&D collaborations. Benefits arise in these collaborations from sharing knowledge about a cost-reducing technology. However, by forming collaborations, firms also change their own competitive position in the market as well as the overall market structure. Thus, there exists a two-way flow of influence from the market structure to the incentives to form R&D collaborations and, in turn, from the formation of collaborations to the market structure. Westbrock [2010] extends their framework to analyze welfare and inequality in R&D collaboration networks, but abstracts from R&D investment decisions. Even though we do not study network formation as, for example, in Goyal and Moraga-Gonzalez [2001], compared to these papers, we are able to provide results for all possible networks with an arbitrary number of firms and a complete characterization of equilibrium output and R&D effort choices in multiple interdependent markets. We also determine policies related to network design and optimal R&D subsidy programs.

From an econometric perspective, there has been a significant progress in the literature on identification and estimation of social network models recently (see Blume et al. [2011] and Chandrasekhar [2016], for recent surveys). One of the most popular models in applied

[4]An exception is the recent paper by Belhaj et al. [2016], who study network design in a game on networks with strategic complements, but neglect competition effects (global substitutes).

[5]There are papers that look at subsidies in industries with technology spillovers but the R&D network is not explicitly modeled. See e.g. Acemoglu et al. [2012]; Akcigit [2009]; Bloom et al. [2002]; Hinloopen [2001]; Leahy and Neary [1997]; Spencer and Brander [1983].

research is the linear social network models. Bramoullé et al. [2009] provide identification conditions for this model based on the intransitivities in the network structure and propose an IV-based estimation strategy exploiting exogenous characteristics of indirect connections. This estimation strategy gains its popularity due to its simplicity. Yet, the validity of the IVs relies on the assumption that the network structure captured by the adjacency matrix is exogenous. If the adjacency matrix depends on some unobserved variables that are correlated with the error term of the social interaction regression, then the adjacency matrix is endogenous and this IV-based estimator would be inconsistent. In this paper, taking advantage of the panel data structure in the empirical analysis, we introduce both firm and time fixed effects into the linear social network model to attenuate the potential asymptotic bias caused by the endogenous adjacency matrix. To further reduce this potential bias, we use the predicted adjacency matrix based on predetermined dyadic characteristics (instead of the observed adjacency matrix) to construct IVs for this model. This allows us to estimate the causal impact of R&D spillovers.

There is a large empirical literature on technology spillovers [see e.g. Bloom et al., 2013; Einiö, 2014; Griffith et al., 2004; Singh, 2005], and R&D collaborations [see e.g. Hanaki et al., 2010]. There is also an extensive literature that estimates the effect of R&D subsidies on private R&D investments and other measures of innovative performance (see e.g. Bloom et al. [2002], Feldman and Kelley [2006], Dechezleprêtre et al. [2016], and, for a survey, see Klette et al. [2000]). However, to the best of our knowledge, our paper is the first that provides a ranking of firms according to the welfare maximizing subsidies that they should receive. We show, in particular, that the highest subsidized firms are not necessarily those with the largest market share, a larger number of patents or the highest (betweenness, eigenvector or closeness) centrality in the network of R&D collaborations. We find, however, that larger firms should receive higher subsidies than smaller firms as they generate more R&D spillovers. This result is in line with that of Bloom et al. [2013] who also find that smaller firms generate lower social returns to R&D because they operate more in technological niches.

Furthermore, contrary to Akcigit [2009] and Acemoglu et al. [2012], we do not focus on entry and exit but instead incorporate the network structure of R&D collaborating firms. This allows us to take into account the R&D spillover effects of incumbent firms, which are typically ignored in studies of the innovative activity of incumbent firms versus entrants. Therefore, we see our analysis as complementary to that of Akcigit [2009] and Acemoglu et al. [2012], and we show that R&D subsidies can trigger considerable welfare gains when technology spillovers through R&D alliances are incorporated.

3. Theoretical Framework

3.1. Network Game

We consider a general Cournot oligopoly game where a set of firms $\mathcal{N} = \{1, \ldots, n\}$ is partitioned in $M \geq 1$ heterogeneous product markets $\mathcal{M}_m$, $m = 1, \ldots, M$. Let $|\mathcal{M}_m|$ denote the size of market $\mathcal{M}_m$. We allow for consumption goods to be imperfect substitutes (and thus differentiated products) by adopting the consumer utility maximization approach of Singh and Vives [1984]. We first consider q_i the demand for the good produced by firm i in market $\mathcal{M}_m$. A representative consumer in market $\mathcal{M}_m$ obtains the following gross utility from consumption of the goods $\{q_i\}_{i\in\mathcal{M}_m}$

$$\bar{U}_m(\{q_i\}_{i\in\mathcal{M}_m}) = \alpha_m \sum_{i\in\mathcal{M}_m} q_i - \frac{1}{2}\sum_{i\in\mathcal{M}_m} q_i^2 - \frac{\rho}{2}\sum_{i\in\mathcal{M}_m}\sum_{j\in\mathcal{M}_m, j\neq i} q_i q_j.$$

In this formulation, the parameter α_m captures the heterogeneity in market sizes, whereas $\rho \in [0, 1)$ measures the degree of substitutability between products. In particular, $\rho \to 1$ depicts a market of perfectly substitutable goods, while $\rho = 0$ represents the case of local monopolies.

The consumer maximizes net utility $U_m = \bar{U}_m - \sum_{i\in\mathcal{M}_m} p_i q_i$, where p_i is the price of good i. This gives the inverse demand function for firm i

$$p_i = \bar{\alpha}_i - q_i - \rho \sum_{j\in\mathcal{M}_m, j\neq i} q_j, \tag{1}$$

where $\bar{\alpha}_i = \sum_{m=1}^{M} \alpha_m \mathbb{1}_{\{i\in\mathcal{M}_m\}}$. In the model, we will study both the general case where $\rho > 0$ but also the special case where $\rho = 0$. The latter case is when firms are local monopolists so that the price of the good produced by each firm i is only determined by its own quantity q_i (and the size of the market) but not by the quantities of other firms, i.e. $p_i = \bar{\alpha}_i - q_i$.

Firms can reduce their production costs by investing in R&D as well as by benefiting from an R&D collaboration with another firm.[6] The amount of this cost reduction depends on the R&D effort e_i of firm i and the R&D efforts of the R&D collaboration partners of firm i. Given the effort level e_i, the marginal cost c_i of firm i is given by:[7]

$$c_i = \bar{c}_i - e_i - \varphi \sum_{j=1}^{n} a_{ij} e_j, \tag{2}$$

The network of R&D collaborations, G, can be represented by a symmetric $n \times n$ *adjacency matrix* $\mathbf{A}$. Its elements $a_{ij} \in \{0, 1\}$ indicate whether there exists a link between nodes i and j.[8] In the context of our model, $a_{ij} = 1$ if firms i and j have an R&D collaboration and $a_{ij} = 0$ otherwise. As a normalization, we set $a_{ii} = 0$. In Equation (2), the total cost reduction for firm i stems from its own research effort e_i and the research effort of all other collaborating firms (via *knowledge spillovers*), which is captured by the term $\sum_{j=1}^{n} a_{ij} e_j$, where $\varphi \geq 0$ is the marginal cost reduction due to a collaborator's R&D effort. We assume that R&D effort is costly. In particular, the cost of R&D effort is given by $\frac{1}{2}e_i^2$, which is increasing in effort and exhibits decreasing returns. Firm i's profit is then given by

$$\pi_i = (p_i - c_i)q_i - \frac{1}{2}e_i^2. \tag{3}$$

Inserting the inverse demand from Equation (1) and the marginal cost from Equation (2) into Equation (3) gives the following strictly quasi-concave profit function for firm i

$$\pi_i = (\bar{\alpha}_i - \bar{c}_i)q_i - q_i^2 - \rho \sum_{j=1}^{n} b_{ij} q_i q_j + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2}e_i^2, \tag{4}$$

where $b_{ij} = 1$ if firms i and j operate in the same market and $b_{ij} = 0$ otherwise. Consequently, the market structure can be represented by an $n \times n$ competition matrix $\mathbf{B} = [b_{ij}]$. If we arrange firms by markets they operate in, the competition matrix $\mathbf{B}$ will be a block diagonal matrix with a zero diagonal and blocks of sizes $|\mathcal{M}_m|$, $m = 1, \ldots, M$. An illustration can be found below.

[6] For example, Bernstein [1988] finds that R&D spillovers decrease the unit costs of production for a sample of Canadian firms.

[7] We assume that the R&D effort independent marginal cost $\bar{c}_i$ is large enough such that marginal costs, c_i, are always positive for all firms $i \in \mathcal{N}$.

[8] See Online Appendix B.1 for definitions and characterizations of networks.

$$
\mathbf{B} = \left(\begin{array}{cccc|cccc|c}
0 & 1 & \cdots & 1 & 0 & \cdots & \cdots & 0 & \cdots \\
1 & 0 & \cdots & \vdots & \vdots & & & \vdots & \\
\vdots & \vdots & \ddots & 1 & \vdots & & & \vdots & \\
1 & \cdots & 1 & 0 & 0 & \cdots & \cdots & 0 & \cdots \\
\hline
0 & \cdots & \cdots & 0 & 0 & 1 & \cdots & 1 & \\
\vdots & & & \vdots & 1 & 0 & \cdots & \vdots & \\
\vdots & & & \vdots & \vdots & \vdots & \ddots & 1 & \\
0 & \cdots & \cdots & 0 & 1 & \cdots & 1 & 0 & \\
\hline
\vdots & & & \vdots & & & & & \ddots
\end{array}\right)_{n\times n}
$$

3.2. Nash Equilibrium

We consider quantity competition among firms à la Cournot.[9] The following proposition establishes the Nash equilibrium where each firm i simultaneously chooses *both* its output q_i and R&D effort e_i in an arbitrary network of R&D collaborations represented by the adjacency matrix $\mathbf{A}$ and an arbitrary market structure represented by the competition matrix $\mathbf{B}$. Throughout the paper, denote by $\mathbf{I}$ the $n \times n$ identity matrix, $\boldsymbol{\iota}$ the $n \times 1$ vector of ones, and $\lambda_{\max}(\mathbf{A})$ the largest eigenvalue of $\mathbf{A}$. Denote by $\mu_i \equiv \bar{\alpha}_i - \bar{c}_i$ for all $i \in \mathcal{N}$, and $\boldsymbol{\mu}$ the corresponding $n \times 1$ vector with components μ_i. Denote also by $\underline{\mu} = \min_i \{\mu_i \mid i \in \mathcal{N}\}$ and $\overline{\mu} = \max_i \{\mu_i \mid i \in \mathcal{N}\}$, with $0 < \underline{\mu} \leq \overline{\mu}$. Finally, denote by $\mathbf{b}_{\boldsymbol{\mu}}(G, \phi) \equiv (\mathbf{I} - \phi\mathbf{A})^{-1}\boldsymbol{\mu}$ the vector of $\boldsymbol{\mu}$-weighted Katz-Bonacich centralities, and $\mathbf{b}_{\boldsymbol{\iota}}(G, \phi) \equiv (\mathbf{I} - \phi\mathbf{A})^{-1}\boldsymbol{\iota}$ the vector of unweighted Katz-Bonacich centralities, where $\phi = \varphi/(1-\rho)$. [10]

Proposition 1. *Consider the n-player simultaneous-move game with the payoff given by Equation (4), where $\varphi \geq 0$, $0 \leq \rho < 1$ and $0 < \underline{\mu} \leq \mu_i \equiv \bar{\alpha}_i - \bar{c}_i \leq \overline{\mu}$.*

(i) If $\varphi = 0$ or

$$
\varphi\lambda_{\max}(\mathbf{A}) + \rho \max_{m=1,\dots,M} \{|\mathcal{M}_m| - 1\} < 1 \tag{5}
$$

then there exists a unique Nash equilibrium with the equilibrium R&D efforts $\mathbf{e}^*$ and outputs $\mathbf{q}^*$ given by

$$
\mathbf{e}^* = \mathbf{q}^* = (\mathbf{I} - \varphi\mathbf{A} + \rho\mathbf{B})^{-1}\boldsymbol{\mu}. \tag{6}
$$

[9] In the Online Appendix E we show that the same functional forms for best response quantities and efforts can be obtained for price setting firms under Bertrand competition as we find them in the case of Cournot competition.

[10] The proof of Proposition 1 is given in the Online Appendix A. See the Online Appendix B.3 for a precise definition of the Katz-Bonacich centrality used in the proposition.

and the equilibrium profits π_i^* *given by*

$$\pi_i^* = \frac{1}{2}(q_i^*)^2, \quad \forall i \in \mathcal{N}. \tag{7}$$

(ii) If $\phi \equiv \varphi/(1-\rho) < \lambda_{\max}(\mathbf{A})^{-1}$, *then there exists a unique Nash equilibrium in the case when all firms operate in a single market (i.e.,* $M = 1$*), with the equilibrium R&D efforts* $\underline{\mathbf{e}}^*$ *and outputs* $\underline{\mathbf{q}}^*$ *given by*

$$\underline{\mathbf{e}}^* = \underline{\mathbf{q}}^* = \frac{1}{1-\rho}\left(\mathbf{b}_{\boldsymbol{\mu}}(G,\phi) - \frac{\rho\,\|\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)\|_1}{(1-\rho)+\rho\,\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\|_1}\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right). \tag{8}$$

In addition, if

$$\phi\lambda_{\max}(\mathbf{A}) + \frac{n\rho}{1-\rho}\left(\frac{\overline{\mu}}{\underline{\mu}} - 1\right) < 1 \tag{9}$$

then, $\underline{\mathbf{e}}^* = \underline{\mathbf{q}}^* > \mathbf{0}$.

(iii) If $\varphi < \lambda_{\max}(\mathbf{A})^{-1}$, *then there exists a unique Nash equilibrium in the case when goods are non-substitutable (i.e.,* $\rho = 0$*), with the equilibrium R&D efforts* $\overline{\mathbf{e}}^*$ *and outputs* $\overline{\mathbf{q}}^*$ *given by* $\overline{\mathbf{e}}^* = \overline{\mathbf{q}}^* = \mathbf{b}_{\boldsymbol{\mu}}(G,\varphi) = (\mathbf{I} - \varphi\mathbf{A})^{-1}\boldsymbol{\mu} > \mathbf{0}$.

(iv) If the conditions stated in (i)-(iii) hold, then $\overline{\mathbf{e}}^* = \overline{\mathbf{q}}^* \geq \mathbf{e}^* = \mathbf{q}^* \geq \underline{\mathbf{e}}^* = \underline{\mathbf{q}}^* > \mathbf{0}$, *where* $\mathbf{e}^* = \mathbf{q}^*$ *is the vector of equilibrium outputs in the general case given by Equation (6).*

Proposition 1 (i) characterizes the Nash equilibrium for the most general case with a general R&D network and product market structure, while (ii) and (iii) characterize the equilibria of two special cases, namely, the case where all firms operate in the same market and the case where goods are non-substitutable, which provide the lower and upper bounds for the equilibrium in the general case as shown in (iv).

The first-order condition of profit maximization with respect to the R&D effort leads to

$e_i = q_i$,[11] while the first-order condition with respect to the output leads to

$$q_i = \mu_i + \varphi \sum_{j=1}^{n} a_{ij} q_j - \rho \sum_{j=1}^{n} b_{ij} q_j, \tag{10}$$

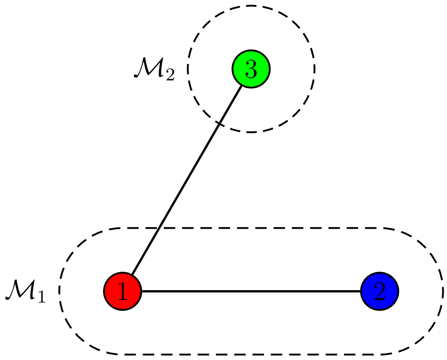
or, in matrix form, $\mathbf{q} = \boldsymbol{\mu} + \varphi \mathbf{A}\mathbf{q} - \rho \mathbf{B}\mathbf{q}$. If $\varphi = 0$ and $0 \leq \rho < 1$, or if the condition given by Equation (5) holds, the matrix $\mathbf{I} - \varphi \mathbf{A} + \rho \mathbf{B}$ is positive definite, and thus there exists a unique Nash equilibrium characterized by Equation (6). This result generalizes those of Ballester et al. [2006], Calvó-Armengol et al. [2009] and Bramoullé et al. [2014] to allow agents to make multivariate choices on R&D effort and output levels in the presence of both network effects and competition effects.[12]

Furthermore, to provide bounds on the equilibrium output in the general case characterized by Equation (6), we consider two special cases, namely, the case where all firms operate in the same market and the case where goods are non-substitutable. In particular, the lower bound is given by the equilibrium output of the case where all firms operate in the same market, which is strictly positive if the condition given by Equation (9) holds. Observe that, when all firms are homogenous, i.e., $\mu_i = \mu$ for all $i \in \mathcal{N}$, then Equation (9) holds if $\phi \lambda_{\max}(\mathbf{A}) < 1$. On the other hand, everything else equal, the higher the discrepancy of marginal payoffs $\overline{\mu}/\underline{\mu}$, the lower is the level of network complementarities $\phi \lambda_{\max}(\mathbf{A})$ that are compatible with this condition. A similar condition is obtained in Calvó-Armengol et al. [2009].

More generally, the key insight of Proposition 1 is the interaction between the *network effect*, through the adjacency matrix $\mathbf{A}$, and the *market effect*, through the competition matrix $\mathbf{B}$, and this is why the first-order condition with respect to q_i given by Equation (10) takes both of them into account. To better understand this result, consider the following simple example where firms 1 and 2 as well as firms 1 and 3 are engaged in R&D collaborations. Suppose that there are two markets where firms 1 and 2 operate in the same market $\mathcal{M}_1$ while firm 3 operates alone in market $\mathcal{M}_2$ (see Figure 1). Then, the adjacency matrix $\mathbf{A}$ and the competition matrix

[11] The proportional relationship between R&D effort levels and outputs has been confirmed in a number of empirical studies [see e.g. Cohen and Klepper, 1996; Klette and Kortum, 2004]. In the data used in our empirical analysis, the Pearson product-moment correlation coefficient of R&D effort levels and outputs is 0.66, which indicates strong linearity between these two variables.

[12] In Online Appendix C we highlight the contribution of our model with respect to the literature on games on networks by, first, shutting the network effects, second, the competition effects, and then comparing our model to that of Ballester et al. [2006] and Bramoullé et al. [2014].



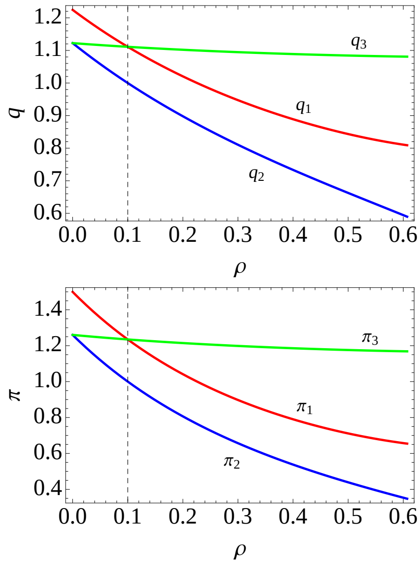


Figure 1: Equilibrium output from Equation (11) and profits for the three firms with $\mu = 1$, $\varphi = 0.1$ and varying values of the competition parameter $0 \leq \rho < 1 - \sqrt{2}\varphi$. Profits of firms 1 and 3 intersect at $\rho = \varphi$ (indicated with a dashed line).

$\mathbf{B}$ are given by

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with $\lambda_{\max}(\mathbf{A}) = \sqrt{2}$ and $\lambda_{\max}(\mathbf{B}) = 1$. Assume that firms are homogeneous such that $\mu_i = \mu$ for $i = 1, 2, 3$. Using Proposition 1, the condition for the existence of a unique Nash equilibrium with positive outputs is $\sqrt{2}\varphi + \rho < 1$. The equilibrium outputs are given by

$$\mathbf{q}^* = \mu(\mathbf{I} - \varphi\mathbf{A} + \rho\mathbf{B})^{-1}\boldsymbol{\iota} = \frac{\mu}{1 - 2\varphi^2 + 2\varphi\rho - \rho^2} \begin{pmatrix} 1 + 2\varphi - \rho \\ (1+\varphi)(1-\rho) \\ (1+\rho)(1+\varphi-\rho) \end{pmatrix}, \tag{11}$$

and equilibrium profits are given by $\pi_i^* = (q_i^*)^2/2$ for $i = 1, 2, 3$. Figure 1 shows an illustration of equilibrium outputs and profits for the three firms with $\mu = 1$, $\varphi = 0.1$ and varying values of the competition parameter $0 \leq \rho < 1 - \sqrt{2}\varphi$. We see that firm 1 has higher profits due to having the largest number of R&D collaborations when competition is weak (ρ is low compared to φ). However, when ρ increases, its profits decrease and become smaller than the profit of firm 3 when $\rho > \varphi$. This result highlights the key trade-off faced by firms between the *technology (or knowledge) spillover effect* and the *product rivalry effect* of R&D [cf. Bloom et al., 2013] since the former increases with φ, which captures the intensity of the spillover effect while the latter increases with ρ, which indicates the degree of competition in the product market.

3.3. Welfare

We next turn to analyzing welfare in the economy. Inserting the inverse demand from Equation (1) into net utility U_m of the consumer in market $\mathcal{M}_m$ shows that

$$U_m = \frac{1}{2}\sum_{i\in\mathcal{M}_m} q_i^2 + \frac{\rho}{2}\sum_{i\in\mathcal{M}_m}\sum_{j\in\mathcal{M}_m, j\neq i} q_i q_j.$$

For given quantities, the consumer surplus is strictly increasing in the degree ρ of substitutability between products. In the special case of non-substitutable goods, when $\rho = 0$, we obtain $U_m = \frac{1}{2}\sum_{i\in\mathcal{M}_m} q_i^2$, while in the case of perfectly substitutable goods, when $\rho \to 1$, we get $U_m = \frac{1}{2}\left(\sum_{i\in\mathcal{M}_m} q_i\right)^2$. The total consumer surplus is then given by $U = \sum_{m=1}^{M} U_m$. The producer surplus is given by aggregate profits $\Pi = \sum_{i=1}^{n}\pi_i$. As a result, the total welfare is equal to $W = U + \Pi$. Inserting profits as a function of equilibrium outputs from Equation (7) leads to the total welfare in the Nash equilibrium given by

$$W = \sum_{i=1}^{n}(q_i^*)^2 + \frac{\rho}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} q_i^* q_j^* = \mathbf{q}^{*\top}\mathbf{q}^* + \frac{\rho}{2}\mathbf{q}^{*\top}\mathbf{B}\mathbf{q}^*. \tag{12}$$

As welfare in Equation (12) is increasing in the output levels of the firms, it is clear that the higher the production levels of the firms, the higher is welfare.[13] Since output is proportional to R&D, this shows that there is a general problem of underinvestment in R&D (see also Online Appendix G.1). In the following section we therefore study the welfare gains from a policy that encourages firms to spend more on R&D.

4. R&D Subsidy Policies

Because of the externalities generated by R&D activities, market resource allocation will typically not be socially optimal. In Online Appendix G.1, we show that, indeed, there is a generic problem of under-investment in R&D, as the *private* returns from R&D are lower than the

[13] A discussion of how welfare is affected by the network structure can be found in the Online Appendix G.2. In particular, we investigate which network structure maximizes welfare.

social returns from R&D. A policy intervention can correct this market failure through R&D subsidy or tax programs. We extend our framework by considering an optimal R&D subsidy program that reduces the firms' R&D costs. For our analysis, we first assume that all firms obtain a homogeneous subsidy per unit of R&D effort spent. Then, we proceed by allowing the social planner to differentiate between firms and implement firm-specific R&D subsidies.[14]

4.1. Homogeneous R&D Subsidies

Following Spencer and Brander [1983] and Hinloopen [2000, 2001], an government (or planner) is introduced that can provide a subsidy, $s \in [0, \bar{s}]$ per unit of R&D effort for some $\bar{s} > 0$. It is assumed that each firm receives the same per unit R&D subsidy. With a homogeneous R&D subsidy, the profit of firm i given by Equation (4) becomes:

$$\pi_i = (\bar{\alpha}_i - \bar{c}_i)q_i - q_i^2 - \rho q_i \sum_{j=1}^{n} b_{ij} q_j + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2} e_i^2 + s e_i. \tag{13}$$

The game consists of two stages. In the first stage, the planner sets a subsidy rate on R&D effort, and in the second stage, the firms choose outputs and R&D efforts given the subsidy rate set in the first stage. The assumption that the planner can pre-commit itself to the subsidy rate and thus can act in this leadership role is fairly natural. In this context, the optimal R&D subsidy s^* determined by the planner is found by maximizing the total welfare $W(G, s)$ less the cost of the subsidy $s \sum_{i=1}^{n} e_i$, taking into account the fact that firms choose outputs and R&D efforts for a given subsidy rate by maximizing profits in Equation (13). If we define the net welfare as $\overline{W}(G, s) \equiv W(G, s) - s \sum_{i=1}^{n} e_i$, the social planner's problem is given by

$$s^* = \arg\max_{s \in [0, \bar{s}]} \overline{W}(G, s).$$

The following proposition characterizes the Nash equilibrium and derives the optimal subsidy rate that solves the planner's problem.[15]

Proposition 2. *Consider the n-player simultaneous-move game with the payoff given by*

[14] We would like to emphasize that, as we have normalized the cost of R&D to one in the profit function of Equation (3), the absolute values of R&D subsidies are not meaningful in the subsequent analysis, but rather relative comparisons across firms are.

[15] The proofs of Propositions 2 and 3 are given in Online Appendix A.

Equation (13), where $\varphi \geq 0$, $0 \leq \rho < 1$ and $0 < \underline{\mu} \leq \mu_i \equiv \bar{\alpha}_i - \bar{c}_i \leq \overline{\mu}$. Let $\mathbf{R} = (\mathbf{I} - \varphi\mathbf{A} + \rho\mathbf{B})^{-1}(\mathbf{I} + \varphi\mathbf{A})$ and $\mathbf{H} = \mathbf{I} + \mathbf{R} + \mathbf{R}^\top - 2\mathbf{R}^\top\mathbf{R} - \rho\mathbf{R}^\top\mathbf{B}\mathbf{R}$.

(i) If $\varphi = 0$ or the condition given by Equation (5) holds, then there exists a unique Nash equilibrium with the equilibrium outputs given by

$$\mathbf{q}^* = (\mathbf{I} - \varphi\mathbf{A} + \rho\mathbf{B})^{-1}\boldsymbol{\mu} + s\mathbf{R}\boldsymbol{\iota}, \tag{14}$$

the equilibrium R&D efforts given by

$$e_i^* = q_i^* + s, \quad \forall i \in \mathcal{N}, \tag{15}$$

and the equilibrium profits given by

$$\pi_i^* = \frac{(q_i^*)^2 + s^2}{2}, \quad \forall i \in \mathcal{N}. \tag{16}$$

(ii) If $\boldsymbol{\iota}^\top\mathbf{H}\boldsymbol{\iota} > 0$, the optimal subsidy level is given by

$$s^* = \frac{\boldsymbol{\iota}^\top(2\mathbf{R} + \rho\mathbf{B}\mathbf{R} - \mathbf{I})^\top(\mathbf{I} - \varphi\mathbf{A} + \rho\mathbf{B})^{-1}\boldsymbol{\mu}}{\boldsymbol{\iota}^\top\mathbf{H}\boldsymbol{\iota}}, \tag{17}$$

provided that $0 < s^* < \bar{s}$.

In part (i) of Proposition 2, we solve the second stage of the game where firms decide their outputs and R&D efforts given the homogenous subsidy s. In part (ii) of the proposition, we solve the first stage of the game where the planner optimally determines the subsidy rate. In the special case that goods are not substitutable, i.e. $\rho = 0$, the optimal subsidy level is $s^* = \boldsymbol{\iota}^\top(2\widetilde{\mathbf{R}} - \mathbf{I})^\top(\mathbf{I} - \varphi\mathbf{A})^{-1}\boldsymbol{\mu}/(\boldsymbol{\iota}^\top\widetilde{\mathbf{H}}\boldsymbol{\iota})$, given that $\boldsymbol{\iota}^\top\widetilde{\mathbf{H}}\boldsymbol{\iota} > 0$, where $\widetilde{\mathbf{R}} = (\mathbf{I} - \varphi\mathbf{A})^{-1}(\mathbf{I} + \varphi\mathbf{A})$ and $\widetilde{\mathbf{H}} = \mathbf{I} + \widetilde{\mathbf{R}} + \widetilde{\mathbf{R}}^\top - 2\widetilde{\mathbf{R}}^\top\widetilde{\mathbf{R}}$.

4.2. Targeted R&D Subsidies

We now consider the case where the planner can offer different subsidy rates to different firms, so that firm i, for all $i = 1, \ldots, n$, receives a subsidy $s_i \in [0, \bar{s}]$ per unit of R&D effort. Let $\mathbf{s}$

be an $n \times 1$ vector with components s_i. With target R&D subsidies, the profit of firm i given by Equation (4) becomes:

$$\pi_i = (\bar{\alpha}_i - \bar{c}_i)q_i - q_i^2 - \rho q_i \sum_{j=1}^{n} b_{ij} q_j + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2} e_i^2 + s_i e_i. \tag{18}$$

As in the case of homogenous subsidies, the optimal R&D subsidies $\mathbf{s}^*$ are found by maximizing the welfare $W(G, \mathbf{s})$ less the cost of the subsidy $\sum_{i=1}^{n} s_i e_i$, given that firms are choosing outputs and R&D efforts for a given subsidy level by maximizing profits in Equation (18). If we define the net welfare as $\overline{W}(G, \mathbf{s}) \equiv W(G, \mathbf{s}) - \sum_{i=1}^{n} e_i s_i$, then the solution to the social planner's problem is given by

$$\mathbf{s}^* = \arg\max_{\mathbf{s} \in [0, \bar{s}]^n} \overline{W}(G, \mathbf{s}).$$

The following proposition characterizes the Nash equilibrium and derives the optimal subsidy rate that solves the planner's problem.

Proposition 3. *Consider the n-player simultaneous-move game with the payoff given by Equation (18), where $\varphi \geq 0$, $0 \leq \rho < 1$ and $0 < \underline{\mu} \leq \mu_i \equiv \bar{\alpha}_i - \bar{c}_i \leq \overline{\mu}$. Let $\mathbf{R} = (\mathbf{I} - \varphi \mathbf{A} + \rho \mathbf{B})^{-1}(\mathbf{I} + \varphi \mathbf{A})$ and $\mathbf{H} = \mathbf{I} + \mathbf{R} + \mathbf{R}^\top - 2\mathbf{R}^\top \mathbf{R} - \rho \mathbf{R}^\top \mathbf{B} \mathbf{R}$.*

(i) If $\varphi = 0$ or the condition given by Equation (5) holds, then there exists a unique Nash equilibrium with the equilibrium outputs given by

$$\mathbf{q}^* = (\mathbf{I} - \varphi \mathbf{A} + \rho \mathbf{B})^{-1} \boldsymbol{\mu} + \mathbf{R}\mathbf{s}, \tag{19}$$

the equilibrium R&D efforts given by

$$\mathbf{e}^* = \mathbf{q}^* + \mathbf{s}, \tag{20}$$

and the equilibrium profits given by

$$\pi_i^* = \frac{(q_i^*)^2 + s_i^2}{2}, \quad \forall i \in \mathcal{N}. \tag{21}$$

(ii) If the matrix $\mathbf{H}$ *is positive definite, the optimal subsidy levels are given by*

$$\mathbf{s}^* = \mathbf{H}^{-1}(2\mathbf{R} + \rho\mathbf{B}\mathbf{R} - \mathbf{I})^\top(\mathbf{I} - \varphi\mathbf{A} + \rho\mathbf{B})^{-1}\boldsymbol{\mu}, \tag{22}$$

provided that $0 < s_i^* < \bar{s}$ *for all* $i = 1, \ldots, n$.

As in the previous proposition, in part (i) of Proposition 3, we solve for the second stage of the game where firms decide their outputs and R&D efforts given the *targeted* subsidy s_i. In part (ii), we solve the first stage of the game where the planner optimally decides the targeted subsidy rate.[16] In the special case that goods are not substitutable, i.e. $\rho = 0$, the optimal subsidy level is $\mathbf{s}^* = \widetilde{\mathbf{H}}^{-1}(2\widetilde{\mathbf{R}} - \mathbf{I})^\top(\mathbf{I} - \varphi\mathbf{A})^{-1}\boldsymbol{\mu}$, given that the matrix $\widetilde{\mathbf{H}}$ is positive definite, where $\widetilde{\mathbf{R}} = (\mathbf{I} - \varphi\mathbf{A})^{-1}(\mathbf{I} + \varphi\mathbf{A})$ and $\widetilde{\mathbf{H}} = \mathbf{I} + \widetilde{\mathbf{R}} + \widetilde{\mathbf{R}}^\top - 2\widetilde{\mathbf{R}}^\top\widetilde{\mathbf{R}}$.

In the following sections we will test the different parts of our theoretical predictions. First, we will test Proposition 1 and try to disentangle between the *technology (or knowledge) spillover effect* and the *product rivalry effect* of R&D. Second, once the parameters of the model have been estimated, we will use Propositions 2 and 3, respectively, to determine which firms should be subsidized, and how large their subsidies should be in order to maximize net welfare.

5. Data

To obtain a comprehensive picture of R&D alliances, we use data on interfirm R&D collaborations stemming from two sources that have been widely used in the literature [cf. Schilling, 2009]. The first one is the Cooperative Agreements and Technology Indicators (CATI) database [cf. Hagedoorn, 2002]. This database only records agreements for which a combined innovative activity or an exchange of technology is at least part of the agreement.[17] The second source is the Thomson Securities Data Company (SDC) alliance database. SDC collects data from the U.S. Securities and Exchange Commission (SEC) filings (and their international counterparts),

[16]Note that when the condition for positive definiteness is not satisfied then we can sill use part (ii) of Proposition 3, respectively, as a candidate for a welfare improving subsidy program. However, there might exist other subsidy programs which yield even higher welfare gains.

[17]Firms might benefit from each other's research beyond what is captured by the network of R&D collaborations. Thus, in Section 6.4, we also define R&D collaborations between firms more broadly by their degree of technological proximity.

trade publications, wires, and news sources. We include only alliances from SDC that are classified explicitly as R&D collaborations. The Online Appendix H.1 provides more information about the different R&D collaboration databases used for this study.

We then merged the CATI database with the Thomson SDC alliance database. For the matching of firms across datasets we used the name matching algorithm developed as part of the NBER patent data project [Atalay et al., 2011; Trajtenberg et al., 2009].[18] The merged datasets allow us to study patterns in R&D partnerships in several industries over an extended period of several decades.

Observe that because of our IV strategy (See Section 6.2.3 below), which is based on R&D tax credits in the U.S., we only consider U.S. firms as in Bloom et al. [2013].[19]

The systematic collection of inter-firm alliances started in 1987 and ended in 2006 for the CATI database. However, information about alliances prior to 1987 is available in both databases, and we use all information available starting from the year 1963 and ending in 2006.[20] We construct the R&D alliance network by assuming that an alliance lasts 5 years. In the Online Appendix (Section J.1), we conduct robustness checks with different specifications of alliance durations.

Some firms might be acquired by other firms due to mergers and acquisitions (M&A) over time, and this will impact the R&D collaboration network [cf. e.g. Hanaki et al., 2010]. We account for M&A activities by assuming that an acquiring firm inherits all the R&D collaborations of the target firm. We use two complementary data sources to obtain comprehensive information about M&As. The first is the Thomson Reuters' SDC M&A database, which has historically been the reference database for empirical research in the field of M&As. The second database for M&As is Bureau van Dijk's Zephyr database, which is an alternative to the SDC M&As database. A comparison and more detailed discussion of the two M&As databases can be found in the Online Appendix H.2.

Figure 2 shows the number of firms, n, participating in an alliance in the R&D network, the

[18]See `https://sites.google.com/site/patentdataproject`. We thank Enghin Atalay and Ali Hortacsu for making their name matching algorithm available to us.

[19]In the working paper version, König et al. [2014], we also consider non-U.S. firms, but with a different estimation strategy.

[20]Fama and French [1992] note that Compustat suffers from a large selection bias prior to 1962, and we discard any data prior to 1962 from our sample.

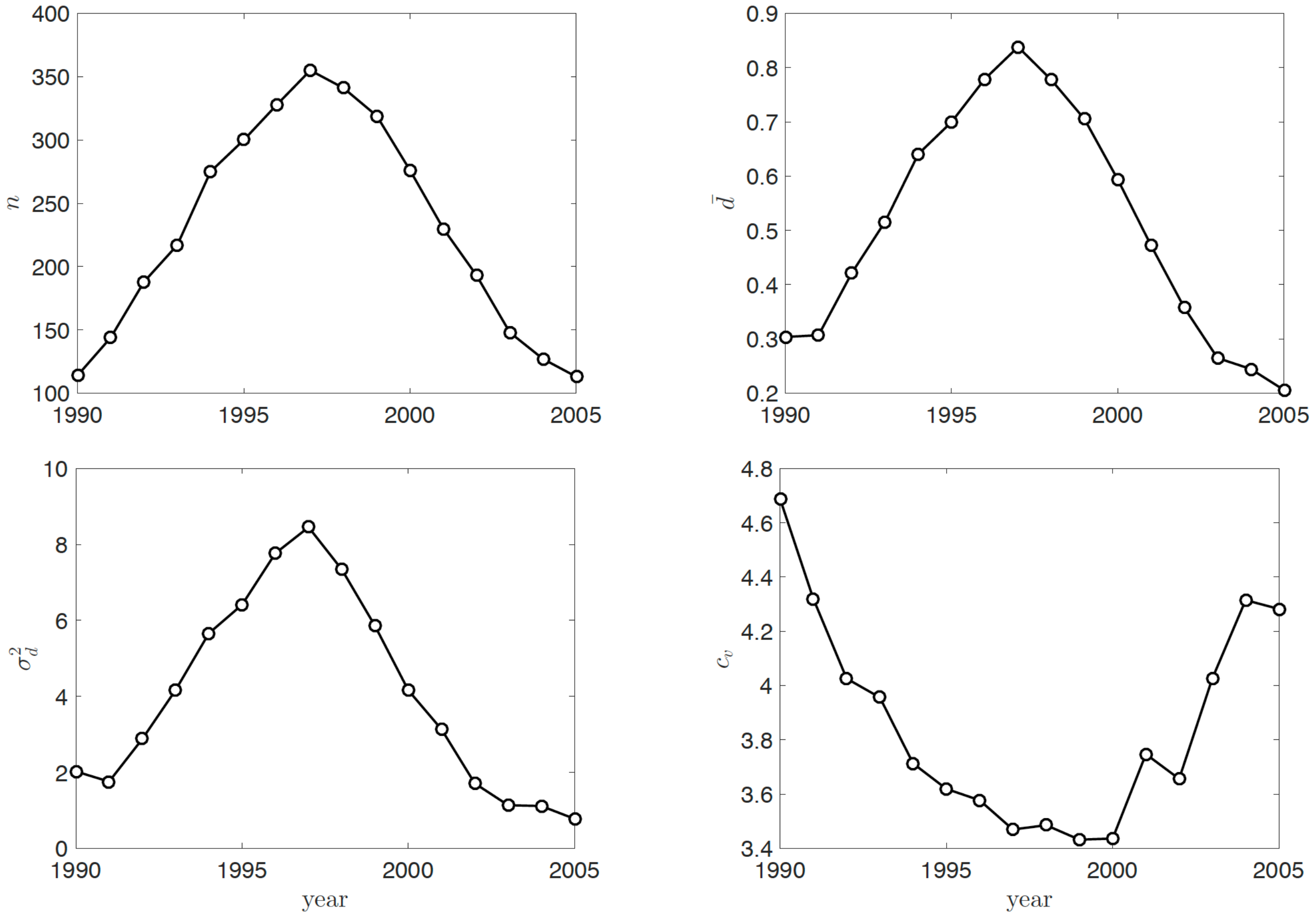


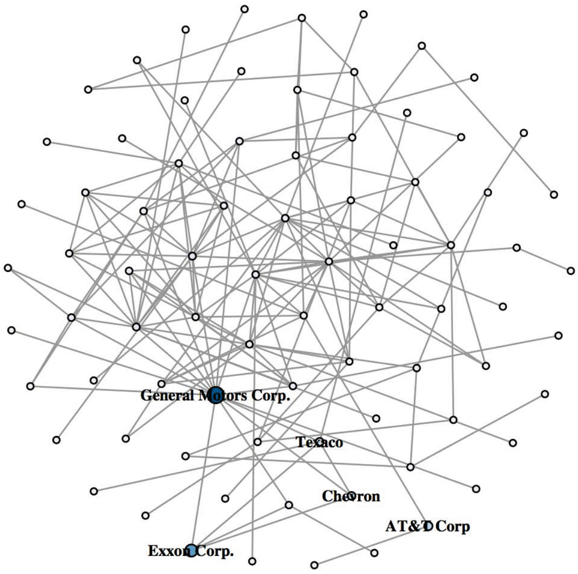
Figure 2: The number of firms, n, participating in an alliance, the average degree, $\bar{d}$, the degree variance, σ_d^2, and the degree coefficient of variation, $c_v = \sigma_d/\bar{d}$.

average degree, $\bar{d}$, the degree variance, σ_d^2, and the degree coefficient of variation, $c_v = \sigma_d/\bar{d}$, over the years 1990 to 2005. It can be seen that there are very large variations over the years in the number of firms having an R&D alliance with other firms. Starting from 1990, we observe a strong increase (due to the IT boom) followed by a steady decline from 1997 onwards. Both, the average number of alliances per firm (captured by the average degree $\bar{d}$) and the degree variance σ_d^2 follow a similar pattern. In contrast, the degree coefficient of variation, c_v, has first decreased and then increased over the years.

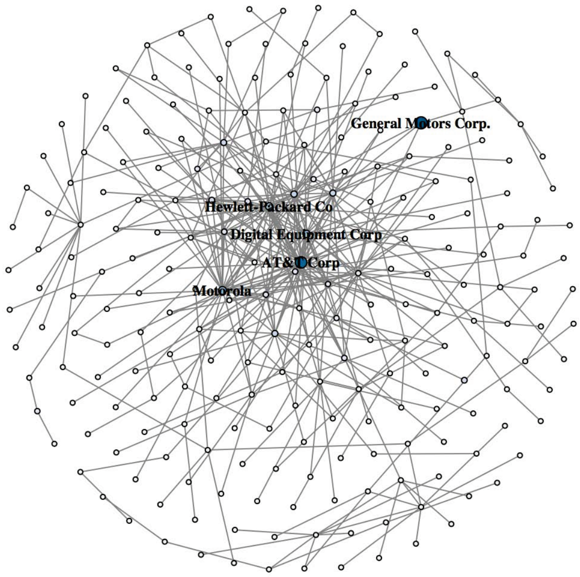
In Figure 3, exemplary plots of the largest connected component in the R&D network for the years 1990, 1995, 2000 and 2005 are shown. The giant component has a core-periphery structure with many R&D interactions between firms from different sectors.[21]

The combined CATI-SDC database provides the names for each firm in an alliance, but does not contain balance sheet information. We thus matched the firms' names in the CATI-SDC database with the firms' names in Standard & Poor's Compustat U.S. annual fundamentals

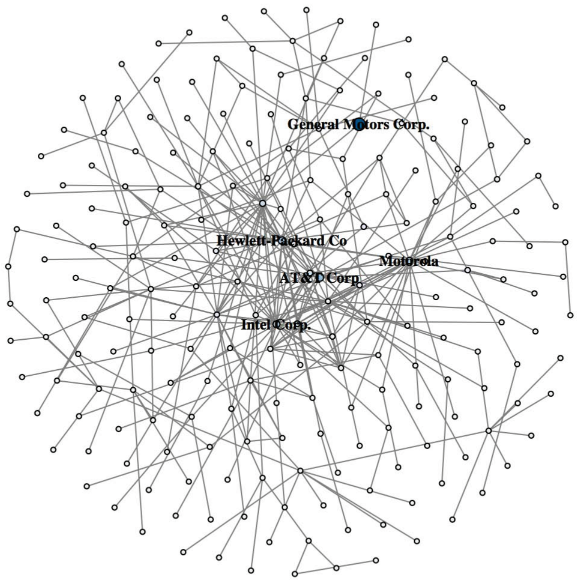
[21]See also Figure H.5 in the Online Appendix H.1.



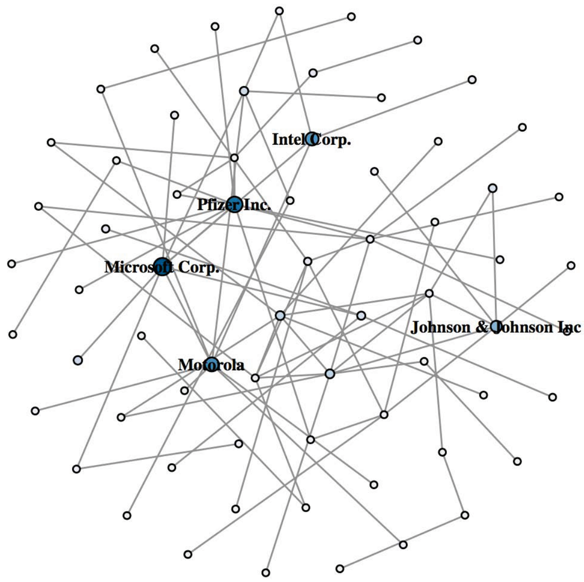
(a) 1990



(b) 1995



(c) 2000



(d) 2005

Figure 3: Network snapshots of the largest connected component for the years (a) 1990, (b) 1995, (c) 2000 and (d) 2005. Nodes' sizes and shades indicate their targeted subsidies (see Section 7). The names of the 5 highest subsidized firms are indicated in the network.

Table 1: Summary statistics computed across the years 1967 to 2006.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.	Compustat Mean
Sales [10^6]	21,067	2,101.56	7,733.29	9.98×10^{-8}	168,055.80	1,085.05
Empl.	19,709	16,694.82	51,299.36	1	876,800.00	4,322.08
Capital [10^6]	20,873	1,629.29	7,388.32	3.82×10^{-8}	170,437.40	663.44
R&D Exp. [10^6]	18,629	70.75	287.42	5.56×10^{-4}	6,621.19	14.71
R&D Exp. / Empl.	17,203	20,207.79	55,887.27	3.37	2,568,507.00	4,060.12
R&D Stock [10^6]	17,584	406.87	1,520.97	5.58×10^{-3}	22,292.97	33.13
Num. Patents	12,177	2,588.31	7,814.59	1	76,644.00	14.39

Notes: Values for sales, capital and R&D expenses are in U.S. dollars with 1983 as the base year. Compustat means are computed across all firms in the Compustat U.S. fundamentals annual database over all non-missing observations over the years 1967 to 2006.

database, as well as Bureau van Dijk's Osiris database, to obtain information about their balance sheets and income statements [see e.g. Dai, 2012]. Compustat and Osiris only contain firms listed on the stock market, so they typically exclude smaller firms. However, they should capture the most R&D intensive firms, as R&D is typically concentrated in publicly listed firms [cf. e.g. Bloom et al., 2013]. The Online Appendix H.3 provides additional details about the accounting databases used in this study.

For the purpose of matching firms across databases, we again use the above mentioned name matching algorithm. We could match roughly 26% of the firms in the alliance data (considering only firms with accounting information available). From our match between the firms' names in the alliance database and the firms' names in the Compustat and Osiris databases, we obtained a firm's sales and R&D expenditures. Individual firms' output levels are computed from deflated sales using 2-SIC digit industry-year specific price deflators from the OECD-STAN database [cf. Gal, 2013].[22] Furthermore, we use information on R&D expenditures to compute R&D capital stocks using a perpetual inventory method with a 15% depreciation rate (following Hall et al. [2000] and Bloom et al. [2013]). Considering only firms with non-missing observations on sales, output and R&D expenditures we end up with a sample of $1,186$ firms and a total of 1010 collaborations over the years 1967 to 2006.[23]

The empirical distributions for output $P(q)$ (using a logarithmic binning of the data with 100 bins) and the degree distribution $P(d)$ are shown in Figure 4. Both are highly skewed,

[22]In Online Appendix J.4, as a robustness check, we consider three alternative specifications of the competition matrix based on the primary and secondary industry classification codes that can be found in the Compustat Segments and Orbis databases [cf. Bloom et al., 2013], or using the Hoberg-Phillips product similarity indicators [cf. Hoberg and Phillips , 2016].

[23]See the Online Appendix H for a discussion about the representativeness of our data sample, and Online Appendix J.5 for a discussion about the impact of missing data on our estimation results.

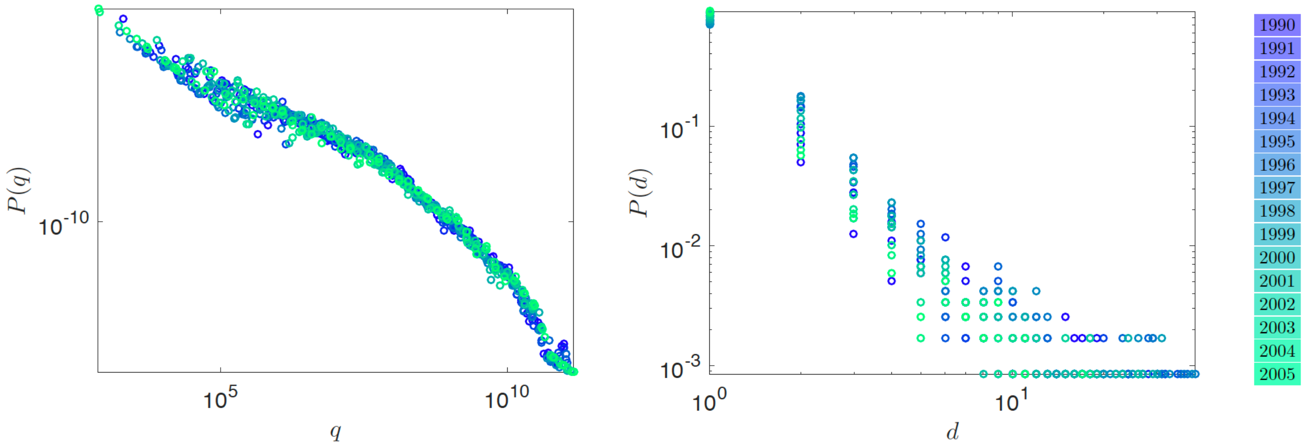


Figure 4: Empirical output distribution $P(q)$ and the distribution of degree $P(d)$ for the years 1990 to 2005. The data for output has been logarithmically binned and non-positive data entries have been discarded. Both distributions are highly skewed.

indicating a large degree of inequality in the number of goods produced as well as the number of R&D collaborations. Industry totals are computed across all firms in the Compustat U.S. fundamentals database (without missing observations). Basic summary statistics can be seen in Table 1. The table shows that the R&D collaborating firms in our sample are typically larger and have higher R&D expenditures than the average across all firms in the Compustat database. This is consistent with previous studies which found that cooperating firms tend to be larger and more R&D intensive [cf. e.g. Belderbos et al., 2004].

6. Econometric Analysis

6.1. Econometric Specification

In this section, we introduce the econometric equivalent to the equilibrium quantity produced by each firm given in Equation (10). Our empirical counterpart of the marginal cost c_{it} of firm i from Equation (2) at period t has a fixed cost equal to $\bar{c}_{it} = \eta_i^* - \epsilon_{it} - x_{it}\beta$, and thus we get

$$c_{it} = \eta_i^* - \epsilon_{it} - \beta x_{it} - e_{it} - \varphi \sum_{j=1}^{n} a_{ij,t} e_{jt}, \tag{23}$$

where x_{it} is a measure for the productivity of firm i, η_i^* captures the unobserved (to the econometrician) time-invariant characteristics of the firm, and ϵ_{it} captures the remaining unobserved (to the econometrician) characteristics of the firm.

Following Equation (1), the inverse demand function for firm i is given by

$$p_{it} = \bar{\alpha}_m + \bar{\alpha}_t - q_{it} - \rho \sum_{j=1}^{n} b_{ij} q_{jt}, \tag{24}$$

where $b_{ij} = 1$ if i and j are in the same market and zero otherwise. In this equation, $\bar{\alpha}_m$ indicates the market-specific fixed effect and $\bar{\alpha}_t$ captures the time fixed effect due to exogenous demand shifters that affect consumer income, number of consumers, consumer taste and preferences, and expectations over future prices of complements and substitutes and future income.

Denote by $\kappa_t \equiv \bar{\alpha}_t$ and $\eta_i \equiv \bar{\alpha}_m - \eta_i^*$. Observe that κ_t captures the time fixed effect while η_i, which includes both $\bar{\alpha}_m$ and η_i^*, captures the firm fixed effect. Then, proceeding as in Section 3 (see, in particular the proof of Proposition 1), adding subscript t for time and using Equations (23) and (24), the econometric equivalent to the best-response quantity in Equation (10) is given by

$$q_{it} = \varphi \sum_{j=1}^{n} a_{ij,t} q_{jt} - \rho \sum_{j=1}^{n} b_{ij} q_{jt} + \beta x_{it} + \eta_i + \kappa_t + \epsilon_{it}. \tag{25}$$

Observe that the econometric specification in Equation (25) has a similar specification as the product competition and technology spillover production function estimation in Bloom et al. [2013] where the estimation of φ will give the intensity of the *technology (or knowledge) spillover effect* of R&D, while the estimation of ρ will give the intensity of the *product rivalry effect*. However, as opposed to that paper, we explicitly model the technology spillovers stemming from R&D collaborations using a network approach.

In vector-matrix form, we can write Equation (25) as

$$\mathbf{q}_t = \varphi \mathbf{A}_t \mathbf{q}_t - \rho \mathbf{B} \mathbf{q}_t + \mathbf{x}_t \beta + \boldsymbol{\eta} + \kappa_t \boldsymbol{\iota}_n + \boldsymbol{\epsilon}_t, \tag{26}$$

where $\mathbf{q}_t = (q_{1t}, \cdots, q_{nt})^\top$, $\mathbf{A}_t = [a_{ij,t}]$, $\mathbf{B} = [b_{ij}]$, $\mathbf{x}_t = (x_{1t}, \cdots, x_{nt})^\top$, $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)^\top$, $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \cdots, \epsilon_{nt})^\top$, and $\boldsymbol{\iota}_n$ is an n-dimensional vector of ones.

For the T periods, Equation (26) can be written as

$$\mathbf{q} = \varphi \mathrm{diag}\{\mathbf{A}_t\} \mathbf{q} - \rho (\mathbf{I}_T \otimes \mathbf{B}) \mathbf{q} + \mathbf{x}\beta + \boldsymbol{\iota}_T \otimes \boldsymbol{\eta} + \boldsymbol{\kappa} \otimes \boldsymbol{\iota}_n + \boldsymbol{\epsilon}, \tag{27}$$

where $\mathbf{q} = (\mathbf{q}_1^\top, \cdots, \mathbf{q}_T^\top)^\top$, $\mathbf{x} = (\mathbf{x}_1^\top, \cdots, \mathbf{x}_T^\top)^\top$, $\boldsymbol{\kappa} = (\kappa_1, \cdots, \kappa_T)^\top$, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \cdots, \boldsymbol{\epsilon}_T^\top)^\top$. The

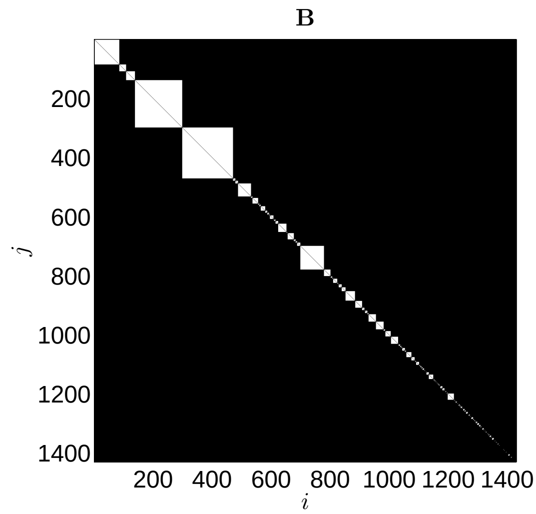


Figure 5: The empirical competition matrix $\mathbf{B} = (b_{ij})_{1 \leq i,j \leq n}$ measured by 4-digit level industry SIC codes.

vectors $\mathbf{q}$, $\mathbf{x}$ and $\boldsymbol{\epsilon}$ are of dimension $(nT \times 1)$, where T is the number of years available in the data.

In terms of data, our main variables will be measured as follows. Output q_{it} is calculated using sales divided by the year-industry price deflators from the OECD-STAN database [cf. Gal, 2013]. The network data stems from the combined CATI-SDC databases and we set $a_{ij,t} = 1$ if there exists an R&D collaboration between firms i and j in the last s years before time t, where s is the duration of an alliance. The exogenous variable x_{it} is the firm's time-lagged R&D stock at the time $t-1$. Finally, we measure b_{ij} as in the theoretical model so that $b_{ij} = 1$ if firms i and j are the same industry (measured by the industry SIC codes at the 4-digit level) and $b_{ij} = 0$ otherwise. The empirical competition matrix $\mathbf{B}$ can be seen in Figure 5. The block-diagonal structure indicating different markets is clearly visible.

6.2. Identification Issues

We adopt a structural approach in the sense that we estimate the first-order condition of the firms' profit maximization problem in terms of output and R&D effort, which lead to Equations (25) and (26). The best-response quantity in Equation (26) then corresponds to a higher-order Spatial Auto-Regressive (SAR) model with two spatial lags, $\mathbf{A}_t\mathbf{q}_t$ and $\mathbf{B}\mathbf{q}_t$ [cf. Lee and Liu, 2010].

There are several potential identification problems in the estimation of Equation (25) or

(26). We face, actually, four sources of potential bias[24] arising from (*i*) *correlated or common-shock effects*, (*ii*) *simultaneity* of q_{it} and q_{jt}, (*iii*) *endogeneity of the R&D stock*, and (*iv*) *endogeneity of the R&D alliance matrix.*

6.2.1. Correlated or Common-Shock Effects

Correlated or common-shock effects arise in network models due to the fact that there may be common environmental factors that cause the individuals in the same network to behave in a similar manner. They may be confounded with the network effects (i.e. φ and ρ) we are trying to identify. To alleviate this problem, we incorporate both firm and time fixed effects (i.e. η_i and κ_t) to the outcome Equation (25).

6.2.2. Simultaneity of Product Outputs

We use instrumental variables when estimating our outcome Equation (25) to deal with the issue of simultaneity between q_{it} and q_{jt}. Indeed, the output of firm i at time t, q_{it}, is a function of the total output of all firms collaborating in R&D with firm i at time t, i.e. $\bar{q}_{a,it} \equiv \sum_{j=1}^{n} a_{ij,t} q_{jt}$, and the total output of all firms that operate in the same market as firm i, i.e. $\bar{q}_{b,it} \equiv \sum_{j=1}^{n} b_{ij} q_{jt}$. Due the feedback effect, q_{jt} also depends on q_{it} and, thus, $\bar{q}_{a,it}$ and $\bar{q}_{b,it}$ are endogenous.

Recall that x_{it} denotes the time-lagged R&D stock of firm i at the time $t-1$. To deal with this issue, we instrument $\bar{q}_{a,it}$ by the time-lagged total R&D stock of all firms with an R&D collaboration with firm i, i.e. $\sum_{j=1}^{n} a_{ij,t} x_{jt}$, and instrument $\bar{q}_{b,it}$ by the time-lagged total R&D stock of all firms that operate in the same industry as firm i, i.e. $\sum_{j=1}^{n} b_{ij} x_{jt}$. The rationale for this IV strategy is that the time-lagged total R&D stock of R&D collaborators and product competitors of firm i *directly* affects the total output of these firms but only *indirectly* affects the output of firm i through the total output of these same firms.

More formally, to estimate Equation (27), first we transform it with the projection matrix $\mathbf{J} = (\mathbf{I}_T - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^\top) \otimes (\mathbf{I} - \frac{1}{n}\boldsymbol{\iota}_n\boldsymbol{\iota}_n^\top)$. The transformed Equation (27) is

$$\mathbf{J}\mathbf{q} = \varphi \mathbf{J}\mathrm{diag}\{\mathbf{A}_t\}\mathbf{q} - \rho \mathbf{J}(\mathbf{I}_T \otimes \mathbf{B})\mathbf{q} + \mathbf{J}\mathbf{x}\beta + \mathbf{J}\boldsymbol{\epsilon}, \tag{28}$$

[24] It should be clear that there is no exogenous contextual effect (and thus no reflection problem) in Equation (25).

where the firm and time fixed effects $\boldsymbol{\eta}$ and $\boldsymbol{\kappa}$ have been eliminated by the projection matrix.[25] Let $\mathbf{Q}_1 = \mathbf{J}[\text{diag}\{\mathbf{A}_t\}\mathbf{x}, (\mathbf{I}_T \otimes \mathbf{B})\mathbf{x}, \mathbf{x}]$ denote the IV matrix and $\mathbf{Z} = \mathbf{J}[\text{diag}\{\mathbf{A}_t\}\mathbf{q}, (\mathbf{I}_T \otimes \mathbf{B})\mathbf{q}, \mathbf{x}]$ denote the matrix of regressors in Equation (28). As there is a single exogenous variable in Equation (28), the model is just-identified. The IV estimator of parameters $(\varphi, -\rho, \beta)^\top$ is given by $(\mathbf{Q}_1^\top \mathbf{Z})^{-1}\mathbf{Q}_1^\top \mathbf{q}$. With the estimated $(\varphi, -\rho, \beta)^\top$, one can recover $\boldsymbol{\eta}$ and $\boldsymbol{\kappa}$ by the least squares dummy variables method.

Obviously, the above IV-based identification strategy is valid only if the time-lagged R&D stock, $x_{i,t-1}$, and the R&D alliance matrix, $\mathbf{A}_t = [a_{ij,t}]$, are exogenous. In Section 6.2.3 we address the potential endogeneity of the time-lagged R&D stock, while the endogeneity of the R&D alliance matrix is discussed in Section 6.2.4.

6.2.3. Endogeneity of the R&D Stock

The R&D stock depends on past R&D efforts, which could be correlated with the error term of Equation (25). However, as the R&D stock is time-lagged and fixed effects are included, the existing literature has argued that the correlation between the (time-lagged) R&D stock and the error term of Equation (25) is likely to be weak. To further alleviate the potential endogeneity issue of the time-lagged R&D stock, we use supply side shocks from tax-induced changes to the user cost of R&D to construct IVs as in Bloom et al. [2013].[26] To be more specific, we use changes in the firm-specific tax price of R&D to construct instrumental variables for R&D expenditures. Let w_{it} denote the time-lagged R&D tax credit firm i received at time $t-1$.[27] We instrument $\bar{q}_{a,it}$ by the time-lagged total R&D tax credits of all firms having R&D collaborations with firm i, i.e. $\sum_{j=1}^n a_{ij,t} w_{jt}$, instrument $\bar{q}_{b,it}$ by the time-lagged total R&D tax credits of all firms that operate in the same industry as firm i, i.e. $\sum_{j=1}^n b_{ij} w_{jt}$, and instrument the time-lagged R&D stock x_{it} by the time-lagged R&D tax credit w_{it}. The rationale for this IV strategy is that the time-lagged total R&D credits of R&D collaborators and product competitors of firm i *directly* affects the total output of these firms but only *indirectly* affects the output of firm i through the total output of these same firms.

[25] For unbalanced panels, the firm and time fixed effects can be eliminated by a projection matrix given in Wansbeek and Kapteyn [1989].

[26] We would like to thank Nick Bloom for making the tax credit data available to us.

[27] See Appendix B.3 in the Supplementary Material of Bloom et al. [2013] for details on the specification of w_{it}.

More formally, let $\mathbf{Q}_2 = \mathbf{J}[\text{diag}\{\mathbf{A}_t\}\mathbf{w}, (\mathbf{I}_T \otimes \mathbf{B})\mathbf{w}, \mathbf{w}]$, where $\mathbf{w} = (\mathbf{w}_1^\top, \cdots, \mathbf{w}_T^\top)^\top$ and $\mathbf{w}_t = (w_{1t}, \cdots, w_{nt})^\top$, denote the IV matrix, and $\mathbf{Z} = \mathbf{J}[\text{diag}\{\mathbf{A}_t\}\mathbf{q}, (\mathbf{I}_T \otimes \mathbf{B})\mathbf{q}, \mathbf{x}]$ denote the matrix of regressors in Equation (28). The IV estimator of parameters $(\varphi, -\rho, \beta)^\top$ is given by $(\mathbf{Q}_2^\top \mathbf{Z})^{-1}\mathbf{Q}_2^\top \mathbf{q}$.

6.2.4. Endogeneity of the R&D Alliance Matrix

The R&D alliance matrix $\mathbf{A}_t = [a_{ij,t}]$ is endogenous if there exists an *unobservable factor* that affects both the outputs, q_{it} and q_{jt}, and the R&D alliance, indicated by $a_{ij,t}$. If the unobservable factor is firm-specific, then it is captured by the firm fixed-effect η_i. If the unobservable factor is time-specific, then it is captured by the time fixed-effect κ_t. Therefore, the fixed effects in the panel data model are helpful for attenuating the potential endogeneity of $\mathbf{A}_t$.

However, it may still be that there are some unobservable firm-specific time-varying factors that affect the formation of R&D collaborations and thus make the R&D alliance matrix $\mathbf{A}_t$ endogenous. To deal with this issue, we run a two-stage IV estimation as in Kelejian and Piras [2014] where, in the first stage, we obtain a predicted R&D alliance matrix based on predetermined dyadic characteristics, and, in the second stage, we employ the IV strategy explained above using IVs constructed with the predicted adjacency matrix from the first stage.

Let us now explain how to obtain a predicted R&D alliance matrix in the first stage. We estimate a logistic regression model with the corresponding log-odds ratio as a function of predetermined dyadic characteristics:

$$\log\left(\frac{\mathbb{P}\left(a_{ij,t}=1 \mid (\mathbf{A}_\tau)_{\tau=1}^{t-s-1}, f_{ij,t-s-1}, city_{ij}, market_{ij}\right)}{1-\mathbb{P}\left(a_{ij,t}=1 \mid (\mathbf{A}_\tau)_{\tau=1}^{t-s-1}, f_{ij,t-s-1}, city_{ij}, market_{ij}\right)}\right)$$
$$= \gamma_0 + \gamma_1 \max_{\tau=1,\ldots,t-s-1} a_{ij,\tau} + \gamma_2 \max_{\substack{\tau=1,\ldots,t-s-1\\ k=1,\ldots,n}} a_{ik,\tau}a_{kj,\tau} + \gamma_3 f_{ij,t-s-1} + \gamma_4 f_{ij,t-s-1}^2 + \gamma_5 city_{ij} + \gamma_6 market_{ij}, \tag{29}$$

In this model, $\max_{\tau=1,\ldots,t-s-1} a_{ij,\tau}$ is a dummy variable, which is equal to 1 if firms i and j had an R&D collaboration before time $t-s$ (s is the duration of an alliance) and 0 otherwise; $\max_{\tau=1,\ldots,t-s-1;k=1,\ldots,n} a_{ik,\tau}a_{kj,\tau}$ is a dummy variable, which is equal to 1 if firms i and j had a common R&D collaborator before time $t-s$ and 0 otherwise; $f_{ij,t-s-1}$ is the time-lagged technological proximities between firms i and j, measured here by either the Jaffe or the Mahalanobis patent similarity indices at time $t-s-1$;[28] $city_{ij}$ is a dummy variable, which is equal to 1 if firms i and j are located in the same city[29] and 0 otherwise; and $market_{ij}$ is a dummy variable, which is equal to 1 if firms i and j are in the same market and 0 otherwise.[30]

The rationale for this IV solution is as follows. Take, for example, the dummy variable, which is equal to 1 if firms i and j had a common R&D collaborator before time $t-s$, and 0 otherwise. This means that, if firms i and j had a common collaborator in the past (i.e. before time $t-s$), then they are more likely to have an R&D collaboration in period t, i.e. $a_{ij,t}=1$, but, conditional on the firm and time fixed effects, having a common collaborator in the past should not *directly* affect the outputs of firms i and j in period t (i.e. the exclusion restriction is satisfied). A similar argument can be made for the other variables in Equation (29). As a result, using IVs based on the predicted adjacency matrix $\widehat{\mathbf{A}}_t$ should alleviate the concern of invalid IVs due to the endogeneity of the adjacency matrix $\mathbf{A}_t$.

Formally, let $\mathbf{Q}_3 = \mathbf{J}[\text{diag}\{\widehat{\mathbf{A}}_t\}\mathbf{x}, (\mathbf{I}_T \otimes \mathbf{B})\mathbf{x}, \mathbf{x}]$ denote the IV matrix based on the predicted R&D alliance matrix and $\mathbf{Z} = [\text{diag}\{\mathbf{A}_t\}\mathbf{q}, (\mathbf{I}_T \otimes \mathbf{B})\mathbf{q}, \mathbf{x}]$ denote the matrix of regressors in Equation (28). Then, the estimator of the parameters $(\varphi, -\rho, \beta)^\top$ with IVs based on the predicted adjacency matrix is given by $(\mathbf{Q}_2^\top \mathbf{Z})^{-1}\mathbf{Q}_3^\top \mathbf{q}$.

6.3. Estimation Results

6.3.1. Main results

Table 2 reports the parameter estimates of Equation (26) with time fixed effects (Model A) and with both firm and time fixed effects (Model B). In these regressions, we assume that the time-lagged R&D stock and the R&D alliance matrix are exogenous. We see that, with both firm and time fixed effects, the estimated parameters in Model B are statistically significant with the

[28] We matched the firms in our alliance data with the owners of patents recorded in the Worldwide Patent Statistical Database (PATSTAT). This allowed us to obtain the number of patents and the patent portfolio held for about 36% of the firms in the alliance data. From the firms' patents, we then computed their technological proximity following Jaffe [1986] as $f_{ij}^{\mathrm{J}} = \frac{\mathbf{P}_i^\top \mathbf{P}_j}{\sqrt{\mathbf{P}_i^\top \mathbf{P}_i}\sqrt{\mathbf{P}_j^\top \mathbf{P}_j}}$, where $\mathbf{P}_i$ represents the patent portfolio of firm i and is a vector whose k-th component P_{ik} counts the number of patents firm i has in technology category k divided by the total number of technologies attributed to the firm. As an alternative measure for technological similarity we also use the Mahalanobis proximity index f_{ij}^{M} introduced in Bloom et al. [2013]. The Online Appendix H.5 provides further details about the match of firms to their patent portfolios and the construction of the technology proximity measures f_{ij}^k, $k \in \{\mathrm{J}, \mathrm{M}\}$.

[29] See Singh [2005] who also tests the effect of geographic distance on R&D spillovers and collaborations.

[30] Observe that the predictors for the link-formation probability are either time-lagged or predetermined so the IVs constructed with $\widehat{\mathbf{A}}_t$ are less likely to suffer from any endogeneity issues.

Table 2: Parameter estimates from a panel regression of Equation (26). Model A includes only time fixed effects, while Model B includes both firm and time fixed effects. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

	Model A		Model B	
φ	-0.0118	(0.0075)	0.0106**	(0.0051)
ρ	0.0114***	(0.0015)	0.0189***	(0.0028)
β	0.0053***	(0.0002)	0.0027***	(0.0002)
# firms	1186		1186	
# observations	16924		16924	
Cragg-Donald Wald F stat.	6454.185		7078.856	
firm fixed effects	no		yes	
time fixed effects	yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

expected signs, i.e., the *technology (or knowledge) spillover effect* (estimate of φ) has a *positive* impact on own output while the *product rivalry effect* (estimate of ρ) has *negative* impact on own output. However, without controlling for firm fixed effects, the estimated technology spillover effect in Model A is negative.

As Equation (6) of the theoretical model suggests, a firm's R&D effort is proportional to its production level, the positive technology spillover effect indicates that the higher a firm's production level (or R&D effort) is, the more its R&D collaborator produces. That is, there exist strategic complementarities between allied firms in production and R&D effort. On the other hand, the negative product rivalry effect indicates the higher a firm's production level (or R&D effort) is, the less its product competitors in the same market produce. Furthermore, this table also shows that a firm's productivity captured by its own time-lagged R&D stock has a positive and significant impact on its own production level. Finally, the Cragg-Donald Wald F statistics for both models are well above the conventional benchmark for weak IVs [cf. Stock and Yogo, 2005].

6.3.2. Endogeneity of R&D Stocks and Tax-Credit Instruments

Table 3 reports the parameter estimates of Equation (26) with tax credits as IVs for the time-lagged R&D stock as discussed in Section 6.2.3. Similarly to the benchmark results reported in Section 6.3.1, with both firm and time fixed effects, the estimated parameters in Model D are statistically significant with the expected signs, i.e., the *technology (or knowledge) spillover*

Table 3: Parameter estimates from a panel regression of Equation (26) with IVs based on time-lagged tax credits. Model C includes only time fixed effects, while Model D includes both firm and time fixed effects. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

	Model C		Model D	
φ	-0.0133	(0.0114)	0.0128*	(0.0069)
ρ	0.0182***	(0.0018)	0.0156**	(0.0076)
β	0.0054***	(0.0004)	0.0023***	(0.0006)
# firms	1186		1186	
# observations	16924		16924	
Cragg-Donald Wald F stat.	138.311		78.791	
firm fixed effects	no		yes	
time fixed effects	yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

effect is positive while the *product rivalry effect* is negative. However, without firm fixed effects, the estimated technology spillover effect in Model C is biased downward to become negative, which is similar to what we obtained without the tax-credit instruments (Table 2). Furthermore, a firm's productivity captured by its own time-lagged R&D stock has a positive and significant impact on its own production level. Finally, the reported Cragg-Donald Wald F statistics for both models suggest the IVs based on tax credits are informative.

6.3.3. Endogeneity of the R&D Alliance Matrix

We also consider IVs based on the predicted R&D alliance matrix, i.e. $\widehat{\mathbf{A}}_t\mathbf{x}_t$, as discussed in Section 6.2.3.

First, we obtain the predicted alliance-formation probability $\hat{a}_{ij,t}$ from the logistic regression given by Equation (29). The logistic regression result, using either the Jaffe or Mahalanobis patent similarity measures, is reported in Table 4. The estimated coefficients are all statistically significant with expected signs. Interestingly, having a past collaboration or a past common collaborator, being established in the same city, or operating in the same industry/market increases the probability that two firms have an R&D collaboration in the current period. Furthermore, being close in technology (measured by either the Jaffe or Mahalanobis patent similarity measure) in the past also increases the chance of having an R&D collaboration in the current period, even though this relationship is concave.

Next, we estimate Equation (26) with IVs based on the predicted alliance matrix. The

Table 4: Link formation regression results. Technological similarity, f_{ij}, is measured using either the Jaffe or the Mahalanobis patent similarity measures. The dependent variable $a_{ij,t}$ indicates if an R&D alliance exists between firms i and j at time t. The estimation is based on the observed alliances in the years 1967–2006.

technological similarity	Jaffe	Mahalanobis
Past collaboration	0.5981***	0.5920***
	(0.0150)	(0.0149)
Past common collaborator	0.1162***	0.1164***
	(0.0238)	(0.0236)
$f_{ij,t-s-1}$	13.6977***	6.0864***
	(0.6884)	(0.3323)
$f^2_{ij,t-s-1}$	-20.4083***	-3.9194***
	(1.7408)	(0.4632)
$city_{ij}$	1.1283***	1.1401***
	(0.1017)	(0.1017)
$market_{ij}$	0.8451***	0.8561***
	(0.0424)	(0.0422)
# observations	3,964,120	3,964,120
McFadden's R^2	0.0812	0.0813

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

estimates are reported in Table 5. We find that the estimates of both the technology spillovers and the product rivalry effect are still significant with the expected signs. Compared to Table 2, the estimate of the technology spillovers (i.e. the estimation of φ) has, however, a larger value and a larger standard error. Finally, the reported Cragg-Donald Wald F statistics suggest the IVs based on the predicted alliance matrix are informative.

6.3.4. Robustness Analysis

In Online Appendix J, we perform some additional robustness checks. First, in Appendix J.1, we estimate our model for alliance durations ranging from 3 to 7 years. Second, in Appendix J.2, we consider a model where the spillover and competition coefficients are *not* identical across markets. We perform a robustness check using two major divisions in our data, namely the manufacturing and services sectors that cover, respectively, 76.8% and 19.3% firms in our sample. Third, in Appendix J.3, we conduct a robustness analysis by directly controlling for potential input-supplier effects. Fourth, in Appendix J.4, we consider three alternative specifications of the competition matrix. Finally, in Appendix J.5, we discuss the issue of possible biases due to sampled network data. We find that the estimates are robust to all these extensions.

Table 5: Parameter estimates from a panel regression of Equation (26) with endogenous R&D alliance matrix. The IVs are based on the predicted links from the logistic regression reported in Table 4, where technological similarity is measured using either the Jaffe or the Mahalanobis patent similarity measures. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

technological similarity	Jaffe		Mahalanobis	
φ	0.0582*	(0.0343)	0.0593*	(0.0341)
ρ	0.0197***	(0.0031)	0.0197***	(0.0031)
β	0.0024***	(0.0002)	0.0024***	(0.0002)
# firms	1186		1186	
# observations	16924		16924	
Cragg-Donald Wald F stat.	48.029		49.960	
firm fixed effects	yes		yes	
time fixed effects	yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

6.4. Direct and Indirect Technology Spillovers

In this section, we extend our empirical model of Equation (25) by allowing for both, direct (between firms with an R&D alliance) and indirect (between firms without an R&D alliance) technology spillovers. The generalized model is given by[31]

$$q_{it} = \varphi \sum_{j=1}^{n} a_{ij,t} q_{jt} + \chi \sum_{j=1}^{n} f_{ij,t} q_{jt} - \rho \sum_{j=1}^{n} b_{ij} q_{jt} + \beta x_{it} + \eta_i + \kappa_t + \epsilon_{it}, \tag{30}$$

where $f_{ij,t}$ are weights characterizing alternative channels for technology spillovers (measured by the technological proximity between firms using either the Jaffe or the Mahalanobis patent similarity measures; see Bloom et al. [2013]) other than R&D collaborations, and the coefficients φ and χ capture the direct and the indirect technology spillover effects, respectively. In vector-matrix form, we then have

$$\mathbf{q}_t = \varphi \mathbf{A}_t \mathbf{q}_t + \chi \mathbf{F}_t \mathbf{q}_t - \rho \mathbf{B} \mathbf{q}_t + \mathbf{x}_t \beta + \boldsymbol{\eta} + \kappa_t \boldsymbol{\iota}_n + \boldsymbol{\epsilon}_t. \tag{31}$$

The results of a fixed-effect panel regression of Equation (31) are shown in Table 6. Both technology spillover coefficients, φ and χ, are positive, while only the direct spillover effect is significant. This suggests R&D alliances are the main channel for technology spillovers.

[31] The theoretical foundation of Equation (30) can be found in Online Appendix F.

Table 6: Parameter estimates from a panel regression of Equation (31) with both firm and time fixed effects. Technological similarity, f_{ij}, is measured using either the Jaffe or the Mahalanobis patent similarity measures. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

technological similarity	Jaffe		Mahalanobis	
φ	0.0102**	(0.0049)	0.0102**	(0.0049)
χ	0.0063	(0.0052)	0.0043	(0.0030)
ρ	0.0189***	(0.0028)	0.0192**	(0.0028)
β	0.0027***	(0.0002)	0.0027***	(0.0002)
# firms	1190		1190	
# observations	17105		17105	
Cragg-Donald Wald F stat.	4791.308		4303.563	
firm fixed effects	yes		yes	
time fixed effects	yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

7. Empirical Implications for the R&D Subsidy Policy

With our estimates from the previous sections – using Model B in Table 2 as our baseline specification – we are now able to empirically determine the optimal subsidy policy, both for the homogenous case, where all firms receive the same subsidy per unit of R&D (see Proposition 2), and for the targeted case, where the subsidy per unit of R&D may vary across firms (see Proposition 3).[32]

As our empirical analysis focuses on U.S. firms, the central planner that would implement such an R&D subsidy policy could be the U.S. government or a U.S. governmental agency. In the U.S., R&D policies have been widely used to foster the firms' R&D activities. In particular, as of 2006, 32 states in the U.S. provided a tax credit on general, company funded R&D [cf. Wilson, 2009]. Moreover, another prominent example in the U.S. is the Advanced Technology Program (ATP), which was administered by the National Institute of Standards and Technology (NIST) [cf. Feldman and Kelley, 2003].

Observe that we provide a *network-contingent* subsidy program, that is, each time an R&D subsidy policy is implemented, it takes into account the prevalent network structure. In other words, we determine how, for any observed network structure, the R&D policy should be

[32] Additional details about the numerical implementation of the optimal subsidies program can be found in Online Appendix I.

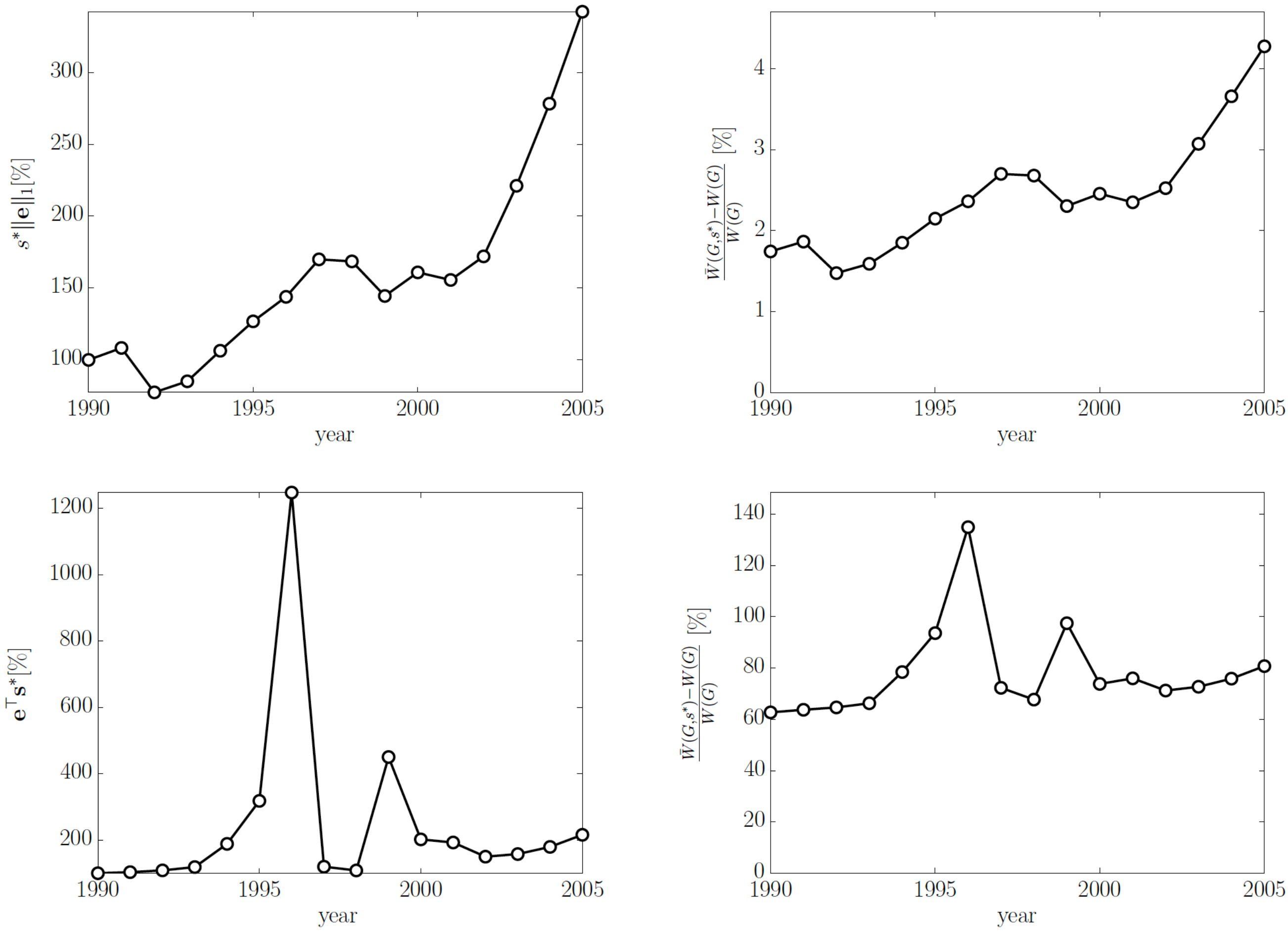


Figure 6: (Top left panel) The total optimal subsidy payments, $s^*\|\mathbf{e}\|_1$, in the homogeneous case over time, using the subsidies in the year 1990 as the base level. (Top right panel) The percentage increase in welfare due to the homogeneous subsidy, s^*, over time. (Bottom left panel) The total subsidy payments, $\mathbf{e}^\top \mathbf{s}^*$, when the subsidies are targeted towards specific firms, using the subsidies in the year 1990 as the base level. (Bottom right panel) The percentage increase in welfare due to the targeted subsidies, $\mathbf{s}^*$, over time.

specified (short-run persepctive). The rationale for this approach is that, in an uncertain and highly dynamic environment such as the R&D intensive industries that we consider, an *optimal contingent policy* is typically preferable over a *fixed policy* [see, e.g. Buiter, 1981].[33] In the following we will then calculate the optimal subsidy for each firm in every year that the network is observed.

In Figure 6, in the top panel, we calculate the optimal homogenous subsidy times R&D effort over time, using the subsidies in the year 1990 as the base level (top left panel), and the percentage increase in welfare due to the homogenous subsidy over time (top right panel). The total subsidized R&D effort more than doubled over the time between 1990 and 2005. In terms of welfare, the highest increase (around 3.5 %) is obtained in the year 2001, while the increase in welfare in 1990 is smaller (below 2.5 %). The bottom panel of Figure 6 does the same exercise

[33]Note that, as the subsidy reacts to changes in the link structure, there is no point in the firms adjusting their links to extract extra subsidies. In particular, if a firm were to form redundant links (with diminishing value added to welfare) then our policy would reduce the subsidies allocated to this firm.

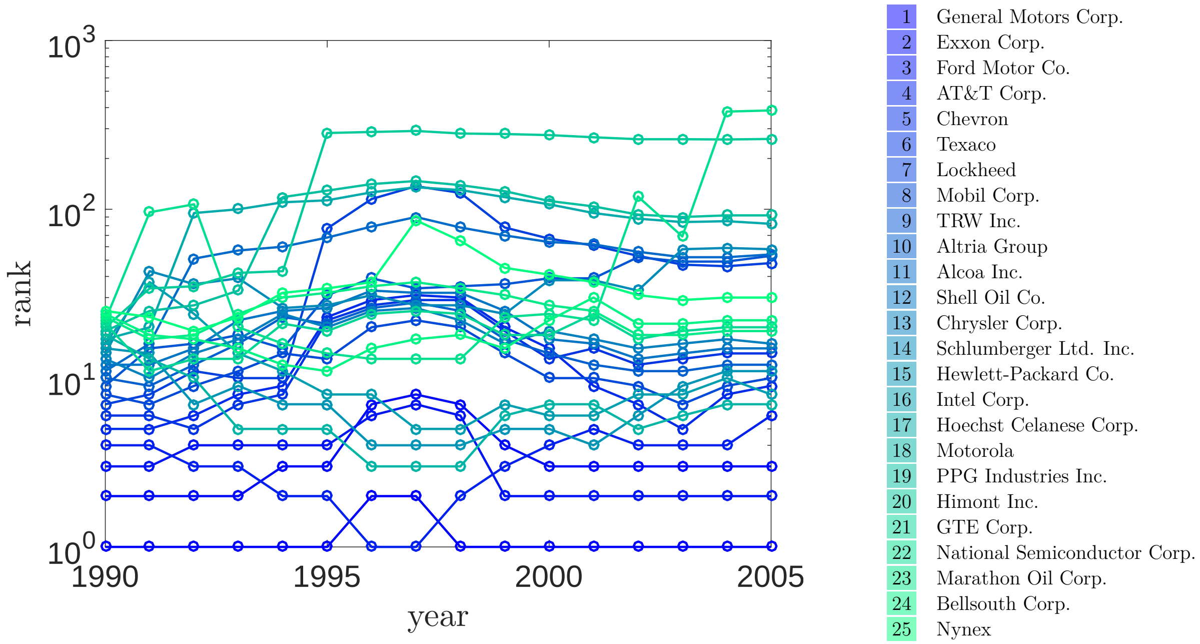


Figure 7: Change in the ranking of the 25 highest subsidized firms (Table 7) from 1990 to 2005.

for the targeted subsidy policy. The largest total expenditures on the targeted subsidies are higher than the ones for the homogeneous subsidies, and they can also vary by several orders of magnitude. The targeted subsidy program also turns out to have a much higher impact on total welfare, as it can improve welfare by up to 80 %, while the homogeneous subsidies can improve total welfare only by up to 3.5 %. Moreover, the optimal subsidy levels show a strong variation over time. Both the homogeneous and the aggregate targeted subsidy seem to follow a cyclical trend (while this pattern seems to be more pronounced for the targeted subsidy), similar to the strong variation we have observed for the number of firms participating in R&D collaborations in a given year in Figure 2. This cyclical trend is also reminiscent of the R&D expenditures observed in the empirical literature on business cycles [cf. Galí, 1999].

We can compare the optimal subsidy level predicted from our model with the R&D tax subsidies actually implemented in the United States and selected other countries between 1979 to 1997 [see Bloom et al., 2002; Impullitti, 2010]. While these time series typically show a steady increase of R&D subsidies over time, they do not seem to incorporate the cyclicality that we obtain for the optimal subsidy levels. Our analysis thus suggests that policy makers should adjust R&D subsidies to these cycles.

We next proceed by providing a ranking of firms in terms of targeted subsidies. Such a ranking can guide a planner who wants to maximize total welfare by introducing an R&D subsidy program, identify which firms should receive the highest subsidies, and how high these subsidies should be. The ranking of the first 25 firms by their optimal subsidy levels in 1990

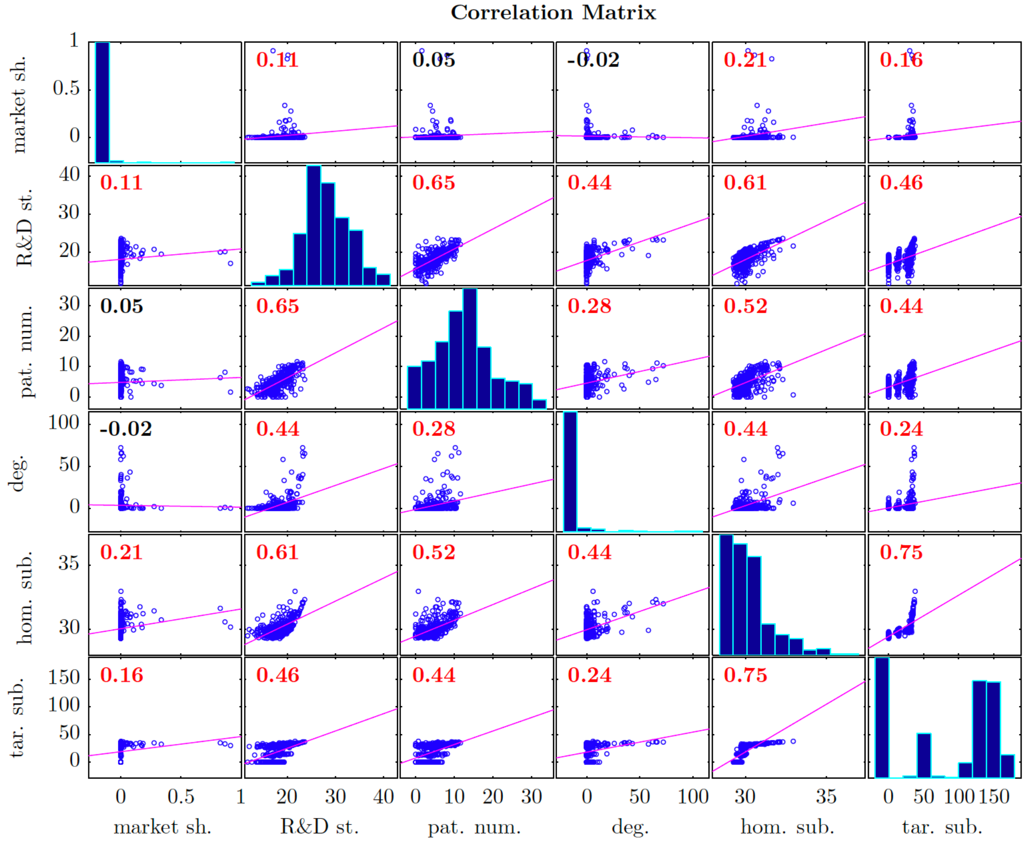


Figure 8: Pair correlation plot of market shares, R&D stocks, the number of patents, the degree, the homogeneous subsidies and the targeted subsidies (cf. Table 8), in the year 2005. The Spearman correlation coefficients are shown for each scatter plot. The data have been log and square root transformed to account for the heterogeneity in across observations.

can be found in Table 7 while the one for 2005 is shown in Table 8.[34] We see that the ranking of firms in terms of subsidies does not correspond to other rankings in terms of network centrality, patent stocks or market share.

There is also volatility in the ranking since many firms that are ranked in the top 25 in 1990 are no longer there in 2005 (for example *TRW Inc.*, *Alcoa Inc.*, *Schlumberger Ltd. Inc.*, etc.). Figure 7 shows the change in the ranking of the 25 highest subsidized firms (Table 7) from 1990 to 2005.

A comparison of market shares, R&D stocks, the number of patents, the degree (i.e. the number of R&D collaborations), the homogeneous subsidy and the targeted subsidy shows a high correlation between the R&D stock and the number of patents, with a (Spearman) correlation coefficient of 0.65 for the year 2005. A high correlation can also be found for the homogeneous subsidy and the targeted subsidy, with a correlation coefficient of 0.75 for the year

[34]The network statistics shown in these tables correspond to the full CATI-SDC network dataset, prior to dropping firms with missing accounting information. See Online Appendix H.1 for more details about the data sources and construction of the R&D alliances network.

2005. The corresponding pair correlation plots for the year 2005 can be seen in Figure 8. We also find that highly subsidized firms tend to have a larger R&D stock, and also a larger number of patents, degree and market share. However, these measures can only partially explain the subsidies ranking of the firms, as the market share is more related to the product market rivalry effect, while the R&D and patent stocks are more related to the technology spillover effect, and both enter into the computation of the optimal subsidy program.

Observe that our subsidy rankings typically favor larger firms as they tend to be better connected in the R&D network than small firms. This adds to the discussion of whether large or small firms are contributing more to the innovativeness of an economy [cf. Mandel, 2011], by adding another dimension along which larger firms can have an advantage over small ones, namely by creating R&D spillover effects that contribute to the overall productivity of the economy. While studies such as Spencer and Brander [1983] and Acemoglu et al. [2012] find that R&D should often be taxed rather than subsidized, we find in line with e.g. Hinloopen [2001] that R&D subsidies can have a significantly positive effect on welfare. As argued by Hinloopen [2001], the reason why our results differ from those of Spencer and Brander [1983] is that we take into account the consumer surplus when deriving the optimal R&D subsidy. Moreover, in contrast to Acemoglu et al. [2012], we do not focus on entry and exit but incorporate the network of R&D collaborating firms. This allows us to take into account the R&D spillover effects of incumbent firms, which are typically ignored in studies of the innovative activity of incumbent firms versus entrants. Therefore, we see our analysis as complementary to that of Acemoglu et al. [2012], and we show that R&D subsidies can trigger considerable welfare gains when technology spillovers through R&D alliances are incorporated.

8. Discussion

In this section we discuss some assumptions of our model and their implications on the empirical and policy analysis.

Inertia of R&D networks One of the underlying assumptions of our model is that the R&D network exhibits inertia. That is, compared to making adjustments to production and R&D expenditures, it is relatively costly – both in terms of money and time – to form new alliances

Table 7: Subsidies ranking for the year 1990 for the first 25 firms.

Firm	Share [%][a]	num pat.	**d**	$\mathbf{v}_{\mathrm{PF}}$	Betweenness[b]	Closeness[c]	**q** [%][d]	hom. sub. [%][e]	tar. sub. [%][f]	SIC[g]	Rank
General Motors Corp.	9.2732	76644	88	0.1009	0.0007	0.0493	6.9866	0.0272	0.3027	3711	1
Exxon Corp.	7.7132	21954	22	0.0221	0.0000	0.0365	5.4062	0.0231	0.1731	2911	2
Ford Motor Co.	7.3456	20378	6	0.0003	0.0000	0.0153	3.7301	0.0184	0.0757	3711	3
AT&T Corp.	9.5360	5692	8	0.0024	0.0000	0.0202	3.2272	0.0156	0.0565	4813	4
Chevron	2.8221	12789	23	0.0226	0.0001	0.0369	2.5224	0.0098	0.0418	2911	5
Texaco	2.9896	9134	22	0.0214	0.0000	0.0365	2.4965	0.0095	0.0415	2911	6
Lockheed	42.3696	2	51	0.0891	0.0002	0.0443	1.5639	0.0035	0.0196	3760	7
Mobil Corp.	4.2265	3	0	0.0000	0.0000	0.0000	1.9460	0.0111	0.0191	2911	8
TRW Inc.	5.3686	9438	43	0.0583	0.0002	0.0415	1.4509	0.0027	0.0176	3714	9
Altria Group	43.6382	0	0	0.0000	0.0000	0.0000	1.4665	0.0073	0.0117	2111	10
Alcoa Inc.	11.4121	4546	36	0.0287	0.0002	0.0372	1.2136	0.0032	0.0114	3350	11
Shell Oil Co.	14.6777	9504	0	0.0000	0.0000	0.0000	1.4244	0.0073	0.0109	1311	12
Chrysler Corp.	2.2414	3712	6	0.0017	0.0000	0.0218	1.3935	0.0075	0.0109	3711	13
Schlumberger Ltd. Inc.	25.9218	9	18	0.0437	0.0000	0.0370	1.1208	0.0029	0.0099	1389	14
Hewlett-Packard Co.	7.1106	6606	64	0.1128	0.0002	0.0417	1.1958	0.0047	0.0093	3570	15
Intel Corp.	9.3900	1132	67	0.1260	0.0003	0.0468	1.0152	0.0018	0.0089	3674	16
Hoechst Celanese Corp.	5.6401	516	38	0.0368	0.0002	0.0406	1.0047	0.0021	0.0085	2820	17
Motorola	14.1649	21454	70	0.1186	0.0004	0.0442	1.0274	0.0028	0.0080	3663	18
PPG Industries Inc.	13.3221	24904	20	0.0230	0.0000	0.0366	0.9588	0.0021	0.0077	2851	19
Himont Inc.	0.0000	59	28	0.0173	0.0001	0.0359	0.8827	0.0014	0.0072	2821	20
GTE Corp.	3.1301	4	0	0.0000	0.0000	0.0000	1.1696	0.0067	0.0070	4813	21
National Semiconductor Corp.	4.0752	1642	43	0.0943	0.0001	0.0440	0.8654	0.0012	0.0068	3674	22
Marathon Oil Corp.	7.9828	202	0	0.0000	0.0000	0.0000	1.1306	0.0060	0.0068	1311	23
Bellsouth Corp.	2.4438	3	14	0.0194	0.0000	0.0329	1.0926	0.0060	0.0064	4813	24
Nynex	2.3143	26	24	0.0272	0.0001	0.0340	0.9469	0.0049	0.0052	4813	25

[a] Market share in the primary 4-digit SIC sector in which the firm is operating. In case of missing data the closest year with sales data available has been used.

[b] The normalized betweenness centrality is the fraction of all shortest paths in the network that contain a given node, divided by $(n-1)(n-2)$, the maximum number of such paths.

[c] The closeness centrality of node i is computed as $\frac{2}{n-1}\sum_{j=1}^{n} 2^{-\ell_{ij}(G)}$, where $\ell_{ij}(G)$ is the length of the shortest path between i and j in the network G and the factor $\frac{2}{n-1}$ is the maximal centrality attained for the center of a star network.

[d] The relative output of a firm i follows from Proposition 1.

[e] The homogeneous subsidy for each firm i is computed as $e_i^* s^*$, relative to the total homogeneous subsidies $\sum_{j=1}^{n} e_j^* s^*$ (see Proposition 2).

[f] The targeted subsidy for each firm i is computed as $e_i^* s_i^*$, relative to the total targeted subsidies $\sum_{j=1}^{n} e_j^* s_j^*$ (see Proposition 3).

[g] The primary 4-digit SIC code according to Compustat U.S. fundamentals database.

Table 8: Subsidies ranking for the year 2005 for the first 25 firms.

Firm	Share [%][a]	num pat.	$\mathbf{d}$	$\mathbf{v}_{\mathrm{PF}}$	Betweenness[b]	Closeness[c]	$\mathbf{q}$ [%][d]	hom. sub.[%][e]	tar. sub. [%][f]	SIC[g]	Rank
General Motors Corp.	3.9590	90652	19	0.0067	0.0002	0.0193	4.1128	0.0174	0.2186	3711	1
Ford Motor Co.	3.6818	27452	7	0.0015	0.0000	0.0139	3.4842	0.0153	0.1531	3711	2
Exxon Corp.	4.0259	53215	6	0.0007	0.0001	0.0167	2.9690	0.0132	0.1108	2911	3
Microsoft Corp.	10.9732	10639	62	0.1814	0.0020	0.0386	1.6959	0.0057	0.0421	7372	4
Pfizer Inc.	3.6714	74253	65	0.0298	0.0034	0.0395	1.6796	0.0069	0.0351	2834	5
AT&T Corp.	0.0000	16284	0	0.0000	0.0000	0.0000	1.5740	0.0073	0.0311	4813	6
Motorola	6.6605	70583	66	0.1598	0.0017	0.0356	1.3960	0.0053	0.0282	3663	7
Intel Corp.	5.0169	28513	72	0.2410	0.0011	0.0359	1.3323	0.0050	0.0249	3674	8
Chevron	2.2683	15049	10	0.0017	0.0001	0.0153	1.3295	0.0058	0.0243	2911	9
Hewlett-Packard Co.	14.3777	38597	7	0.0288	0.0000	0.0233	1.1999	0.0055	0.0183	3570	10
Altria Group	20.4890	5	2	0.0000	0.0000	0.0041	1.1753	0.0054	0.0178	2111	11
Johnson & Johnson Inc.	3.6095	31931	40	0.0130	0.0015	0.0346	1.1995	0.0051	0.0173	2834	12
Texaco	0.0000	10729	0	0.0000	0.0000	0.0000	1.0271	0.0055	0.0124	2911	13
Shell Oil Co.	0.0000	12436	0	0.0000	0.0000	0.0000	0.9294	0.0045	0.0108	1311	14
Chrysler Corp.	0.0000	5112	0	0.0000	0.0000	0.0000	0.9352	0.0052	0.0101	3711	15
Bristol-Myers Squibb Co.	1.3746	16	35	0.0052	0.0009	0.0326	0.8022	0.0034	0.0077	2834	16
Merck & Co. Inc.	1.5754	52036	36	0.0023	0.0007	0.0279	0.8252	0.0038	0.0077	2834	17
Marathon Oil Corp.	5.5960	229	0	0.0000	0.0000	0.0000	0.7817	0.0039	0.0076	1311	18
GTE Corp.	0.0000	5	0	0.0000	0.0000	0.0000	0.7751	0.0041	0.0073	4813	19
Pepsico	36.6491	991	0	0.0000	0.0000	0.0000	0.7154	0.0035	0.0066	2080	20
Bellsouth Corp.	0.9081	2129	0	0.0000	0.0000	0.0000	0.7233	0.0039	0.0063	4813	21
Johnson Controls Inc.	22.0636	304	11	0.0027	0.0001	0.0159	0.6084	0.0021	0.0063	2531	22
Dell	18.9098	80	2	0.0190	0.0000	0.0216	0.6586	0.0028	0.0061	3571	23
Eastman Kodak Co	5.5952	109714	17	0.0442	0.0001	0.0262	0.6171	0.0023	0.0060	3861	24
Lockheed	48.9385	9817	44	0.0434	0.0003	0.0223	0.6000	0.0028	0.0049	3760	25

[a] Market share in the primary 4-digit SIC sector in which the firm is operating. In case of missing data the closest year with sales data available has been used.

[b] The normalized betweenness centrality is the fraction of all shortest paths in the network that contain a given node, divided by $(n-1)(n-2)$, the maximum number of such paths.

[c] The closeness centrality of node i is computed as $\frac{2}{n-1}\sum_{j=1}^{n} 2^{-\ell_{ij}(G)}$, where $\ell_{ij}(G)$ is the length of the shortest path between i and j in the network G and the factor $\frac{2}{n-1}$ is the maximal centrality attained for the center of a star network.

[d] The relative output of a firm i follows from Proposition 1.

[e] The homogeneous subsidy for each firm i is computed as $e_i^* s^*$, relative to the total homogeneous subsidies $\sum_{j=1}^{n} e_j^* s^*$ (see Proposition 2).

[f] The targeted subsidy for each firm i is computed as $e_i^* s_i^*$, relative to the total targeted subsidies $\sum_{j=1}^{n} e_j^* s_j^*$ (see Proposition 3).

[g] The primary 4-digit SIC code according to Compustat U.S. fundamentals database.

in the R&D network. Therefore, we consider a short-run policy analysis, where we treat the R&D network as given and design the optimal subsidy program by taking into account the equilibrium production and R&D investment decisions of the firms. In the long run, the R&D network might itself also respond to the subsidy program, and thus, the design of a long-run subsidy program should take the evolution of the R&D network into account. However, such a dynamic forward looking network formation game would be very hard to solve, and despite the recent developments in analyzing network formation models, there does not exist a commonly agreed procedure for solving such dynamic problems. For this reason, we focus on a short-run policy analysis in this paper, leaving the long-run policy analysis for future work.

Independent markets In our basic model, we consider independent markets, i.e., firms only compete against firms in the same product market, but not against firms from different product markets. This assumption can be relaxed, however, in our theoretical framework. In Proposition 1 we characterize the Nash equilibrium with a single product market (i.e., $M = 1$), where all firms compete against each other. Furthermore, by allowing the elements of the competition matrix $\mathbf{B}$ to take arbitrary weights instead of the binary values 0 or 1, the competition matrix can be flexibly specified to represent more general market structures.

Based on these ideas, we conduct a robustness analysis for our empirical results with alternative specifications of the competition matrix. First, in Section J.2 of the Online Appendix, we re-estimate Equation (27) using two major sectors in our data, namely the manufacturing and services sectors, that, respectively, cover 76.8% and 19.3% firms in our sample. The estimated spillover and competition parameters of these two sectors are largely the same as those in our benchmark specification.

Next, in Section J.4 of the Online Appendix, we consider a richer specification of the $\mathbf{B}$ matrix. This extension follows Bloom et al. [2013] by considering three alternative specifications for the competition matrix based on the primary and secondary industry classification codes that can be found in (i) the Compustat Segments database, (ii) the Orbis database [cf. Bloom et al., 2013], or (iii) the Hoberg-Phillips product similarity database [cf. Hoberg and Phillips , 2016]. These alternative competition matrices capture (in a reduced form) the product portfolio of a firm by taking into account the different industries a firm is operating in. We find that irrespective of what type of competition matrix is being used, the estimated technology spillover effect is positively significant, with the magnitude similar to that obtained in the benchmark model. Moreover, the product rivalry effect with alternative specifications of the competition matrix is also statistically significant with the expected sign.

No input-output linkages Our theoretical model considers horizontally related firms, while it does not incorporate the possible vertical relationships of firms through input-output linkages. To test for potential R&D spillovers between vertically related firms, we conduct a robustness analysis by directly controlling for potential input-supplier effects. We obtain information about firms' buyer-supplier relationships from two data sources. The first is the Compustat Segments database [cf. e.g. Atalay et al., 2011; Barrot and Sauvagnat, 2016]. Compustat Segments provides business details, product information and customer data for over 70% of the companies in the Compustat North American database, with firms' coverage starting in the year 1976. We also use as a second datasource the Capital IQ Business Relationships database [Barrot and Sauvagnat, 2016; Lim, 2016; Mizuno et al., 2014]. The Capital IQ data includes any customers/suppliers that are mentioned in the firms' annual reports, news, websites surveys etc, with firms coverage starting in the year 1990. We then merged these two datasources to obtain a more complete picture of the potential buyer-supplier linkages between the firms in our R&D network. Aggregated over all years we obtained a total of $2,573$ buyer-supplier relationships for the firms matched with our R&D network dataset. Using these data on firms' buyer-supplier relationships, we find that, after controlling for the input-supplier effect, the spillover and competition effects remain statistically significant with the expected signs.

In terms or our policy analysis, the additional R&D spillover effects through input-output relationships could be easily incorporated in our optimal subsidy program by adding another technology spillover matrix from input-output linkages (as in Appendix F). However, as the focus of the current paper is on R&D collaborations, we consider only spillovers stemming from R&D alliances between firms.

No market entry and exit As we focus on a short-run policy analysis in this paper, we consider only incumbent firms and abstract from the complication of market entry and exit. This allows us to study the R&D spillover effects using a network approach, which are typically ignored in studies of innovative activities of incumbent firms versus entrants as for example

Acemoglu et al. [2012]. Therefore, we see our analysis as complementary to that of Acemoglu et al. [2012], and we show that R&D subsidies can trigger considerable welfare gains when technology spillovers through R&D alliances are taken into account.

No foreign firms Another possible extension of the current model is to partition the firms into domestic firms and foreign firms, and consider a subsidy program that only subsidizes domestic firms. This extension would be possible under our current framework as our targeted subsidy program is very flexible. In particular, it is allowed to assign zero subsidies to certain firms (e.g. foreign firms). However, we do not pursue this extension in this paper as in the data we only consider U.S. firms.

9. Conclusion

In this paper, we have developed a model where firms benefit form R&D collaborations (networks) to lower their production costs while at the same time competing on the product market. We have highlighted the positive role of the network in terms of technology spillovers and the negative role of product rivalry in terms of market competition. We have also determined the importance of targeted subsidies on the total welfare of the economy.

Using a panel of R&D alliance networks and annual reports, we have then tested our theoretical results and first showed that both, the technology spillover effect and the market competition effect have the expected signs and are significant. We have also identified the firms in our data that should be subsidized the most to maximize welfare in the economy. Finally, we have drawn some policy conclusions about optimal R&D subsidies from the results obtained over different sectors, as well as their temporal variation.

We believe that the methodology developed in this paper offers a fruitful way of analyzing the existence of R&D spillovers and their policy implications in terms of firms' subsidies across and within different industries. We also believe that putting forward the role of networks in terms of R&D collaborations is important to understanding the different aspects of these markets.

References

Acemoglu, D., Akcigit, U., Bloom, N. and W. Kerr (2012). Innovation, reallocation and growth. Stanford University Working Paper.

Akcigit, U. (2009). Firm size, innovation dynamics and growth. memo. University of Chicago.

Atalay, E., Hortacsu, A., Roberts, J. and C. Syverson (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13), 5199.

Ballester, C., Calvó-Armengol, A. and Y. Zenou (2006). Who's who in networks. wanted: The key player. *Econometrica* 74(5), 1403–1417.

Belderbos, R., Carree, M. and Lokshin, B. (2004). Cooperative R&D and firm performance. *Research policy* 33, 1477–1492.

Belhaj, M., Bervoets, S. and F. Deroïan (2016). Efficient networks in games with local complementarities. *Theoretical Economics* 11(1), 357–380.

games under strategic complementarities. *Games and Economic Behavior* 88, 310–319.

Bernstein, J.I. (1988). Costs of production, intra-and interindustry R&D spillovers: Canadian evidence. *Canadian Journal of Economics*, 324–347.

Bloom, N., Griffith, R. and J. Van Reenen (2002). Do R&D tax credits work? evidence from a panel of countries 1979–1997. *Journal of Public Economics* 85(1), 1–31.

Bloom, N., Schankerman, M. and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica* 81(4), 1347–1393.

Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), Identification of social interactions, In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics*, Vol. 1B, Amsterdam: Elsevier Science, pp. 853-964.

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

Bramoullé, Y., Kranton, R. and M. D'amours (2014). Strategic interaction and networks. *American Economic Review* 104 (3), 898–930

Buiter, Willem H. (1981). The Superiority of Contingent Rules Over Fixed Rules in Models with Rational Expectations. *The Economic Journal* 91(363), 647–670.

Byong-Hun, A. (1983). Iterative methods for linear complementarity problems with upperbounds on primary variables. *Mathematical Programming* 26(3), 295–315.

Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies* 76, 1239–1267.

Chandrasekhar, A. (2016). Econometrics of network formation. In: Y. Bramoullé, B.W. Rogers and A. Galeotti (Eds.), *Oxford Handbook on the Economics of Networks*, Oxford: Oxford University Press.

Chen, J. and S. Burer (2012). Globally solving nonconvex quadratic programming problems via completely positive programming. *Mathematical Programming Computation* 4(1), 33–52.

Cohen, L. (1994). When can government subsidize research joint ventures? Politics, economics, and limits to technology policy. *The American Economic Review*, 84(2):159–63.

Cohen, W. and S. Klepper (1996). A reprise of size and R&D. *Economic Journal* 106, 925–951.

Czarnitzki, D., Ebersberger, B., and Fier, A. (2007). The relationship between R&D collaboration, subsidies and R&D performance: Empirical evidence from Finland and Germany. *Journal of Applied Econometrics*, 22(7):1347–1366.

Cottle, R.W., Pang, J.-S. and R.E. Stone (1992). *The Linear Complementarity Problem.* Boston: Academic Press.

Cvetkovic, D., Doob, M. and H. Sachs (1995). *Spectra of Graphs: Theory and Applications.* Johann Ambrosius Barth.

Dai, R. (2012). International accounting databases on wrds: Comparative analysis. Working paper, Wharton Research Data Services, University of Pennsylvania.

D'Aspremont, C. and A. Jacquemin (1988). Cooperative and noncooperative R&D in duopoly with spillovers. *American Economic Review* 78(5), 1133–1137.

Dechezleprêtre, Antoine and Einiö, Elias and Martin, Ralf and Nguyen, Kieu-Trang and Van Reenen, John (2016). Do tax incentives for research increase firm innovation? An RD design for R&D. National Bureau of Economic Research, Working Paper No. w22405.

Einiö, E., 2014. R&D subsidies and company performance: Evidence from geographic variation in government funding based on the ERDF population-density rule. *Review of Economics and Statistics* 96(4), 710–728.

Fama, E.F. and K.R. French (1992). The cross-section of expected stock returns. *Journal of Finance* 47, 427–465.

Feldman, M. P., Kelley, M. R. (2003). Leveraging research and development: Assessing the impact of the U.S. Advanced Technology Program. *Small Business Economics* 20(2):153–165.

Feldman, M. P., Kelley, M. R., 2006. The ex ante assessment of knowledge spillovers: Government R&D policy, economic incentives and private firm behavior. *Research Policy* 35(10), 1509–1521.

Frank, M., Wolfe, P., 1956. An algorithm for quadratic programming. *Naval Research Logistics Quarterly* 3(1-2), 95–110.

Gal, P. N., 2013. Measuring total factor productivity at the firm level using OECD-ORBIS. OECD Working Paper, ECO/WKP(2013)41.

Galí, J., 1999. Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review* 89(1), 249–271.

Goyal, S. and J.L. Moraga-Gonzalez (2001). R&D networks. *RAND Journal of Economics* 32 (4), 686–707.

Griffith, R., Redding, S. and J. Van Reenen (2004). Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *Review of Economics and Statistics* 86 (4), 883–895.

Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research Policy* 31(4), 477–492.

Hall, B. H., Jaffe, A. B., Trajtenberg, M., 2000. Market value and patent citations: A first look. National Bureau of Economic Research, Working Paper No. 7741.

Hanaki, N., Nakajima, R., Ogura, Y., 2010. The dynamics of R&D network in the IT industry. *Research Policy* 39(3), 386–399.

Hinloopen, J., 2000. More on subsidizing cooperative and noncooperative R&D in duopoly with spillovers. *Journal of Economics* 72(3), 295–308.

Hinloopen, J., 2001. Subsidizing R&D cooperatives. *De Economist* 149(3), 313–345.

Horn, R. A., Johnson, C. R., 1990. *Matrix Analysis*. Cambridge University Press.

Impullitti, G. (2010). International competition and U.S. R&D subsidies: A quantitative welfare analysis. *International Economic Review* 51(4), 1127–1158.

Hoberg, Gerard and Phillips, Gordon (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124(5), 1423–1465.

Jackson, M. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O., Rogers, B., Zenou, Y. (2017). The economic consequences of social network structure. *Journal of Economic Literature* 55, 49–95.

Jackson, M. O. and Y. Zenou (2015). Games on networks. In: P. Young and S. Zamir (Eds.), *Handbook of Game Theory with Economic Applications* vol. 4, pp. 95–163.

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

Kelejian, H.H. and G. Piras (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics* 46, 140–149.

Klette, T. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy* 112(5), 986–1018.

Klette, T. J., Møen, J. and Z. Griliches (2000). Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies. *Research Policy* 29(4), 471–495.

König, M.D. and Liu, X. and Y. Zenou (2014). R&D Networks: Theory, Empirics and Policy Implications. CEPR Discussion Paper No. 9872.

Leahy, D. and J.P. Neary (1997). Public policy towards R&D in oligopolistic industries. *American Economic Review*, 642–662.

Lee, L. and X. Liu (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26(1), 187–230.

Mandel, M., 2011. Scale and innovation in today's economy. Progressive Policy Institute Policy Memo.

Roijakkers, N., Hagedoorn, J., April 2006. Inter-firm R&D partnering in pharmaceutical biotechnology since 1975: Trends, patterns, and networks. *Research Policy* 35(3), 431–446.

Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal* 30(3), 233–260.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51(5):756–770.
Singh, N., Vives, X., 1984. Price and quantity competition in a differentiated duopoly. *RAND Journal of Economics* 15(4), 546–554.
Spencer, B. J., Brander, J. A., 1983. International R & D rivalry and industrial strategy. *Review of Economic Studies* 50(4), 707–722.
Stock, J. H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: D.W.K. Andrews (Ed.), Identification and Inference for Econometric Models. New York: Cambridge University Press, pp. 80–108.
Suzumura, K. (1992). Cooperative and Noncooperative R&D in an Oligopoly with Spillovers. *American Economic Review* 82(5), 1307–1320.
Takalo, T., Tanayama, T. and O. Toivanen (2013). Market failures and the additionality effects of public support to private R&D.*International Journal of Industrial Organization* 31(5), 634–642.
Takalo, T., Tanayama, T., and O. Toivanen (2017). Welfare effects of R&D support policies. CEPR Discussion Paper No. 12155.
Trajtenberg, M., Shiff, G. and R. Melamed (2009). The "names game": Harnessing inventors, patent data for economic research. *Annals of Economics and Statistics*, 79–108.
Wansbeek, T. and A. Kapteyn (1989). Estimation of the error-components model with incomplete panels. *Journal of Econometrics* 41(3), 341–361.
Westbrock, B. (2010). Natural concentration in industrial research collaboration. *RAND Journal of Economics* 41(2), 351–371.
Wilson, D.J. (2009). Beggar thy neighbor? The in-state, out-of-state, and aggregate effects of R&D tax credits. *Review of Economics and Statistics* 91(2), 431–436.
Zunica-Vicente, J.A., Alonso-Borrego, C., Forcadell, F. and J. Galan 2014. Assessing the effect of public subsidies on firm R&D investment: A survey. *Journal of Economic Surveys* 28, 36–67.

Online Appendix for "R&D Networks: Theory, Empirics and Policy Implications"

Michael D. König[a], Xiaodong Liu[b], Yves Zenou[c]

[a] *Department of Economics, University of Zurich, Schönberggasse 1, CH-8001 Zurich, Switzerland.*
[b] *Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309–0256, United States.*
[c] *Department of Economics, Monash University, Caulfield VIC 3145, Australia, and IFN.*

Contents

A. Proofs

Proof of Proposition 1 (i) The FOCs of maximizing the profit function given by Equation (4) with respect to the R&D effort e_i and the output q_i of firm i are given by

$$\frac{\partial \pi_i}{\partial e_i} = q_i - e_i = 0,$$
$$\frac{\partial \pi_i}{\partial q_i} = \mu_i - 2q_i - \rho \sum_{j=1}^{n} b_{ij} q_j + e_i + \varphi \sum_{j=1}^{n} a_{ij} e_j = 0,$$

where $\mu_i \equiv \bar{\alpha}_i - \bar{c}_i$. Solving the FOCs gives

$$e_i = q_i, \tag{A.1}$$
$$q_i = \mu_i - \rho \sum_{j=1}^{n} b_{ij} q_j + \varphi \sum_{j=1}^{n} a_{ij} q_j, \tag{A.2}$$

or, in vector-matrix form,

$$\mathbf{e} = \mathbf{q},$$
$$\mathbf{q} = \boldsymbol{\mu} - \rho \mathbf{B}\mathbf{q} + \varphi \mathbf{A}\mathbf{q}.$$

Therefore, there exists a unique Nash equilibrium with the equilibrium outputs and R&D efforts given by Equation (6) if the matrix $\mathbf{I} + \rho\mathbf{B} - \varphi\mathbf{A}$ is positive definite. The symmetric matrix $\mathbf{I}_n + \rho\mathbf{B} - \varphi\mathbf{A}$ is positive definite if its smallest eigenvalue is positive, that is when

$$1 + \lambda_{\min}(\rho\mathbf{B} - \varphi\mathbf{A}) > 0. \tag{A.3}$$

First we consider the case of $\varphi = 0$. In this case, Equation (A.3) becomes $1 + \rho\lambda_{\min}(\mathbf{B}) > 0$. Since $\mathbf{B}$ can be written as a block diagonal matrix with a zero diagonal and blocks of sizes $|\mathcal{M}_m|$, $m = 1, \ldots, M$, the spectrum (set of eigenvalues) of $\mathbf{B}$ is given by $\{|\mathcal{M}_1| - 1, |\mathcal{M}_2| - 1, ..., |\mathcal{M}_M| - 1, -1, \ldots, -1\}$, with $\lambda_{\min}(\mathbf{B}) = -1$. As $0 \leq \rho < 1$, $1 + \rho\lambda_{\min}(\mathbf{B}) > 0$ and thus Equation (A.3) holds. Next we consider the general case that φ may not be zero. In this case, Equation (A.3) is equivalent to $\lambda_{\max}(\varphi\mathbf{A} - \rho\mathbf{B}) < 1$. Since $\lambda_{\max}(\varphi\mathbf{A} - \rho\mathbf{B}) \leq \varphi\lambda_{\max}(\mathbf{A}) + \rho\lambda_{\max}(\mathbf{B})$ and $\lambda_{\max}(\mathbf{B}) = \max_{m=1,\ldots,M}\{|\mathcal{M}_m| - 1\}$,[1] a sufficient condition for Equation (A.3) to hold is given by Equation (5). Finally, substitution of the equilibrium outputs and R&D efforts given by Equation (6) into the profit function (4) gives the equilibrium profits in Equation (7).

(ii) When all firms operate in the same market so that $M = 1$, the best response function given by Equation (A.2) can be written as

$$q_i = \frac{1}{1-\rho}\mu_i - \frac{\rho}{1-\rho}\hat{q} + \frac{\varphi}{1-\rho}\sum_{j=1}^{n} a_{ij} q_j. \tag{A.4}$$

where $\hat{q} \equiv \sum_{j=1}^{n} q_j$ corresponds to the total output of all firms. Observe that $0 < 1 - \rho \leq 1$

[1] Let $\|\cdot\|$ be any matrix norm, including the spectral norm, which is just the largest eigenvalue. Then we have that $\|\sum_{i=1}^{n} \alpha_i \mathbf{A}_i\| \leq \sum_{i=1}^{n} |\alpha_i| \|\mathbf{A}_i\| \leq (\sum_{i=1}^{n} |\alpha_i|) \max_i \|\mathbf{A}_i\|$ by Weyl's theorem [cf. e.g. Horn and Johnson, 1990, Theorem 4.3.1].

as $0 \leq \rho < 1$. In matrix form, Equation (A.4) can be written as

$$(\mathbf{I} - \phi\mathbf{A})\mathbf{q} = \frac{1}{1-\rho}(\boldsymbol{\mu} - \rho\hat{q}\boldsymbol{\iota}),$$

where $\phi = \varphi/(1-\rho)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, and $\boldsymbol{\iota} = (1, \ldots, 1)^\top$. If $\phi < \lambda_{\max}(\mathbf{A})^{-1}$, this is equivalent to

$$\mathbf{q} = \frac{1}{1-\rho}\left(\mathbf{b}_{\boldsymbol{\mu}}(G,\phi) - \rho\hat{q}\,\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right), \tag{A.5}$$

where $\mathbf{b}_{\boldsymbol{\iota}}(G,\phi) = (\mathbf{I} - \phi\mathbf{A})^{-1}\boldsymbol{\iota}$ is the vector of unweighted Katz-Bonacich centralities and $\mathbf{b}_{\boldsymbol{\mu}}(G,\phi) = (\mathbf{I} - \phi\mathbf{A})^{-1}\boldsymbol{\mu}$ is the vector of weighted Katz-Bonacich centralities with the weights given by μ_i for $i = 1, \ldots, n$. Premultiplying Equation (A.5) by $\boldsymbol{\iota}^\top$, we obtain

$$(1-\rho)\,\hat{q} = \left\|\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)\right\|_1 - \rho\hat{q}\,\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1,$$

where $\left\|\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)\right\|_1 = \boldsymbol{\iota}^\top\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)$ is the sum of the weighted Katz-Bonacich centralities and $\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1 = \boldsymbol{\iota}^\top\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)$ is the sum of the unweighted Katz-Bonacich centralities. Solving this equation, we get

$$\hat{q} = \frac{\left\|\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)\right\|_1}{(1-\rho) + \rho\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1}.$$

Plugging this value of $\hat{q}$ into Equation (A.5), we finally obtain Equation (8) in the proposition.

In the following we provide a condition which guarantees that the equilibrium outputs given by Equation (8) are positive. According to Equation (8), $\underline{\mathbf{q}}^* > \mathbf{0}$ if and only if

$$\mathbf{b}_{\boldsymbol{\mu}}(G,\phi) > \frac{\rho\left\|\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)\right\|_1}{(1-\rho) + \rho\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1}\mathbf{b}_{\boldsymbol{\iota}}(G,\phi). \tag{A.6}$$

Denote by $\underline{\mu} = \min_i \{\mu_i \mid i \in N\}$ and $\overline{\mu} = \max_i \{\mu_i \mid i \in N\}$, with $\underline{\mu} \leq \overline{\mu}$. Then, we have

$$\begin{aligned}\left\|\mathbf{b}_{\boldsymbol{\mu}}(G,\phi)\right\|_1 &\leq \overline{\mu}\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1,\\ \mathbf{b}_{\boldsymbol{\mu}}(G,\phi) &\geq \underline{\mu}\,\mathbf{b}_{\boldsymbol{\iota}}(G,\phi).\end{aligned}$$

Thus, a sufficient condition for Equation (A.6) to hold is

$$\underline{\mu}\,\mathbf{b}_{\boldsymbol{\iota}}(G,\phi) > \frac{\rho\overline{\mu}\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1}{(1-\rho) + \rho\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1}\mathbf{b}_{\boldsymbol{\iota}}(G,\phi),$$

or equivalently

$$1-\rho > \rho\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1\left(\frac{\overline{\mu}}{\underline{\mu}} - 1\right). \tag{A.7}$$

Next, observe that, by definition

$$\left\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\right\|_1 = \sum_{p=0}^{\infty}\phi^p\boldsymbol{\iota}^\top\mathbf{A}^p\boldsymbol{\iota}. \tag{A.8}$$

We know that $\lambda_{\max}(\mathbf{A}^p) \leq \lambda_{\max}(\mathbf{A})^p$, for all $p \geq 0$.[2] Also, $\boldsymbol{\iota}^\top\mathbf{A}^p\boldsymbol{\iota}/n$ is the average connec-

[2]Observe that the relationship $\lambda_{\max}(\mathbf{A}^p) = \lambda_{\max}(\mathbf{A})^p$, $p \geq 0$, holds true for both symmetric as well as asymmetric adjacency matrices $\mathbf{A}$ as long as $\mathbf{A}$ has non-negative entries, $a_{ij} \geq 0$.

tivity in the matrix $\mathbf{A}^p$ of paths of length p in the original network $\mathbf{A}$, which is smaller than its spectral radius $\lambda_{\max}(\mathbf{A}^p)$ [Cvetkovic et al., 1995], i.e. $\boldsymbol{\iota}^\top \mathbf{A}^p \boldsymbol{\iota}/n \leq \lambda_{\max}(\mathbf{A}^p) \leq \lambda_{\max}(\mathbf{A})^p$. Therefore, Equation (A.8) leads to the following inequality

$$\|\mathbf{b}_{\boldsymbol{\iota}}(G,\phi)\|_1 = \sum_{p=0}^{\infty} \phi^p \boldsymbol{\iota}^\top \mathbf{A}^p \boldsymbol{\iota} \leq n \sum_{p=0}^{\infty} \phi^p \lambda_{\max}(\mathbf{A})^p = \frac{n}{1-\phi\lambda_{\max}(\mathbf{A})}.$$

A sufficient condition for Equation (A.7) to hold is thus given by Equation (9). In the case that all firms are homogenous, $\overline{\mu}/\underline{\mu} = 1$, and Equation (A.7) holds as $0 \leq \rho < 1$.

(iii) When $\rho = 0$, if $\varphi < \lambda_{\max}(\mathbf{A})^{-1}$, the matrix $\mathbf{I} - \varphi\mathbf{A}$ is nonsingular. From the FOCs of profit maximization, the equilibrium R&D efforts and outputs are given by

$$\overline{\mathbf{e}}^* = \overline{\mathbf{q}}^* = (\mathbf{I} - \varphi\mathbf{A})^{-1}\boldsymbol{\mu} = \sum_{p=0}^{\infty} \varphi^p \mathbf{A}^p \boldsymbol{\mu} > \mathbf{0}.$$

(iv) Let $\mathbf{B}$ denote the competition matrix with an arbitrary number of markets. Under the competition matrix $\mathbf{B}$, the Nash equilibrium output levels are the solution to the following system of equations

$$q_i = f_i(\mathbf{q}) \equiv \mu_i - \rho\sum_{j=1}^{n} b_{ij}q_j + \varphi\sum_{j=1}^{n} a_{ij}q_j. \tag{A.9}$$

We can compare this to the Nash equilibrium output levels with a single market, which solve

$$q_i = \underline{f}_i(\mathbf{q}) \equiv \mu_i - \rho\sum_{j=1, j\neq i}^{n} q_j + \varphi\sum_{j=1}^{n} a_{ij}q_j,$$

and the Nash equilibrium output levels with non-substitutable goods, which solve

$$q_i = \overline{f}_i(\mathbf{q}) \equiv \mu_i + \varphi\sum_{j=1}^{n} a_{ij}q_j.$$

As $\overline{f}_i(\mathbf{q}) \geq f_i(\mathbf{q}) \geq \underline{f}_i(\mathbf{q})$ when $\mathbf{q} > \mathbf{0}$, the desired result follows by the comparison lemma (cf. Lemma 3.4 in Khalil [2002]).

$\square$

Proof of Propositions 2 and 3 As Proposition 2 is a special case of Proposition 3 with $s_i = s$ for $i = 1, \ldots, n$, we give the proof of the two propositions together.

(i) The FOCs of maximizing the profit function given by Equation (18) with respect to the R&D effort e_i and the output q_i of firm i are given by

$$\frac{\partial \pi_i}{\partial e_i} = q_i - e_i + s_i = 0,$$
$$\frac{\partial \pi_i}{\partial q_i} = \mu_i - 2q_i - \rho\sum_{j=1}^{n} b_{ij}q_j + e_i + \varphi\sum_{j=1}^{n} a_{ij}e_j = 0,$$

where $\mu_i \equiv \bar{\alpha}_i - \bar{c}_i$. Solving the FOCs gives

$$
\begin{aligned}
e_i &= q_i + s_i, \\
q_i &= \mu_i - \rho \sum_{j=1}^{n} b_{ij} q_j + \varphi \sum_{j=1}^{n} a_{ij} q_j + s_i + \varphi \sum_{j=1}^{n} a_{ij} s_j,
\end{aligned}
$$

or, in vector-matrix form,

$$
\begin{aligned}
\mathbf{e} &= \mathbf{q} + \mathbf{s}, \\
\mathbf{q} &= \boldsymbol{\mu} - \rho \mathbf{B} \mathbf{q} + \varphi \mathbf{A} \mathbf{q} + \mathbf{s} + \varphi \mathbf{A} \mathbf{s}.
\end{aligned}
$$

Therefore, there exists a unique Nash equilibrium with the equilibrium outputs and R&D efforts given by Equations (19) and (20) if the matrix $\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A}$ is positive definite. From the proof of Proposition 1, a sufficient condition for the matrix $\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A}$ to be positive definite is $\varphi = 0$ or the condition given by Equation (5) holds. Substitution of Equations (19) and (20) into the profit function given by Equation (18) gives the equilibrium profits in Equation (21). Equations (14) and (15) can be obtained by replacing $\mathbf{s}$ in Equations (19) and (20) by $s\boldsymbol{\iota}$. Substitution of Equations (14) and (15) into the profit function given by Equation (13) gives the equilibrium profits in Equation (16).

(ii) The net welfare can be written as

$$
\begin{aligned}
\overline{W}(G, \mathbf{s}) &= \frac{1}{2} \left(\sum_{i=1}^{n} (q_i^*)^2 + \rho \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} q_i^* q_j^* \right) + \sum_{i=1}^{n} \pi_i^* - \sum_{i=1}^{n} s_i e_i^* \\
&= \sum_{i=1}^{n} (q_i^*)^2 - \sum_{i=1}^{n} q_i^* s_i - \frac{1}{2} \sum_{i=1}^{n} s_i^2 + \frac{\rho}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} q_i^* q_j^* \\
&= \mathbf{q}^{*\top} \mathbf{q}^* - \frac{1}{2} (\mathbf{q}^{*\top} \mathbf{s} + \mathbf{s}^\top \mathbf{q}*) - \frac{1}{2} \mathbf{s}^\top \mathbf{s} + \frac{\rho}{2} \mathbf{q}^{*\top} \mathbf{B} \mathbf{q}^*.
\end{aligned}
$$

Using the fact that $\mathbf{q}^* = \tilde{\mathbf{q}} + \mathbf{R}\mathbf{s}$, where $\tilde{\mathbf{q}} \equiv (\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A})^{-1} \boldsymbol{\mu}$ and $\mathbf{R} \equiv (\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A})^{-1} (\mathbf{I} + \varphi \mathbf{A})$, we can write the net welfare as

$$
\overline{W}(G, \mathbf{s}) = \tilde{\mathbf{q}}^\top \tilde{\mathbf{q}} + \frac{\rho}{2} \tilde{\mathbf{q}}^\top \mathbf{B} \tilde{\mathbf{q}} + \mathbf{s}^\top (2\mathbf{R} + \rho \mathbf{B} \mathbf{R} - \mathbf{I})^\top \tilde{\mathbf{q}} - \frac{1}{2} \mathbf{s}^\top \mathbf{H} \mathbf{s}, \tag{A.10}
$$

where

$$
\mathbf{H} = \mathbf{I} + \mathbf{R} + \mathbf{R}^\top - 2\mathbf{R}^\top \mathbf{R} - \rho \mathbf{R}^\top \mathbf{B} \mathbf{R}.
$$

Observe that the matrix $\mathbf{H}$ is symmetric. The FOC of maximizing the net welfare with respect to $\mathbf{s}$ is given by

$$
\frac{\partial \overline{W}(G, \mathbf{s})}{\partial \mathbf{s}} = (2\mathbf{R} + \rho \mathbf{B} \mathbf{R} - \mathbf{I})^\top \tilde{\mathbf{q}} - \mathbf{H} \mathbf{s} = 0,
$$

with the hessian given by $\frac{\partial^2 \overline{W}(G, \mathbf{s})}{\partial \mathbf{s} \partial \mathbf{s}^\top} = -\mathbf{H}$. When the matrix $\mathbf{H}$ is positive definite, we obtain a global maximum for the concave quadratic optimization problem with the optimal subsidy levels given by Equation (22). To obtain the optimal homogenous subsidy level given by Equation (17), replace $\mathbf{s}$ in the net welfare given by Equation (A.10) by $s\boldsymbol{\iota}$ and maximize the net welfare with respect to s.

□

B. Definitions and Characterizations

B.1. Network Definitions

A *network* (*graph*) $G \in \mathcal{G}^n$ is the pair $(\mathcal{N}, \mathcal{E})$ consisting of a set of *nodes* (*vertices*) $\mathcal{N} = \{1, \ldots, n\}$ and a set of edges (*links*) $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ between them, where $\mathcal{G}^n$ denotes the family of undirected graphs with n nodes. A link (i, j) is *incident* with nodes i and j. The *neighborhood* of a node $i \in \mathcal{N}$ is the set $\mathcal{N}_i = \{j \in \mathcal{N} : (i, j) \in \mathcal{E}\}$. The *degree* d_i of a node $i \in \mathcal{N}$ gives the number of links incident to node i. Clearly, $d_i = |\mathcal{N}_i|$. Let $\mathcal{N}_i^{(2)} = \bigcup_{j \in \mathcal{N}_i} \mathcal{N}_j \setminus (\mathcal{N}_i \cup \{i\})$ denote the second-order neighbors of node i. Similarly, the k-th order neighborhood of node i is defined recursively from $\mathcal{N}_i^{(0)} = \{i\}$, $\mathcal{N}_i^{(1)} = \mathcal{N}_i$ and $\mathcal{N}_i^{(k)} = \bigcup_{j \in \mathcal{N}_i^{(k-1)}} \mathcal{N}_j \setminus \left(\bigcup_{l=0}^{k-1} \mathcal{N}_i^{(l)} \right)$. A *walk* in G of length k from i to j is a sequence $\langle i_0, i_1, \ldots, i_k \rangle$ of nodes such that $i_0 = i$, $i_k = j$, $i_p \neq i_{p+1}$, and i_p and i_{p+1} are (directly) linked, that is $i_p i_{p+1} \in \mathcal{E}$, for all $0 \leq p \leq k-1$. Nodes i and j are said to be *indirectly linked* in G if there exists a walk from i to j in G containing nodes other than i and j. A pair of nodes i and j is connected if they are either directly or indirectly linked. A node $i \in \mathcal{N}$ is *isolated* in G if $\mathcal{N}_i = \emptyset$. The network G is said to be *empty* (denoted by $\overline{K}_n$) when all its nodes are isolated.

A *subgraph*, G', of G is the graph of subsets of the nodes, $\mathcal{N}(G') \subseteq \mathcal{N}(G)$, and links, $\mathcal{E}(G') \subseteq \mathcal{E}(G)$. A graph G is *connected*, if there is a path connecting every pair of nodes. Otherwise G is disconnected. The *components* of a graph G are the maximally connected subgraphs. A component is said to be *minimally connected* if the removal of any link makes the component disconnected.

A *dominating set* for a graph $G = (\mathcal{N}, \mathcal{E})$ is a subset $\mathcal{S}$ of $\mathcal{N}$ such that every node not in $\mathcal{S}$ is connected to at least one member of S by a link. An *independent set* is a set of nodes in a graph in which no two nodes are adjacent. For example the central node in a star $K_{1,n-1}$ forms a dominating set while the peripheral nodes form an independent set.

Let $G = (\mathcal{N}, \mathcal{E})$ be a graph whose distinct positive degrees are $d_{(1)} < d_{(2)} < \ldots < d_{(k)}$, and let $d_0 = 0$ (even if no agent with degree 0 exists in G). Furthermore, define $\mathcal{D}_i = \{v \in \mathcal{N} : d_v = d_{(i)}\}$ for $i = 0, \ldots, k$. Then the set-valued vector $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ is called the *degree partition* of G.

Consider a *nested split graph* $G = (\mathcal{N}, \mathcal{E})$ and let $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ be its degree partition [cf. Cvetkovic and Rowlinson, 1990; Mahadev and Peled, 1995]. Then the nodes $\mathcal{N}$ can be partitioned in independent sets $\mathcal{D}_i$, $i = 1, \ldots, \left\lfloor \frac{k}{2} \right\rfloor$ and a dominating set $\bigcup_{i=\left\lfloor \frac{k}{2} \right\rfloor+1}^{k} \mathcal{D}_i$ in the graph $G' = (\mathcal{N} \setminus \mathcal{D}_0, \mathcal{E})$. Moreover, the neighborhoods of the nodes are nested, such that the set of neighbors of each node is contained in the set of neighbors of each higher degree node. In particular, for each node $v \in \mathcal{D}_i$, $\mathcal{N}_v = \bigcup_{j=1}^{i} \mathcal{D}_{k+1-j}$ if $i = 1, \ldots, \left\lfloor \frac{k}{2} \right\rfloor$ if $i = 1, \ldots, k$, while $\mathcal{N}_v = \bigcup_{j=1}^{i} \mathcal{D}_{k+1-j} \setminus \{v\}$ if $i = \left\lfloor \frac{k}{2} \right\rfloor + 1, \ldots, k$.

In a *complete* graph K_n, every node is adjacent to every other node. The graph in which no pair of nodes is adjacent is the empty graph $\overline{K}_n$. A *clique* $K_{n'}$, $n' \leq n$, is a complete subgraph of the network G. A graph is *k-regular* if every node i has the same number of links $d_i = k$ for all $i \in \mathcal{N}$. The complete graph K_n is $(n-1)$-regular. The cycle C_n is 2-regular. In a *bipartite graph* there exists a partition of the nodes in two disjoint sets $\mathcal{S}_1$ and $\mathcal{S}_2$ such that each link connects a node in $\mathcal{S}_1$ to a node in $\mathcal{S}_2$. $\mathcal{S}_1$ and $\mathcal{S}_2$ are independent sets with cardinalities n_1 and n_2, respectively. In a complete bipartite graph K_{n_1,n_2} each node in $\mathcal{S}_1$ is connected to each other node in $\mathcal{S}_2$. The *star* $K_{1,n-1}$ is a complete bipartite graph in which $n_1 = 1$ and $n_2 = n-1$.

The *complement* of a graph G is a graph $\overline{G}$ with the same nodes as G such that any two nodes of $\overline{G}$ are adjacent if and only if they are not adjacent in G. For example the complement of the complete graph K_n is the empty graph $\overline{K}_n$.

Let $\mathbf{A}$ be the symmetric $n \times n$ *adjacency matrix* of the network G. The element $a_{ij} \in \{0, 1\}$ indicates if there exists a link between nodes i and j such that $a_{ij} = 1$ if $(i, j) \in \mathcal{E}$ and

$a_{ij} = 0$ if $(i,j) \notin \mathcal{E}$. The k-th power of the adjacency matrix is related to walks of length k in the graph. In particular, $\left(\mathbf{A}^k\right)_{ij}$ gives the number of walks of length k from node i to node j. The *eigenvalues* of the adjacency matrix $\mathbf{A}$ are the numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ has a nonzero solution vector $\mathbf{v}_i$, which is an *eigenvector* associated with λ_i for $i = 1, \ldots, n$. Since the adjacency matrix $\mathbf{A}$ of an undirected graph G is real and symmetric, the eigenvalues of $\mathbf{A}$ are real, $\lambda_i \in \mathbb{R}$ for all $i = 1, \ldots, n$. Moreover, if $\mathbf{v}_i$ and $\mathbf{v}_j$ are eigenvectors for different eigenvalues, $\lambda_i \neq \lambda_j$, then $\mathbf{v_i}$ and $\mathbf{v}_j$ are orthogonal, i.e. $\mathbf{v}_i^\top \mathbf{v}_j = 0$ if $i \neq j$. In particular, $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of $\mathbf{A}$. Since $\mathbf{A}$ is a real symmetric matrix, there exists an orthogonal matrix $\mathbf{S}$ such that $\mathbf{S}^\top \mathbf{S} = \mathbf{S}\mathbf{S}^\top = \mathbf{I}$ (that is $\mathbf{S}^\top = \mathbf{S}^{-1}$) and $\mathbf{S}^\top \mathbf{A}\mathbf{S} = \mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix of eigenvalues of $\mathbf{A}$ and the columns of $\mathbf{S}$ are the corresponding eigenvectors. The *Perron-Frobenius eigenvalue* $\lambda_{\mathrm{PF}}(G)$ is the *largest real eigenvalue* of $\mathbf{A}$ associated with G, i.e. all eigenvalues λ_i of $\mathbf{A}$ satisfy $|\lambda_i| \leq \lambda_{\mathrm{PF}}(G)$ for $i = 1, \ldots, n$ and there exists an associated nonnegative eigenvector $\mathbf{v}_{\mathrm{PF}} \geq 0$ such that $\mathbf{A}\mathbf{v}_{\mathrm{PF}} = \lambda_{\mathrm{PF}}(G)\mathbf{v}_{\mathrm{PF}}$. For a connected graph G the adjacency matrix $\mathbf{A}$ has a unique largest real eigenvalue $\lambda_{\max}(G)$ and a positive associated eigenvector $\mathbf{v}_{\mathrm{PF}} > 0$. The largest eigenvalue $\lambda_{\max}(G)$ has been suggested to measure the irregularity of a graph [Bell, 1992], and the components of the associated eigenvector $\mathbf{v}_{\mathrm{PF}}$ are a measure for the centrality of a node in the network. A measure $C_v : \mathcal{G} \to [0,1]$ for the centralization of the network G has been introduced by Freeman [1979] for generic centrality measures $\mathbf{v}$. In particular, the *centralization* C_v of G is defined as $C_v(G) \equiv \sum_{i \in G} (v_{i^*} - v_i) / \max_{G' \in \mathcal{G}^n} \sum_{j \in G'} (v_{j^*} - v_j)$, where i^* and j^* are the nodes with the highest values of centrality in the networks G, G', respectively, and the maximum in the denominator is computed over all networks $G' \in \mathcal{G}^n$ with the same number n of nodes. There also exists a relation between the number of walks in a graph and its eigenvalues. The number of closed walks of length k from a node i in G to herself is given by $\left(\mathbf{A}^k\right)_{ii}$ and the total number of closed walks of length k in G is $\mathrm{tr}\left(\mathbf{A}^k\right) = \sum_{i=1}^n \left(\mathbf{A}^k\right)_{ii} = \sum_{i=1}^n \lambda_i^k$. We further have that $\mathrm{tr}\left(\mathbf{A}\right) = 0$, $\mathrm{tr}\left(\mathbf{A}^2\right)$ gives twice the number of links in G and $\mathrm{tr}\left(\mathbf{A}^3\right)$ gives six times the number of triangles in G.

A nested split graph is characterized by a *stepwise adjacency matrix* $\mathbf{A}$, which is a symmetric, binary $(n \times n)$-matrix with elements a_{ij} satisfying the following condition: if $i < j$ and $a_{ij} = 1$ then $a_{hk} = 1$ whenever $h < k \leq j$ and $h \leq i$. Both, the complete graph, K_n, as well as the star $K_{1,n-1}$, are particular examples of nested split graphs. Nested split graphs are also the graphs which maximize the largest eigenvalue, $\lambda_{\max}(G)$, [Brualdi and Solheid, 1986], and they are the ones that maximize the degree variance [Peled et al., 1999].[3]

The *cores* of a graph are defined as follows: Given a network G, the induced subgraph $G_k \subseteq G$ is the k-core of G if it is the largest subgraph such that the degree of all nodes in G_k is at least k. Note that the cores of a graph are nested such that $G_{k+1} \subseteq G_k$. Cores can be used as a measure of centrality in the network G, and the largest k-core centrality across all nodes in the graph is called the *degeneracy* of G. Note that k-cores can be obtained by a simple pruning algorithm: at each step, we remove all nodes with degree less than k. We repeat this procedure until there exist no such nodes or all nodes are removed. We define the coreness of each node as follows: The coreness of node i, cor_i, is k if and only if $i \in G_k$ and $i \notin G_{k+1}$. We have that $\mathrm{cor}_i \leq d_i$. However, there is no other relation between the degree and coreness of nodes in a graph.

B.2. Walk Generating Functions

Denote by $\boldsymbol{\iota} = (1, \ldots, 1)^\top$ the n-dimensional vector of ones and define $\mathbf{M}(G, \phi) = (\mathbf{I} - \phi\mathbf{A})^{-1}$. Then, the quantity $N_G(\phi) = \boldsymbol{\iota}^\top \mathbf{M}(G, \phi)\boldsymbol{\iota}$ is the *walk generating function* of the graph G [cf.

[3] See for example König et al. [2014] for a discussion of further properties of nested split graphs.

Cvetkovic et al., 1995]. Let N_k denote the number of walks of length k in G. Then we can write N_k as follows

$$N_k = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^{[k]} = \boldsymbol{\iota}^\top \mathbf{A}^k \boldsymbol{\iota},$$

where $a_{ij}^{[k]}$ is the ij-th element of $\mathbf{A}^k$. The walk generating function is then defined as

$$N_G(\phi) \equiv \sum_{k=0}^{\infty} N_k \phi^k = \boldsymbol{\iota}^\top \left(\sum_{k=0}^{\infty} \phi^k \mathbf{A}^k \right) \boldsymbol{\iota} = \boldsymbol{\iota}^\top \left(\mathbf{I} - \phi \mathbf{A}\right)^{-1} \boldsymbol{\iota} = \boldsymbol{\iota}^\top \mathbf{M}(G, \phi) \boldsymbol{\iota}.$$

For a k-regular graph G_k, the walk generating function is equal to

$$N_{G_k}(\phi) = \frac{n}{1 - k\phi}.$$

For example, the cycle C_n on n nodes (see Figure B.1, left panel) is a 2-regular graph and its walk generating function is given by $N_{C_n}(\phi) = \frac{1}{1-2\phi}$. As another example, consider the star $K_{1,n-1}$ with n nodes (see Figure B.1, middle panel). Then the walk generating function is given by

$$N_{K_{1,n-1}}(\phi) = \frac{n + 2(n-1)\phi}{1 - (n-1)\phi^2}.$$

In general, it holds that $N_G(0) = n$, and one can show that $N_G(\phi) \geq 0$. We further have that

$$\mathbf{M}(G, \phi) = (\mathbf{I} - \phi \mathbf{A})^{-1} = \sum_{k=0}^{\infty} \phi^k \mathbf{A}^k = \sum_{k=0}^{\infty} \phi^k \mathbf{S} \boldsymbol{\Lambda}^k \mathbf{S}^\top,$$

where $\boldsymbol{\Lambda} \equiv \text{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix containing the eigenvalues of the real, symmetric matrix $\mathbf{A}$, and $\mathbf{S}$ is an orthogonal matrix with columns given by the orthogonal eigenvectors of $\mathbf{A}$ (with $\mathbf{S}^\top = \mathbf{S}^{-1}$), and we have used the fact that $\mathbf{A} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^\top$ [Horn and Johnson, 1990]. The eigenvectors $\mathbf{v}_i$ have the property that $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ and are normalized such that $\mathbf{v}_i^\top \mathbf{v}_i = 1$. Note that $\mathbf{A} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^\top$ is equivalent to $\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^\top$. It then follows that

$$\boldsymbol{\iota}^\top \mathbf{M}(G, \phi) \boldsymbol{\iota} = \boldsymbol{\iota}^\top \mathbf{S} \sum_{k=0}^{\infty} \phi^k \boldsymbol{\Lambda}^k \mathbf{S}^\top \boldsymbol{\iota},$$

where

$$\mathbf{S}^\top \boldsymbol{\iota} = \left(\boldsymbol{\iota}^\top \mathbf{v}_1, \ldots, \boldsymbol{\iota}^\top \mathbf{v}_n \right)^\top,$$

and

$$\boldsymbol{\Lambda}^k = \begin{pmatrix} \lambda_1^k & 0 & \ldots & 0 \\ 0 & \lambda_2^k & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & & \lambda_n^k \end{pmatrix} = \lambda_1^k \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \left(\frac{\lambda_2}{\lambda_1}\right)^k & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & & \left(\frac{\lambda_n}{\lambda_1}\right)^k \end{pmatrix}.$$

We then can write

$$\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} = \sum_{k=0}^{\infty} \phi^k \lambda_1^k \left(\boldsymbol{\iota}^\top \mathbf{v}_1, \ldots, \boldsymbol{\iota}^\top \mathbf{v}_n\right) \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \left(\frac{\lambda_2}{\lambda_1}\right)^k & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & & \left(\frac{\lambda_n}{\lambda_1}\right)^k \end{pmatrix} \left(\boldsymbol{\iota}^\top \mathbf{v}_1, \ldots, \boldsymbol{\iota}^\top \mathbf{v}_n\right)^\top,$$

which gives

$$\begin{aligned} \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} &= \sum_{k=0}^{\infty} \phi^k \lambda_1^k \left((\boldsymbol{\iota}^\top \mathbf{v}_1)^2 + \left(\frac{\lambda_2}{\lambda_1}\right)^k (\boldsymbol{\iota}^\top \mathbf{v}_2)^2 + \ldots + \left(\frac{\lambda_n}{\lambda_1}\right)^k (\boldsymbol{\iota}^\top \mathbf{v}_n)^2 \right) \\ &= \sum_{i=1}^{n} (\boldsymbol{\iota}^\top \mathbf{v}_i)^2 \sum_{k=0}^{\infty} \phi^k \lambda_i^k \\ &= \sum_{i=1}^{n} \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{1-\phi\lambda_i}. \end{aligned}$$

The above computation also shows that

$$N_k = \boldsymbol{\iota}^\top \mathbf{A}^k \boldsymbol{\iota} = \sum_{i=1}^{n} (\boldsymbol{\iota}^\top \mathbf{v}_i)^2 \lambda_i^k.$$

Hence, we can write the walk generating function as follows

$$N_G(\phi) = \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} = \sum_{k=0}^{\infty} N_k \phi^k = \sum_{i=1}^{n} \frac{(\mathbf{v}_i^\top \mathbf{u})^2}{1-\lambda_i\phi}.$$

If λ_1 is much larger than λ_j for all $j \geq 2$, then we can approximate

$$N_G(\phi) \approx (\boldsymbol{\iota}^\top \mathbf{v}_1)^2 \sum_{k=0}^{\infty} \phi^k \lambda_1^k = \frac{(\boldsymbol{\iota}^\top \mathbf{v}_1)^2}{1-\phi\lambda_1}.$$

Moreover, there exists the following relationship between the largest eigenvalue $\lambda_{\max}$ of the adjacency matrix and the number of walks of length k in G [cf. Van Mieghem, 2011, p. 47]

$$\lambda_{\max}(G) \geq \left(\frac{N_k(G)}{n}\right)^{\frac{1}{k}},$$

and, in particular,

$$\lim_{k\to\infty} \left(\frac{N_k(G)}{n}\right)^{\frac{1}{k}} = \lambda_{\max}(G).$$

Hence, we have that $n\lambda_{\max}(G)^k \geq N_k(G)$, and

$$N_G(\phi) = \sum_{k=0}^{\infty} N_k \phi^k \leq n \sum_{k=0}^{\infty} (\lambda_{\max}(G)\phi)^k = \frac{n}{1-\phi\lambda_{\max}(G)}. \tag{B.11}$$

To derive a lower bound, note that for $\phi \geq 0$, $N_G(\phi)$ is increasing in ϕ, so that $N_G(\phi) \geq N_0 + \phi N_1 + \phi^2 N_2$. Using the fact that $N_0 = n$, $N_1 = 2m = n\bar{d}$ and $N_2 = \sum_{i=1}^{n} d_i^2 = n(\bar{d}^2 + \sigma_d^2)$, we then get the lower bound

$$N_G(\phi) \geq n + 2m\phi + n(\bar{d}^2 + \sigma_d^2)\phi^2. \tag{B.12}$$

Finally, Cvetkovic et al. [1995, p. 45] have found an alternative expression for the walk generating function given by

$$N_G(\phi) = \frac{1}{\phi}\left((-1)^n \frac{c_{\mathbf{A}^c}\left(-\frac{1}{\phi}-1\right)}{c_{\mathbf{A}}\left(\frac{1}{\phi}\right)} - 1\right),$$

where $c_{\mathbf{A}}(\phi) \equiv \det\left(\mathbf{A} - \phi\mathbf{I}_n\right)$ is the characteristic polynomial of the matrix $\mathbf{A}$, whose roots are the eigenvalues of $\mathbf{A}$. It can be written as $c_{\mathbf{A}}(\phi) = \phi^n - a_1\phi^{n-1} + \ldots + (-1)^n a_n$, where $a_1 = \mathrm{tr}(\mathbf{A})$ and $a_n = \det(\mathbf{A})$. Furthermore, $\mathbf{A}^c = \boldsymbol{\iota}\boldsymbol{\iota}^\top - \mathbf{I} - \mathbf{A}$ is the complement of $\mathbf{A}$, and $\boldsymbol{\iota}\boldsymbol{\iota}^\top$ is an $n \times n$ matrix of ones. This is a convenient expression for the walk generating function, as there exist fast algorithms to compute the characteristic polynomial [Samuelson, 1942].

B.3. Bonacich Centrality

In the following we introduce a network measure capturing the centrality of a firm in the network due to Katz [1953] and later extended by Bonacich [1987]. Let $\mathbf{A}$ be the symmetric $n \times n$ *adjacency matrix* of the network G and λ_{PF} its largest real eigenvalue. The matrix $\mathbf{M}(G,\phi) = (\mathbf{I} - \phi\mathbf{A})^{-1}$ exists and is non-negative if and only if $\phi < 1/\lambda_{\mathrm{PF}}$.[4] Then

$$\mathbf{M}(G,\phi) = \sum_{k=0}^{\infty} \phi^k \mathbf{A}^k. \tag{B.13}$$

The Bonacich centrality vector is given by

$$\mathbf{b}_{\boldsymbol{\iota}}(G,\phi) = \mathbf{M}(G,\phi) \cdot \boldsymbol{\iota}, \tag{B.14}$$

where $\boldsymbol{\iota} = (1,\ldots,1)^\top$. We can write the Bonacich centrality vector as

$$\mathbf{b}_{\boldsymbol{\iota}}(G,\phi) = \sum_{k=0}^{\infty} \phi^k \mathbf{A}^k \cdot \boldsymbol{\iota} = (\mathbf{I} - \phi\mathbf{A})^{-1} \cdot \boldsymbol{\iota}.$$

For the components $b_{\boldsymbol{\iota},i}(G,\phi)$, $i = 1,\ldots,n$, we get

$$b_{\boldsymbol{\iota},i}(G,\phi) = \sum_{k=0}^{\infty} \phi^k (\mathbf{A}^k \cdot \boldsymbol{\iota})_i = \sum_{k=0}^{\infty} \phi^k \sum_{j=1}^{n} \left(\mathbf{A}^k\right)_{ij}. \tag{B.15}$$

The sum of the Bonacich centralities is then exactly the walk generating function we have introduced in Section B.2

$$\sum_{i=1}^{n} b_{\boldsymbol{\iota},i}(G,\phi) = \boldsymbol{\iota}^\top \mathbf{b}_{\mathbf{u}}(G,\phi) = \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} = N_G(\phi).$$

Moreover, because $\sum_{j=1}^{n}\left(\mathbf{A}^k\right)_{ij}$ counts the number of all walks of length k in G starting from i, $b_{\mathbf{u},i}(G,\phi)$ is the number of all walks in G starting from i, where the walks of length k

[4] The proof can be found e.g. in Debreu and Herstein [1953].

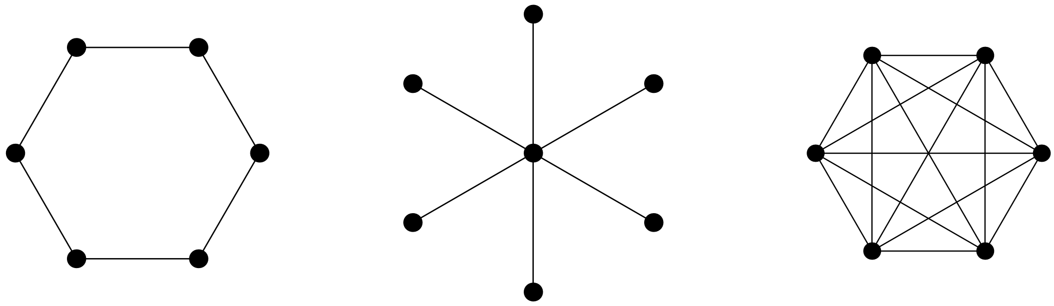


Figure B.1: Illustration of a cycle C_6, a star $K_{1,6}$ and a complete graph, K_6.

are weighted by their geometrically decaying factor ϕ^k. In particular, we can decompose the Bonacich centrality as follows

$$b_i(G,\rho) = \underbrace{b_{ii}(G,\phi)}_{\text{closed walks}} + \underbrace{\sum_{j\neq i} b_{ij}(G,\phi)}_{\text{out-walks}}, \tag{B.16}$$

where $b_{ii}(G,\phi)$ counts all closed walks from firm i to i and $\sum_{j\neq i} b_{ij}(G,\phi)$ counts all the other walks from i to every other firm $j \neq i$. Similarly, Ballester et al. [2006] define the *intercentrality* of firm $i \in \mathcal{N}$ as

$$c_i(G,\phi) = \frac{b_i(G,\phi)^2}{b_{ii}(G,\phi)}, \tag{B.17}$$

where the factor $b_{ii}(G,\phi)$ measures all closed walks starting and ending at firm i, discounted by the factor ϕ, whereas $b_i(G,\phi)$ measures the number of walks emanating at firm i, discounted by the factor ϕ. The intercentrality index hence expresses the ratio of the (square of the) number of walks leaving a firm i relative to the number of walks returning to i.

We give two examples in the following to illustrate the Bonacich centrality. The graphs used in these examples are depicted in Figure B.1. First, consider the star $K_{1,n-1}$ with n nodes (see Figure B.1, middle panel) and assume w.l.o.g. that 1 is the index of the central node with maximum degree. We now compute the Bonacich centrality for the star $K_{1,n-1}$. We have that

$$\begin{aligned}
\mathbf{M}(K_{1,n-1},\phi) = (\mathbf{I}-\phi\mathbf{A})^{-1} &= \begin{pmatrix} 1 & -\phi & \cdots & \cdots & -\phi \\ -\phi & 1 & 0 & & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ & & \ddots & & \vdots \\ \vdots & \vdots & & & 0 \\ -\phi & 0 & \cdots & 0 & 1 \end{pmatrix}^{-1} \\
&= \frac{1}{1-(n-1)\phi^2} \begin{pmatrix} 1 & \phi & \cdots & \cdots & \phi \\ \phi & 1-(n-2)\phi^2 & \phi^2 & & \phi^2 \\ \vdots & \phi^2 & \ddots & \ddots & \vdots \\ & & \ddots & & \vdots \\ \vdots & \vdots & & & \phi^2 \\ \phi & \phi^2 & \cdots & \phi^2 & 1-(n-2)\phi^2 \end{pmatrix}.
\end{aligned}$$

Since $\mathbf{b} = \mathbf{M}\cdot\boldsymbol{\iota}$ we then get

$$\mathbf{b}(K_{1,n-1},\phi) = \frac{1}{1-(n-1)\phi^2}\left(1+(n-1)\phi, 1+\phi, \ldots, 1+\phi\right)^\top. \tag{B.18}$$

Next, consider the complete graph K_n with n nodes (see Figure B.1, right panel). We have

$$\mathbf{M}(K_n,\phi) = (\mathbf{I}-\phi\mathbf{A})^{-1} = \begin{pmatrix} 1 & -\phi & \cdots & \cdots & -\phi \\ -\phi & 1 & -\phi & & -\phi \\ \vdots & -\phi & \ddots & \ddots & \vdots \\ & & \ddots & & \vdots \\ \vdots & \vdots & & & -\phi \\ -\phi & -\phi & \cdots & -\phi & 1 \end{pmatrix}^{-1}$$

$$= \frac{1}{1-(n-2)\phi-(n-1)\phi^2} \begin{pmatrix} 1-(n-2)\phi & \phi & \cdots & \cdots & \phi \\ \phi & 1-(n-2)\phi & \phi & & \phi \\ \vdots & \phi & \ddots & \ddots & \vdots \\ & & \ddots & & \vdots \\ \vdots & \vdots & & & \phi \\ \phi & \phi & \cdots & \phi & 1-(n-2)\phi \end{pmatrix}.$$

With $\mathbf{b} = \mathbf{M}\cdot\boldsymbol{\iota}$ we then have that

$$\mathbf{b}(K_n,\phi) = \frac{1}{1-(n-1)\phi}(1,\ldots,1)^\top . \tag{B.19}$$

The Bonacich matrix of Equation (B.13) is also a measure of structural similarity of the firms in the network, called *regular equivalence*. Leicht et al. [2006] define a similarity score b_{ij}, which is high if nodes i and j have neighbors that themselves have high similarity, given by $b_{ij} = \phi\sum_{k=1}^n a_{ik}b_{kj} + \delta_{ij}$. In matrix-vector notation this reads $\mathbf{M} = \phi\mathbf{A}\mathbf{M} + \mathbf{I}$. Rearranging yields $\mathbf{M} = (\mathbf{I}-\phi\mathbf{A})^{-1} = \sum_{k=0}^{\infty}\phi^k\mathbf{A}^k$, assuming that $\phi < 1/\lambda_{\mathrm{PF}}$. We hence obtain that the similarity matrix $\mathbf{M}$ is equivalent to the Bonacich matrix from Equation (B.13). The average similarity of firm i is $\frac{1}{n}\sum_{j=1}^n b_{ij} = \frac{1}{n}b_{\boldsymbol{\iota},i}(G,\phi)$, where $b_{\boldsymbol{\iota},i}(G,\phi)$ is the Bonacich centrality of i. It follows that the Bonacich centrality of i is proportional to the average regular equivalence of i. Firms with a high Bonacich centrality are then the ones which also have a high average structural similarity with the other firms in the R&D network.

The interpretation of eingenvector-like centrality measures as a similarity index is also important in the study of correlations between observations in principal component analysis and factor analysis [cf. Rencher and Christensen, 2012]. Variables with similar factor loadings can be grouped together. This basic idea has also been used in the economics literature on segregation [e.g. Ballester and Vorsatz, 2013].

There also exists a connection between the Bonacich centrality of a node and its coreness in the network (see Appendix B.1). The following result, due to Manshadi and Johari [2010], relates the Nash equilibrium to the k-cores of the graph: If $\mathrm{cor}_i = k$ then $b_i(G,\phi) \geq \frac{1}{1-\phi k}$, where the inequality is tight when i belongs to a disconnected clique of size $k+1$. The coreness of networks of R&D collaborating firms has also been studied empirically in Kitsak et al. [2010] and Rosenkopf and Schilling [2007]. In particular, Kitsak et al. [2010] find that the coreness of a firm correlates with its market value. We can easily explain this from our model because we know that firms in higher cores tend to have higher Bonacich centrality, and therefore higher sales and profits (cf. Proposition 1).

C. Games on Networks: The contribution of our model

In this section, we show how our model embeds standard models of games on networks. Our profit function is given by Equation (4), that is

$$\pi_i = \mu_i\, q_i - q_i^2 - \rho \sum_{j=1}^{n} b_{ij} q_i q_j + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2} e_i^2,$$

where $\mu_i = \overline{\alpha}_i - \overline{c}_i$.

C.1. A Model without Network Effects

Let us consider a model with the product market alone, i.e. $\varphi = 0$. In that case, the profit function in Equation (4) of firm i reduces to

$$\pi_i = \mu_i\, q_i - q_i^2 - \rho \sum_{j=1}^{n} b_{ij} q_i q_j + q_i e_i - \frac{1}{2} e_i^2. \tag{C.20}$$

This is, for example, a model that is commonly used in the industrial organization literature to study product differentiation [cf. Singh and Vives, 1984]. In that case, the first-order condition with respect to e_i leads to $e_i = q_i$, while the first-order condition with respect to q_i can be written as:

$$q_i = \mu_i - \rho \sum_{j=1}^{n} b_{ij} q_j.$$

Let $\boldsymbol{\mu}$ be the $n \times 1$ vector of μ_i's.

Lemma 1. *Consider the profit function in Equation (C.20). If* $\left(\frac{\overline{\mu}}{\underline{\mu}} - 1\right) < \frac{1-\rho}{n\rho}$ *then there exists a unique interior Nash equilibrium, which is given by*

$$\mathbf{q} = \left(\mathbf{I} + \rho \mathbf{B}\right)^{-1} \boldsymbol{\mu}.$$

Proof of Lemma 1 First, the condition for existence and uniqueness of the Nash equilibrium is that the matrix $\mathbf{I} + \rho\mathbf{B}$ has to be positive definite. A sufficient condition is that all eigenvalues of this matrix are positive, which is guaranteed by $\lambda_{\min}(\mathbf{B}) > -1/\rho$. Since $\lambda_{\min}(\mathbf{B}) = -1$, this is equivalent to $\rho < 1$, which is always true by assumption. Second, Equation (9) in part (ii) of Proposition 1 requires that the inequality $\frac{n\rho}{1-\rho}\left(\frac{\overline{\mu}}{\underline{\mu}} - 1\right) < 1$ is satisfied for an interior solution to exist. □

We can see that this is a special case of our Proposition 1, when $\varphi = 0$.

C.2. A Model without Competition Effects

Let us now consider a model with no competition effect so that $\rho = 0$. In that case, the profit function in Equation (4) of firm i reduces to:

$$\pi_i = \mu_i\, q_i - q_i^2 + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2} e_i^2.$$

The first-order with respect to e_i leads to: $e_i = q_i$ while that with respect to q_i is given by:

$$\mu_i - 2q_i + e_i + \varphi \sum_{j=1}^{n} a_{ij} e_j = 0.$$

Using the fact that $e_i = q_i$, we easily obtain:

$$q_i = \mu_i + \varphi \sum_{j=1}^{n} a_{ij} q_j.$$

If $\varphi \lambda_{\max}(\mathbf{A}) < 1$, there exists a unique Nash equilibrium given by

$$\mathbf{q}^* = \mathbf{b}_{\boldsymbol{\mu}}(G, \varphi) \equiv (\mathbf{I} - \varphi \mathbf{A})^{-1} \boldsymbol{\mu},$$

where $\mathbf{b}_{\boldsymbol{\mu}}(G, \varphi)$ is the $\boldsymbol{\mu}$-weighted Katz-Bonacich centrality. This is part (iii) of our Proposition 1.

C.3. Comparison of our model with Ballester et al. [2006] and Bramoullé et al. [2014]

Ballester et al. [2006] (BCZ) consider a single market (i.e., $M = 1$) without R&D investment decisions. They also assume that firms are ex ante homogenous with $\mu_i = \mu$. The equilibrium best response function in their case is given by

$$q_i = \mu - \rho \sum_{j=1, j \neq i}^{n} q_j + \varphi \sum_{j=1}^{n} a_{ij} q_j.$$

This is a special case of part (ii) of our Proposition 1 when $\mu_i = \mu$.

Bramoullé et al. [2014] generalize Ballester et al. [2006] by allowing for ex ante heterogeneity. [5] However, they still assume a single market (i.e., $M = 1$), and abstract away from R&D investment decisions. Their equilibrium best response function is

$$q_i = \mu_i - \rho \sum_{j=1, j \neq i}^{n} q_j + \varphi \sum_{j=1}^{n} a_{ij} q_j.$$

In that case, their main result (their Proposition 3) corresponds to part (ii) of our Proposition 1.[6]

D. Herfindahl Index and Market Concentration

The Herfindahl-Hirschman industry concentration index is defined as $H = \sum_{i=1}^{n} s_i^2$, where the market share of firm i is given by $s_i = \frac{q_i}{\sum_{j=1}^{n} q_j}$ [cf. e.g. Hirschman, 1964; Tirole, 1988]. Hence, we can write

$$H = \sum_{i=1}^{n} \left(\frac{q_i}{\sum_{j=1}^{n} q_j} \right)^2 = \frac{\|\mathbf{q}\|_2^2}{\|\mathbf{q}\|_1^2}, \tag{D.21}$$

[5] See also Calvó-Armengol et al. [2009].

[6] The condition for existence and uniqueness of equilibrium in Bramoullé et al. [2014] is slightly different since it involves $\lambda_{\min}(\mathbf{A})$, the lowest eigenvalue of $\mathbf{A}$, rather than $\lambda_{\max}(\mathbf{A})$, the largest eigenvalue of $\mathbf{A}$. Observe that, in our paper, it can be seen from the proof of Proposition 1 that we have another condition for the existence and uniqueness of equilibrium, which is given by: $\lambda_{\min}(\rho \mathbf{B} - \varphi \mathbf{A}) + 1 > 0$, which is similar to that of Bramoullé et al. [2014]. We then write an equivalent condition in terms of $\lambda_{\max}(\mathbf{A})$. Also, in most of their paper, Bramoullé et al. [2014] assume that $\rho = 0$ so that they do not have to worry about the interiority of the solution.

With $\mathbf{q} = \mathbf{b}_{\boldsymbol{\iota}}(G, \phi) = \mathbf{M}(G, \phi)\boldsymbol{\iota}$ in the Nash equilibrium (see Proposition 1), we can write the Herfindahl index of Equation (D.21) as follows

$$H(G) = \frac{\boldsymbol{\iota}^\top \mathbf{M}(G, \phi)^2 \boldsymbol{\iota}}{(\boldsymbol{\iota}^\top \mathbf{M}(G, \phi)\boldsymbol{\iota})^2} = \frac{\|\mathbf{b}\|_2^2}{\|\mathbf{b}\|_1^2} = \frac{\sum_{i=1}^n b_i^2}{\left(\sum_{i=1}^n |b_i|\right)^2} = \gamma(\mathbf{b})^{-1},$$

which is the inverse of the *participation ratio* $\gamma(\cdot)$. The participation ratio $\gamma(\mathbf{x})$ measures the number of elements of $\mathbf{x}$ which are dominant. We have that $1 \leq \gamma(\mathbf{x}) \leq n$, where a value of $\gamma(\mathbf{x}) = n$ corresponds to a fully homogenous case, while $\gamma(\mathbf{x}) = 1$ corresponds to a fully concentrated case (note that, if all x_i are identical then $\gamma(\mathbf{x}) = n$, while if one x_i is much larger than all others we have $\gamma(\mathbf{x}) = 1$). Moreover, $\gamma(\mathbf{x})$ is scale invariant, that is, $\gamma(\alpha\mathbf{x}) = \gamma(\mathbf{x})$ for any $\alpha \in \mathbb{R}_+$. The participation ratio $\gamma(\mathbf{x})$ is further related to the *coefficient of variation* $c_v(\mathbf{x}) = \frac{\sigma(\mathbf{x})}{\mu(\mathbf{x})}$, where $\sigma(\mathbf{x})$ is the standard deviation and $\mu(\mathbf{x})$ the mean of the components of $\mathbf{x}$, via the relationship $c_v(\mathbf{x})^2 = \frac{n}{\gamma(\mathbf{x})} - 1$. This implies that

$$H(G) = \frac{\boldsymbol{\iota}^\top \mathbf{M}(G, \phi)^2 \boldsymbol{\iota}}{(\boldsymbol{\iota}^\top \mathbf{M}(G, \phi)\boldsymbol{\iota})^2} = \frac{c_v(\mathbf{b})^2 + 1}{n} \sim \frac{c_v(\mathbf{b})^2}{n}.$$

Hence, the Herfindhal index is maximized for the graph G with the highest coefficient of variation in the components of the Bonacich centrality $\mathbf{b}_{\boldsymbol{\iota}}(G, \phi)$. Finally, as for small values of ϕ the Bonacich centrality becomes proportional to the degree, the variance of the Bonacich centrality will be determined by the variance of the degree. It is known that the graphs that maximize the degree variance are nested split graphs [cf. Peled et al., 1999].

E. Bertrand Competition

In the case of price setting firms we obtain from the profit function in Equation (3) the FOC with respect to price p_i for firm i

$$\frac{\partial \pi_i}{\partial p_i} = (p_i - c_i)\frac{\partial q_i}{\partial p_i} - q_i = 0.$$

When $i \in \mathcal{M}_m$, then observe that from the inverse demand in Equation (1) we find that

$$q_i = \frac{\alpha_m(1 - \rho_m) - (1 - (n_m - 2)\rho_m)p_i + \rho_m \sum_{j \in \mathcal{M}_m, j \neq i} p_j}{(1 - \rho)(1 + (n_m - 1)\rho_m)},$$

where $n_m \equiv |\mathcal{M}_m|$. It then follows that

$$\frac{\partial q_i}{\partial p_i} = -\frac{1 - (n_m - 2)\rho_m}{(1 - \rho_m)(1 + (n_m - 1)\rho_m)}.$$

Inserting into the FOC with respect to p_i gives

$$q_i = -\frac{1 - (n_m - 2)\rho_m}{(1 - \rho_m)(1 + (n_m - 1)\rho_m)}(p_i - c_i).$$

Inserting Equations (1) and (2) yields

$$
\begin{aligned}
q_i = & \frac{(1-(n_m-2)\rho_m)(\alpha_m-\bar{c}_i)}{\rho_m(4-(2-\rho_m)n_m-\rho_m)} - \frac{1-(n_m-2)\rho_m}{4-(2-\rho_m)n_m-\rho_m} \sum_{j\in\mathcal{M}_m, j\neq i} q_j \\
& + \frac{(1-(n_m-2)\rho_m)}{\rho_m(4-(2-\rho_m)n_m-\rho_m} e_i + \frac{(1-(n_m-2)\rho_m)\varphi}{\rho_m(4-(2-\rho_m)n_m-\rho_m} \sum_{j=1}^n a_{ij} e_j.
\end{aligned}
$$

The FOC with respect to R&D effort is the same as in the case of perfect competition, so that we get $e_i = q_i$. Inserting equilibrium effort and rearranging terms gives

$$
\begin{aligned}
q_i = & \frac{(1-(n_m-2)\rho_m)(\alpha_m-\bar{c}_i)}{\rho_m(4-(2-\rho_m)n_m-\rho_m) - 1(1-(n_m-2)\rho_m)} \\
& - \frac{\rho_m(1-(n_m-2)\rho_m)}{\rho_m(4-(2-\rho_m)n_m-\rho_m) - 1(1-(n_m-2)\rho_m)} \sum_{j\in\mathcal{M}_m, j\neq i} q_j \\
& + \frac{\varphi(1-(n_m-2)\rho_m)}{\rho_m(4-(2-\rho_m)n_m-\rho_m) - 1(1-(n_m-2)\rho_m)} \sum_{j=1}^n a_{ij} q_j.
\end{aligned}
$$

If we denote by

$$
\begin{aligned}
\mu_i & \equiv \frac{(1-(n_m-2)\rho_m)(\alpha_m-\bar{c}_i)}{\rho_m(4-(2-\rho_m)n_m-\rho_m) - 1(1-(n_m-2)\rho_m)}, \\
\rho & \equiv \frac{\rho_m(1-(n_m-2)\rho_m)}{\rho_m(4-(2-\rho_m)n_m-\rho_m) - 1(1-(n_m-2)\rho_m)}, \\
\lambda & \equiv \frac{\varphi(1-(n_m-2)\rho_m)}{\rho_m(4-(2-\rho_m)n_m-\rho_m) - 1(1-(n_m-2)\rho_m)}.
\end{aligned}
$$

Then we can write equilibrium quantities as follows

$$
q_i = \mu_i - \rho \sum_{j=1}^n b_{ij} q_j + \lambda \sum_{j=1}^n a_{ij} q_j. \tag{E.22}
$$

Observe that the reduced form Equation (E.22) is identical to the Cournot case in Equation (10).

F. Equilibrium Characterization with Direct and Indirect Technology Spillovers

We extend our model by allowing for direct (between collaborating firms) and indirect (between non-collaborating firms) technology spillovers. The profit of firm $i \in \mathcal{N}$ is still given by $\pi_i = (p_i - c_i)q_i - \frac{1}{2}e_i^2$, where the inverse demand is $p_i = \bar{\alpha}_i - q_i - \rho\sum_{j=1}^n b_{ij} q_j$. The main change is in the marginal cost of production, which is now equal to[7]

$$
c_i = \bar{c}_i - e_i - \varphi \sum_{j=1}^n a_{ij} e_j - \chi \sum_{j=1}^n w_{ij} e_j, \tag{F.23}
$$

[7] See also Eq. (1) in Goyal and Moraga-Gonzalez [2001].

where w_{ij} are weights characterizing alternative channels for technology spillovers than R&D collaborations (representing for example a patent cross-citation, a flow of workers, or technological proximity measured by the matrix P_{ij} introduced in Footnote 28). Inserting this marginal cost of production into the profit function gives

$$\pi_i = (\bar{\alpha}_i - \bar{c}_i)q_i - q_i^2 - \rho q_i \sum_{j=1}^n b_{ij} q_j + q_i e_i + \varphi q_i \sum_{j=1}^n a_{ij} e_j + \chi q_i \sum_{j=1}^n w_{ij} e_j - \frac{1}{2} e_i^2.$$

As above, from the first-order condition with respect to R&D effort, we obtain $e_i = q_i$. Inserting this optimal effort into the first-order condition with respect to output, we obtain

$$q_i = \bar{\alpha}_i - \bar{c}_i - \rho \sum_{j=1}^n b_{ij} q_j + \varphi \sum_{j=1}^n a_{ij} q_j + \chi \sum_{j=1}^n w_{ij} q_j.$$

Denoting by $\mu_i \equiv \bar{\alpha}_i - \bar{c}_i$, we can write this as

$$q_i = \mu_i - \rho \sum_{j=1}^n b_{ij} q_j + \varphi \sum_{j=1}^n a_{ij} q_j + \chi \sum_{j=1}^n w_{ij} q_j. \tag{F.24}$$

If the matrix $\mathbf{I} + \rho\mathbf{B} - \varphi\mathbf{A} - \chi\mathbf{W}$ is invertible, this gives us the equilibrium quantities

$$\mathbf{q} = (\mathbf{I} + \rho\mathbf{B} - \varphi\mathbf{A} - \chi\mathbf{W})^{-1} \boldsymbol{\mu}.$$

Let us now write the econometric equivalent of Equation (F.24). Proceeding as in Section 6.1, using Equations (23) and (24) and introducing time t, we get

$$\mu_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \eta_i + \kappa_t + \epsilon_{it}.$$

Plugging this value of μ_{it} into Equation (F.24), we obtain

$$q_{it} = \varphi \sum_{j=1}^n a_{ij,t} q_{jt} + \chi \sum_{j=1}^n w_{ij,t} q_{jt} - \rho \sum_{j=1}^n b_{ij} q_{jt} + \mathbf{x}_{it}^\top \beta + \eta_i + \kappa_t + \epsilon_{it}.$$

This is Equation (30) in Section 6.4.

G. Additional Results on Welfare and Efficiency

In the following sections we illustrate how the private returns from R&D can be lower than the social returns (Appendix G.1), and we show which network structures are efficient (Appendix G.2).

G.1. Private vs. Social Returns to R&D

The aim of this section is to show that the choice of q_i by each firm i at the Nash equilibrium is not efficient so that the *private* returns of R&D effort and output are different from the *social* returns of R&D effort and output.

Let us first calculate the Nash equilibrium as in the main text in Section 3. The profit function is given by Equation (4), that is

$$\pi_i = \mu_i \, q_i - q_i^2 - \rho \sum_{j=1}^n b_{ij} q_i q_j + q_i e_i + \varphi q_i \sum_{j=1}^n a_{ij} e_j - \frac{1}{2} e_i^2, \tag{G.25}$$

where $\mu_i := \overline{\alpha}_i - \overline{c}_i$. The first-order condition with respect to e_i yields $q_i = e_i$, so that the first-order condition with respect to q_i leads to:

$$q_i = \mu_i - \rho \sum_{j=1}^{n} b_{ij} q_j + \varphi \sum_{j=1}^{n} a_{ij} q_j. \tag{G.26}$$

In part (i) and (ii) of Proposition 1, we showed that if Equations (5) and (9) hold, then there exists a unique interior Nash equilibrium, which is given by Equation (G.26). Under these conditions we can write the output levels as

$$\mathbf{q}^{NE} = (\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A})^{-1} \boldsymbol{\mu}, \tag{G.27}$$

where the superscript NE refers to the "Nash equilibrium ". Let us now show that the Nash equilibrium defined by Equation (G.27) is not efficient. For this purpose we consider a planner who chooses both R&D efforts, $\mathbf{e} \in \mathbb{R}_+^n$, and output levels, $\mathbf{q} \in \mathbb{R}_+^n$, in order to maximize welfare W, defined as the sum of producer and consumer surplus, U and Π, respectively. Consumer surplus is given by $U = \frac{1}{2} \sum_{i=1}^n q_i^2 + \frac{\rho}{2} \sum_{i=1}^n \sum_{j=1}^n b_{ij} q_i q_j$ while producer surplus is defined as the sum of firms' profits, $\Pi = \sum_{i=1}^n \pi_i$, with π_i given by Equation (G.25). That is, the planner solves the following program:[8]

$$\begin{aligned}
\max_{\mathbf{e},\mathbf{q} \in \mathbb{R}_+^n} W &= \max_{\mathbf{e},\mathbf{q} \in \mathbb{R}_+^n} (U + \Pi) \\
&= \max_{\mathbf{e},\mathbf{q} \in \mathbb{R}_+^n} \sum_{i=1}^{n} \left(\frac{1}{2} q_i^2 + \frac{\rho}{2} \sum_{j=1}^{n} b_{ij} q_i q_j + \mu_i \, q_i - q_i^2 - \rho \sum_{j=1}^{n} b_{ij} q_i q_j + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2} e_i^2 \right) \\
&= \max_{\mathbf{e},\mathbf{q} \in \mathbb{R}_+^n} \sum_{i=1}^{n} \left(\mu_i \, q_i - \frac{1}{2} q_i^2 - \frac{\rho}{2} \sum_{j=1}^{n} b_{ij} q_i q_j + q_i e_i + \varphi q_i \sum_{j=1}^{n} a_{ij} e_j - \frac{1}{2} e_i^2 \right) \\
&= \max_{\mathbf{e},\mathbf{q} \in \mathbb{R}_+^n} \left[\sum_{i=1}^{n} \left(\mu_i \, q_i - \frac{1}{2} q_i^2 + q_i e_i - \frac{1}{2} e_i^2 \right) - \frac{\rho}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} q_i q_j + \varphi \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} q_i e_j \right].
\end{aligned}$$

From the first-order condition with respect to R&D effort, e_i, given by

$$\frac{\partial W}{\partial e_i} = q_i - e_i + \varphi \sum_{j=1}^{n} a_{ij} q_j = 0,$$

we see that

$$e_i = q_i + \varphi \sum_{j=1}^{n} a_{ij} q_j. \tag{G.28}$$

Compared to the Nash equilibrium effort levels ($e_i = q_i$) we see that firms do not spend enough on R&D as compared to what is socially optimal. This is because they do not take into account the spillovers they generate on other connected firms (captured by the term $\varphi \sum_{j=1}^{n} a_{ij} q_j$ in Equation (G.28)). That is, there is a generic problem of *under-investment in R&D*, as the private returns from R&D are lower than the social returns from R&D. This motivates policies for fostering R&D investments as we have introduced them in Section 4 in the paper.

[8] We consider an interior solution such that the conditions in the proof of Proposition 1 are implicitly assumed to be satisfied.

Similarly, the first-order condition with respect to output is given by

$$\frac{\partial W}{\partial q_i} = \mu_i - q_i + e_i - \rho \sum_{j=1}^{n} b_{ij} q_j + 2\varphi \sum_{j=1}^{n} a_{ij} e_j = 0.$$

Inserting the socially optimal R&D effort levels from Equation (G.28) yields

$$\mu_i - q_i + q_i + \varphi \sum_{j=1}^{n} a_{ij} q_j - \rho \sum_{j=1}^{n} b_{ij} q_j + 2\varphi \sum_{j=1}^{n} a_{ij} \left(q_j + \varphi \sum_{k=1}^{n} a_{jk} q_k \right) = 0.$$

This can be written as follows

$$\mu_i + 3\varphi \sum_{j=1}^{n} a_{ij} q_j - \rho \sum_{j=1}^{n} b_{ij} q_j + 2\varphi^2 \sum_{j=1}^{n} a_{ij} \sum_{k=1}^{n} a_{jk} q_k = \mathbf{0}.$$

In vector-matrix notation this is

$$\boldsymbol{\mu} + 3\varphi \mathbf{A}\mathbf{q} - \rho \mathbf{B}\mathbf{q} + 2\varphi^2 \mathbf{A}^2 \mathbf{q} = \mathbf{0},$$

or equivalently

$$\boldsymbol{\mu} = \left(\rho \mathbf{B} - 3\varphi \mathbf{A} - 2\varphi^2 \mathbf{A}^2 \right) \mathbf{q} = \mathbf{0}.$$

When the matrix $\rho \mathbf{B} - 3\varphi \mathbf{A} - 2\varphi^2 \mathbf{A}^2$ is invertible, we get

$$\mathbf{q}^O = \left(\rho \mathbf{B} - 3\varphi \mathbf{A} - 2\varphi^2 \mathbf{A}^2 \right)^{-1} \boldsymbol{\mu}, \tag{G.29}$$

where the superscript O refers to the "social optimum". An examination of (G.27) and (G.29) shows that the two solutions differ and that the Nash equilibrium in such a game is *inefficient*, as there are negative and positive externalities in output (and R&D efforts) due to competition and spillover effects that are not internalized by the firms.

G.2. Efficient Network Structure

The aim of this section is to determine the optimal network structure, i.e. the network structure that maximizes total welfare. We will assume in the following that there is only a single market (with $M = 1$, $b_{ij} = 0$ for $i \neq j$ and $b_{ii} = 1$ for all $i, j \in \mathcal{N}$) and make the homogeneity assumption that $\mu_i = \mu$ for all $i \in \mathcal{N}$. Then, welfare can be written as follows

$$W(G) = \frac{2-\rho}{2} \|\mathbf{q}\|_2^2 + \frac{\rho}{2} \|\mathbf{q}\|_1^2,$$

where $\|\mathbf{q}\|_p \equiv (\sum_{i=1}^n q_i^p)^{\frac{1}{p}}$ is the L^p-norm of $\mathbf{q}$. Further, note that the Herfindahl-Hirschman industry concentration index is given by[9]

$$H = \sum_{i=1}^{n} \left(\frac{q_i}{\sum_{j=1}^{n} q_j} \right)^2 = \frac{\|\mathbf{q}\|_2^2}{\|\mathbf{q}\|_1^2},$$

[9]For more discussion of the Herfindahl index in the Nash equilibrium see Appendix D.

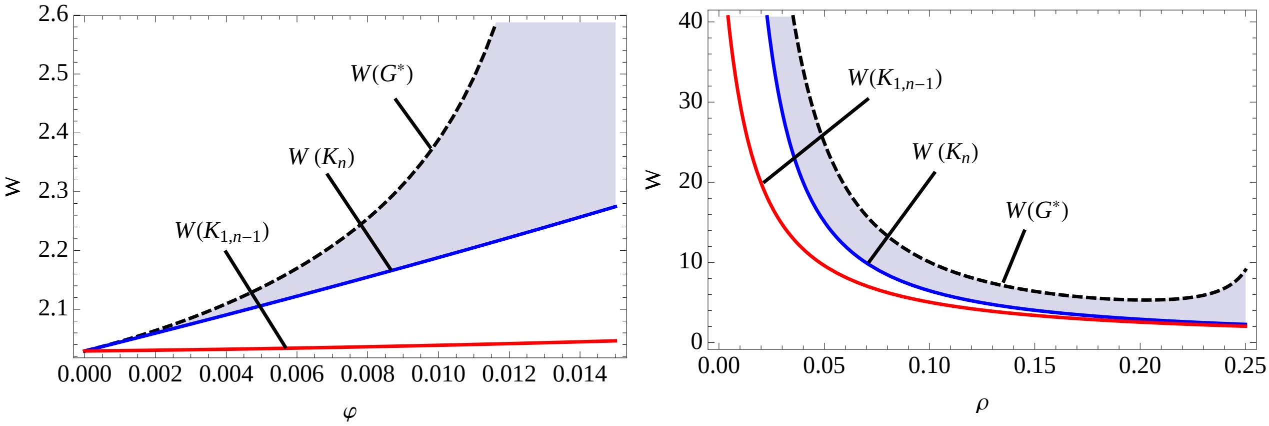


Figure G.2: (Left panel) The upper and lower bounds of Equation (G.31) with $n = 50$, $\rho = 0.25$ for varying values of φ. (Right panel) The upper and lower bounds of Equation (G.31) with $n = 50$, $\varphi = 0.015$ for varying values of ρ.

and denoting total output by $Q = \|\mathbf{q}\|_1$, we can write welfare as follows

$$W(G) = \frac{1}{2}\|\mathbf{q}\|_1^2 \left((2-\rho)\frac{\|\mathbf{q}\|_2^2}{\|\mathbf{q}\|_1^2} + \rho \right) = \frac{Q^2}{2}\left((2-\rho)H + \rho\right). \tag{G.30}$$

One can show that total output Q is largest in the complete graph [cf. Ballester et al., 2006]. However, as welfare depends on both, output Q and industry concentration H, it is not obvious that the complete graph (where $H = 1/n$ is small) is also maximizing welfare. As the following proposition illustrates, we can conclude that the complete graph is welfare maximizing (i.e. efficient) when externalities are weak, but this may no longer be the case when ρ or φ are high.

Proposition 4. *Assume that $\mu_i = \mu$ for all $i = 1, \dots, n$, and let ρ, μ, φ and ϕ satisfy the restrictions of Proposition 1. Denote by $\mathcal{G}^n$ the class of graphs with n nodes, $K_n \in \mathcal{G}^n$ the complete graph, $K_{1,n-1} \in \mathcal{G}^n$ the star network, and let the efficient graph be denoted by $G^* = \operatorname{argmax}_{G\in\mathcal{G}^n} W(G)$.*

(i) Welfare of the efficient graph G^* can be bounded from above and below as follows:

$$\frac{\mu^2 n(2+(n-1)\rho)}{2(1+(n-1)(\rho-\varphi))^2} \leq W(G^*) \leq \frac{\mu^2 n\left((1-\rho)^2(2+(n-1)\rho) - n(n-1)^2\rho\varphi^2\right)}{2((1+(n-1)(\rho-\varphi))^2\left((1-\rho)^2-(n-1)^2\varphi^2\right)}. \tag{G.31}$$

(ii) In the limit of independent markets, when $\rho \to 0$, the complete graph is efficient, $K_n = G^*$.

(iii) In the limit of weak R&D spillovers, when $\varphi \to 0$, the complete graph is efficient, $K_n = G^*$.

(iv) There exists a $\varphi^*(n,\rho) > 0$ (which is decreasing in ρ) such that $W(K_n) < W(K_{1,n-1})$ for all $\varphi > \varphi^*(n,\rho)$, and the complete graph is not efficient, $K_n \neq G^*$.

Proof of Proposition 4 (ii) Assuming that $\mu_i = \mu$ for all $i = 1, \dots, n$, at the Nash equilibrium, and that $\rho = 0$, we have that $\mathbf{q} = \mu\mathbf{M}(G,\varphi)\boldsymbol{\iota}$, where we have denoted by $\mathbf{M}(G,\varphi) \equiv (\mathbf{I} - \varphi\mathbf{A})^{-1}$.[10] We then obtain $W(G) = \mathbf{q}^\top\mathbf{q} = \mu^2\boldsymbol{\iota}^\top\mathbf{M}(G,\varphi)^2\boldsymbol{\iota}$. Observe that the quantity $\boldsymbol{\iota}^\top\mathbf{M}(G,\varphi)\boldsymbol{\iota}$ is the walk generating function, $N_G(\varphi)$, of G that we defined in detail in

[10] Note that there exists a relationship between the matrix $\mathbf{M}(G,\varphi)$ with elements $m_{ij}(G,\varphi)$ and the length of the shortest path $\ell_{ij}(G)$ between nodes i and j in the network G. Namely $\ell_{ij}(G) = \lim_{\varphi\to 0} \frac{\partial \ln m_{ij}(G,\varphi)}{\partial \ln \varphi} = \lim_{\varphi\to 0} \frac{\varphi}{m_{ij}(G,\varphi)}\frac{\partial m_{ij}(G,\varphi)}{\partial\varphi}$. See also Newman [2010, Chap. 6]. This means that the length of the shortest path between i and j is given by the relative percentage change in the weighted number of walks between nodes i and j in G with respect to a relative percentage change in φ in the limit of $\varphi \to 0$.

Appendix B.2. Using the results of Appendix B.2, we obtain

$$\begin{aligned}
\boldsymbol{\iota}^\top \mathbf{M}(G,\varphi)^2 \boldsymbol{\iota} &= \boldsymbol{\iota}^\top \left(\sum_{k=0}^{\infty} \varphi^k \mathbf{A}^k \right)^2 \boldsymbol{\iota} \\
&= \boldsymbol{\iota}^\top \left(\sum_{k=0}^{\infty} \sum_{l=0}^{k} \varphi^l \mathbf{A}^l \varphi^{k-l} \mathbf{A}^{k-l} \right) \boldsymbol{\iota} \\
&= \sum_{k=0}^{\infty} (k+1) \varphi^k \boldsymbol{\iota}^\top \mathbf{A}^k \boldsymbol{\iota} \\
&= N_G(\varphi) + \sum_{k=0}^{\infty} k \varphi^k \boldsymbol{\iota}^\top \mathbf{A}^k \boldsymbol{\iota}.
\end{aligned}$$

Alternatively, we can write

$$\sum_{k=0}^{\infty} (k+1) \varphi^k \boldsymbol{\iota}^\top \mathbf{A}^k \boldsymbol{\iota} = \sum_{k=0}^{\infty} (k+1) N_k \varphi^k = \frac{d}{d\varphi} (\varphi N_G(\varphi)),$$

so that

$$\boldsymbol{\iota}^\top \mathbf{M}(G,\varphi)^2 \boldsymbol{\iota} = \frac{d}{d\varphi} (\varphi N_G(\varphi)) = N_G(\varphi) + \varphi \frac{d}{d\varphi} N_G(\varphi).$$

In the k-regular graph G_k it holds that $N_G(\varphi) = \frac{n}{1-k\varphi}$ and $\frac{d}{d\varphi}(\varphi N_G(\varphi)) = N_G(\varphi) + \varphi \frac{d}{d\varphi} = N_G(\varphi) = \frac{n}{1-k\varphi} + \frac{nk\varphi}{(1-k\varphi)^2} = \frac{n}{1-k\varphi}\left(1 + \frac{k\varphi}{1-k\varphi}\right) = \frac{n}{(1-k\varphi)^2}$. Using the fact that the number of links in a k-regular graph is given by $m = \frac{nk}{2}$ we obtain a lower bound on welfare in the efficient graph given by $\frac{\mu^2 n}{(1-\frac{2m}{n}\varphi)^2} \leq W(G^*)$. This lower bound is highest for the complete graph K_n where $m = n(n-1)/2$, so that[11]

$$\frac{\mu^2 n}{(1-(n-1)\varphi)^2} \leq W(G^*).$$

In order to derive an upper bound, observe that

$$\begin{aligned}
\boldsymbol{\iota}^\top \mathbf{A}^k \boldsymbol{\iota} &= \sum_{i=1}^{n} (\boldsymbol{\iota}^\top \mathbf{v}_i)^2 \lambda_i^k, \\
N_G(\varphi) &= \sum_{i=1}^{n} \frac{(\mathbf{v}_i^\top \boldsymbol{\iota})^2}{1-\lambda_i \varphi},
\end{aligned}$$

[11] Using Rayleigh's inequality, one can show that $\frac{d}{d\varphi}(\varphi N_G(\varphi)) \geq \frac{1}{\lambda_1} \frac{d}{d\varphi}$ [Van Mieghem, 2011, p. 51]. From this we can obtain a lower bound on welfare given by $W(G) \geq \mu^2 \frac{1}{\lambda_1} \frac{d}{d\varphi}(N_G(\varphi))$.

so that we can write

$$\begin{aligned}
\boldsymbol{\iota}^\top \mathbf{M}(G,\varphi)^2\boldsymbol{\iota} &= \sum_{i=1}^n \frac{(\mathbf{v}_i^\top \boldsymbol{\iota})^2}{1-\lambda_i\varphi} + \sum_{i=1}^n (\boldsymbol{\iota}^\top \mathbf{v}_i)^2 \sum_{k=0}^\infty k\varphi^k\lambda_i^k \\
&= \sum_{i=1}^n \frac{(\mathbf{v}_i^\top \boldsymbol{\iota})^2}{1-\lambda_i\varphi} + \sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2\varphi\lambda_i}{(1-\varphi\lambda_i)^2} \\
&= \sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{1-\varphi\lambda_i}\left(1+\frac{\varphi\lambda_i}{1-\varphi\lambda_i}\right) \\
&= \sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{(1-\varphi\lambda_i)^2}.
\end{aligned}$$

From the above it follows that welfare can also be written as

$$W(G) = \mu^2 \frac{d}{d\varphi}(\varphi N_G(\varphi)) = \mu^2 \sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{(1-\varphi\lambda_i)^2}.$$

This expression shows that gross welfare is highest in the graph where λ_1 approaches $1/\varphi$. We then can upper bound welfare as follows[12]

$$W(G) = \mu^2 \sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{(1-\varphi\lambda_i)^2} \le \mu^2 \frac{\sum_{i=1}^n (\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{(1-\varphi\lambda_1)^2} \le \mu^2 \frac{n}{(1-\varphi\lambda_1)^2},$$

where we have used the fact that $N_G(0) = \sum_{i=1}^n (\boldsymbol{\iota}^\top \mathbf{v}_i)^2 = n$ so that $(\boldsymbol{\iota}^\top \mathbf{v}_1)^2 < n$. Note that the largest eigenvalue λ_1 is upper bounded by the largest eigenvalue of the complete graph K_n, where it is equal to $n-1$. In this case, upper and lower bounds coincide, and the efficient graph is therefore complete, that is $K_n = \text{argmax}_{G\in\mathcal{G}^n} W(G)$.

(i) Welfare can be written as

$$W(G) = \frac{2-\rho}{2}\frac{\mu^2}{\rho^2}\frac{\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota} + \frac{\rho}{2-\rho}(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2}{\left(\frac{1-\rho}{\rho} + \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota}\right)^2}.$$

For the k-regular graph G_k we have that

$$\begin{aligned}
\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} &= \frac{n}{1-(k-1)\phi}, \\
\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota} &= \frac{n}{(1-(k-1)\phi)^2},
\end{aligned}$$

and welfare is given by

$$W(G_k) = \frac{\mu^2 n((n-1)\rho+2)}{2(\rho(k\phi+n-1)-k\phi+1)^2}.$$

[12] An alternative proof uses the fact that $\lambda_1 \ge \left(\frac{N_k(G)}{n}\right)^{\frac{1}{k}}$ [cf. Van Mieghem, 2011, p. 47], so that $\frac{d}{d\varphi}(\varphi N_G(\varphi)) = \sum_{k=0}^\infty \varphi^k(k+1)N_k(\varphi) \le n\sum_{k=0}^\infty (\lambda_1\varphi)^k(k+1) = n\sum_{k=0}^\infty(\lambda_1\varphi)^k + n\sum_{k=0}^\infty k(\lambda_1\varphi)^k = n\left(\frac{1}{1+\varphi\lambda_1} + \frac{\varphi\lambda_1}{(1+\varphi\lambda_1)^2}\right) = \frac{n}{(1+\varphi\lambda_1)^2}$.

As $k = 2m/n$ this is

$$W(G_k) = \frac{\mu^2 n^3((n-1)\rho + 2)}{2(2m(\rho-1)\phi + (n-1)n\rho + n)^2}.$$

Together with the definition of the average degree $\bar{d} = \frac{2m}{n}$ this gives us the lower bound on welfare for all graphs with m links. For the complete graph K_n we get

$$\begin{aligned}
\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} &= \frac{n}{1-(n-1)\phi}, \\
\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota} &= \frac{n}{(1-(n-1)\phi)^2},
\end{aligned}$$

so that we obtain for welfare in the complete graph

$$W(K_n) = \frac{\mu^2 n(2+(n-1)\rho)}{2((n-1)\rho(\phi+1)-(n-1)\phi+1)^2}.$$

Using the fact that $\phi = \frac{\varphi}{1-\rho}$ we can write this as follows

$$W(K_n) = \frac{\mu^2 n(2+(n-1)\rho)}{2((n-1)\rho-(n-1)\varphi+1)^2}.$$

This gives us the lower bound on welfare $W(K_n) \leq W(G^*)$. To obtain an upper bound, note that welfare can be written as

$$W(G) = \frac{\mu^2}{2\rho^2} \frac{(2-\rho)\frac{\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota}}{(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2} + \rho}{\frac{\left(\frac{1-\rho}{\rho} + \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota}\right)^2}{(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2}}.$$

Next, observe that

$$\frac{\left(\frac{1-\rho}{\rho} + \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota}\right)^2}{(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2} = \left(1 + \frac{1-\rho}{\rho}\frac{1}{\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota}}\right)^2 \geq \left(1 + \frac{1-\rho}{\rho}\frac{1-\lambda_1\phi}{n}\right)^2,$$

where we have used the fact that $\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} = N_G(\phi) \leq \frac{n}{1-\lambda_1\phi}$. This implies that

$$W(G) \leq \frac{\mu^2}{2\rho^2} \frac{(2-\rho)\frac{\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota}}{(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2} + \rho}{\left(1 + \frac{1-\rho}{\rho}\frac{1-\lambda_1\phi}{n}\right)^2} \tag{G.32}$$

Next, observe that the Herfindahl industry concentration index is defined as $H = \sum_{i=1}^n s_i^2$, where the market share of firm i is given by $s_i = \frac{q_i}{\sum_{j=1}^n q_j}$ [cf. e.g. Tirole, 1988]. Using our equilibrium characterization from Equation (8) we can write

$$H(G) = \sum_{i=1}^n \left(\frac{q_i}{\sum_{j=1}^n q_j}\right)^2 = \frac{\sum_{i=1}^n b_i(G,\phi)^2}{\left(\sum_{j=1}^n b_j(G,\phi)\right)^2} = \frac{\mathbf{b}(G,\phi)^\top \mathbf{b}(G,\phi)}{(\boldsymbol{\iota}^\top \mathbf{b}(G,\phi))^2} = \frac{\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota}}{(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2}. \tag{G.33}$$

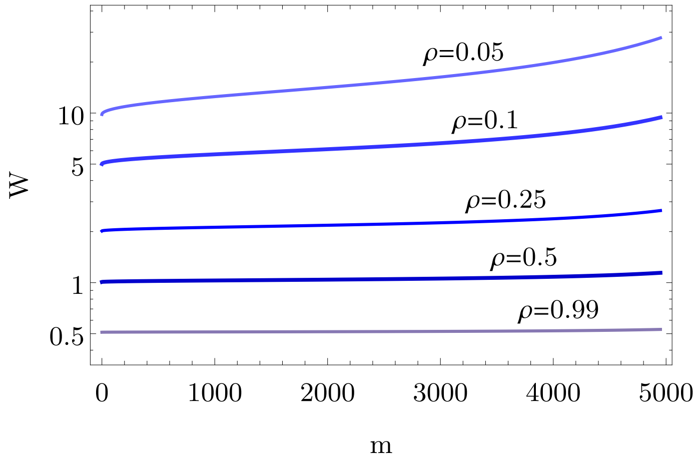


Figure G.3: The RHS in Equation (G.35) with varying values of $m \in \{0, 1, \ldots, n(n-1)/2\}$ for $n = 100$, $\varphi = 0.9(1-\rho)/n$ and $\rho \in \{0.05, 0.1, 0.25, 0.5, 0.99\}$.

The upper bound for welfare can then be written more compactly as follows

$$W(G) \leq \frac{\mu^2}{2\rho^2} \frac{(2-\rho)H(G) + \rho}{\left(1 + \frac{1-\rho}{\rho}\frac{1-\lambda_1\phi}{n}\right)^2}. \tag{G.34}$$

Further, we have that

$$\begin{aligned} H(G) &= \frac{\boldsymbol{\iota}^\top \mathbf{M}^2(G,\phi)\boldsymbol{\iota}}{(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2} = \frac{\frac{d}{d\phi}\left(\phi N_G(\phi)\right)}{N_G(\phi)^2} = \frac{\sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{(1-\phi\lambda_i)^2}}{\left(\sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{1-\phi\lambda_i}\right)^2} \leq \frac{\frac{1}{1-\phi\lambda_1}\sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{1-\phi\lambda_i}}{\left(\sum_{i=1}^n \frac{(\boldsymbol{\iota}^\top \mathbf{v}_i)^2}{1-\phi\lambda_i}\right)^2} \\ &= \frac{1}{(1-\phi\lambda_1)N_G(\phi)} \leq \frac{1}{(1-\phi\lambda_1)(n+2m\phi)} \leq \frac{1}{(1-\phi\sqrt{\frac{2m(n-1)}{n}})(n+2m\phi)}, \end{aligned}$$

where we have used the fact that $N_G(\phi) \geq n + 2m\phi$ for $\phi \in [0, 1/\lambda_1)$, and the upper bound $\lambda_1 \leq \sqrt{\frac{2m(n-1)}{n}}$ [cf. Van Mieghem, 2011, p. 52]. Inserting into the upper bound in Equation (G.32) and substituting $\phi = (1-\rho)/\varphi$ gives

$$W(G^*) \leq \frac{\mu^2 n^2}{2} \frac{\rho + (2-\rho)\frac{(1-\rho)^2}{(n(1-\rho)+2m\varphi)\left(1-\rho-\varphi\sqrt{\frac{2m(n-1)}{n}}\right)}}{\left(1+(n-1)\rho - \varphi\sqrt{\frac{2m(n-1)}{n}}\right)^2}. \tag{G.35}$$

The RHS in Equation (G.35) is increasing in m (see Figure G.3) and attains its maximum at $m = n(n-1)/2$, where we get

$$W(G^*) \leq \frac{\mu^2 n\left((\rho-1)^2((n-1)\rho+2) - (n-1)^2 n\rho\varphi^2\right)}{2((n-1)\rho - n\varphi + \varphi + 1)^2\left((\rho-1)^2 - (n-1)^2\varphi^2\right)}.$$

(iii) Assuming that $\mu_i = \mu$ for all $i = 1, \ldots, n$, we have that

$$\mathbf{q} = \frac{\mu}{1 + \rho(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} - 1)}\mathbf{M}(G,\phi)\boldsymbol{\iota},$$

with $\mathbf{M}(G,\phi) \equiv (\mathbf{I}-\phi\mathbf{A})^{-1}$, and we can write

$$W(G) = \frac{\mu^2}{2(1+\rho(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota}-1))^2}\left((2-\rho)\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota} + \rho(\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2\right).$$

Using the fact that $\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota} = N_G(\phi)$ and $\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota} = \frac{d}{d\phi}(\phi N_G(\phi))$, we then can write welfare in terms of the walk generating function $N_G(\phi)$ as

$$W(G) = \frac{\mu^2}{2(1+\rho(N_G(\phi)-1))^2}\left((2-\rho)\frac{d}{d\phi}(\phi N_G(\phi)) + \rho N_G(\phi)^2\right).$$

Next, observe that

$$N_G(\phi) = N_0 + N_1\phi + N_2\phi^2 + O(\phi^3),$$

and consequently

$$\frac{d}{d\phi}(\phi N_G(\phi)) = N_0 + 2N_1\phi + 3N_2\phi^2 + O(\phi^3).$$

Inserting into welfare gives

$$W(G) = \frac{\mu^2 N_0((N_0-1)\rho+2)}{2((N_0-1)\rho+1)^2} - \frac{\mu^2 N_1(\rho-1)((N_0-1)\rho+2)}{((N_0-1)\rho+1)^3}\phi + O(\phi)^2.$$

Using the fact that $N_0 = n$ and $N_1 = 2m$ we get

$$W(G) = \frac{\mu^2 n((n-1)\rho+2)}{2((n-1)\rho+1)^2} + \frac{2\mu^2 m(1-\rho)(2+(n-1)\rho)}{(1+(n-1)\rho)^3}\phi + O(\phi)^2.$$

Up to terms linear in ϕ this is an increasing function of m, and hence is largest in the complete graph K_n.

(iv) Welfare can be written as

$$W(G) = \frac{\mu^2\left((\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota})^2\rho + \boldsymbol{\iota}^\top \mathbf{M}(G,\phi)^2\boldsymbol{\iota}(2-\rho)\right)}{2((\boldsymbol{\iota}^\top \mathbf{M}(G,\phi)\boldsymbol{\iota}-1)\rho+1)^2}.$$

For the complete graph we obtain

$$\boldsymbol{\iota}^\top \mathbf{M}(K_n,\phi)\boldsymbol{\iota} = \frac{n}{1-(n-1)\phi},$$
$$\boldsymbol{\iota}^\top \mathbf{M}(K_n,\phi)^2\boldsymbol{\iota} = \frac{n}{(1-(n-1)\phi)^2}.$$

With $\phi = \frac{\varphi}{1-\rho}$ welfare in the complete graph is given by

$$W(K_n) = \frac{\mu^2 n((n-1)\rho+2)}{2((n-1)\rho - n\varphi + \varphi + 1)^2},$$

For the star $K_{1,n-1}$

$$\boldsymbol{\iota}^\top \mathbf{M}(K_{1,n-1},\phi)\boldsymbol{\iota} = \frac{2(n-1)\phi+n}{1-(n-1)\phi^2},$$
$$\boldsymbol{\iota}^\top \mathbf{M}(K_{1,n-1},\phi)^2\boldsymbol{\iota} = \frac{(n-1)n\phi^2+4(n-1)\phi+n}{((n-1)\phi^2-1)^2}.$$

Inserting $\phi = \frac{\varphi}{1-\rho}$, welfare in the star is then given by

$$
\begin{aligned}
&W(K_{1,n-1}) \\
&= \frac{\mu^2\left((n-1)\varphi^2(n(3\rho+2)-4\rho)-4(n-1)(\rho-1)\varphi((n-1)\rho+2)+n(\rho-1)^2((n-1)\rho+2)\right)}{2\left(-2(n-1)\rho\varphi+(\rho-1)((n-1)\rho+1)+(n-1)\varphi^2\right)^2}.
\end{aligned} \tag{G.36}
$$

Welfare of the star $K_{1,n-1}$ for varying values of ρ can be seen in Figure G.4, right panel. For the ratio of welfare in the complete graph and the star we then obtain

$$
\begin{aligned}
&\frac{W(K_n)}{W(K_{1,n-1})} = n(2+(n-1)\rho)\left(2(n-1)\rho\varphi+(1-\rho)((n-1)\rho+1)-(n-1)\varphi^2\right)^2 \\
&\times\left((1+(n-1)\rho-(n-1)\varphi)^2\left((n-1)\varphi^2(n(3\rho+2)-4\rho)\right.\right. \\
&\left.\left.+4(n-1)(1-\rho)\varphi((n-1)\rho+2)+n(1-\rho)^2((n-1)\rho+2)\right)\right)^{-1}.
\end{aligned}
$$

This ratio equals one when $\varphi = \varphi^*(n,\rho)$, which is given by

$$
\begin{aligned}
&\varphi^*(n,\rho) = \frac{1}{6A(n-1)((n-1)\rho+n)} \\
&\times\left(\sqrt[3]{2}A^2+2A(n-1)(2-\rho(3(n-1)\rho+5))+2^{2/3}(n-1)\right) \\
&\times\left(6n^2-(n-1)(15(n-2)n+8)\rho^2+(n(3(n-16)n+76)-16)\rho-32n+8\right),
\end{aligned}
$$

where we have denoted by

$$
\begin{aligned}
&A = \left(-3(n-1)^2\left(n\left(3n\left(6n^2-33n+86\right)-248\right)+32\right)\right. \\
&\times\rho^2-27(n-2)(n-1)^4n\rho^4+(n-1)^3(9(n-2)n(3n-19)-32)\rho^3 \\
&\left.+3\sqrt{3}B-12n(n(5n(3(n-5)n+31)-153)+66)\rho-16n(n(n(9n-29)+33)-15)+96\rho-32\right)^{\frac{1}{3}},
\end{aligned}
$$

and

$$
\begin{aligned}
&B = \left((n-2)(n-1)^3n((n-1)\rho+n)^2\right. \\
&\times\left(27(n-2)(n-1)^3n\rho^6-2(n-1)^2(9(n-2)n(6n-19)-32)\rho^5\right. \\
&+(n-1)(n(n(2n(37n-526)+3283)-3046)+384)\rho^4 \\
&+2(n(n(n(n(n+242)-1936)+4384)-3264)+448)\rho^3 \\
&\left.\left.+4((n-2)n(n(3n+302)-786)-256)\rho^2+24(n-2)(n(n+56)-12)\rho+16(n(n+34)-8)\right)\right)^{\frac{1}{2}}.
\end{aligned}
$$

We then have that $W(K_n) > W(K_{1,n-1})$ if $\varphi < \varphi^*(n,\rho)$ and $W(K_n) < W(K_{1,n-1})$ otherwise. An illustration can be seen in Figure G.4, left panel.

□

The upper and lower bounds of case (i) in Proposition 4 on welfare can be seen in Figure G.2. The bounds indicate that welfare is typically increasing in strength of technology spillovers, φ, and decreasing in the degree of competition, ρ, at least when these are not too high. The figure is also consistent with cases (ii) and (iii), where it is shown that for weak spillovers the complete graph is efficient. However, Proposition 4, case (iv), shows that in the presence of stronger externalities through R&D spillovers and competition, the star network generates higher welfare than the complete network. This happens when the welfare gains through concentration, which enter the welfare function through the Herfindahl index H in Equation

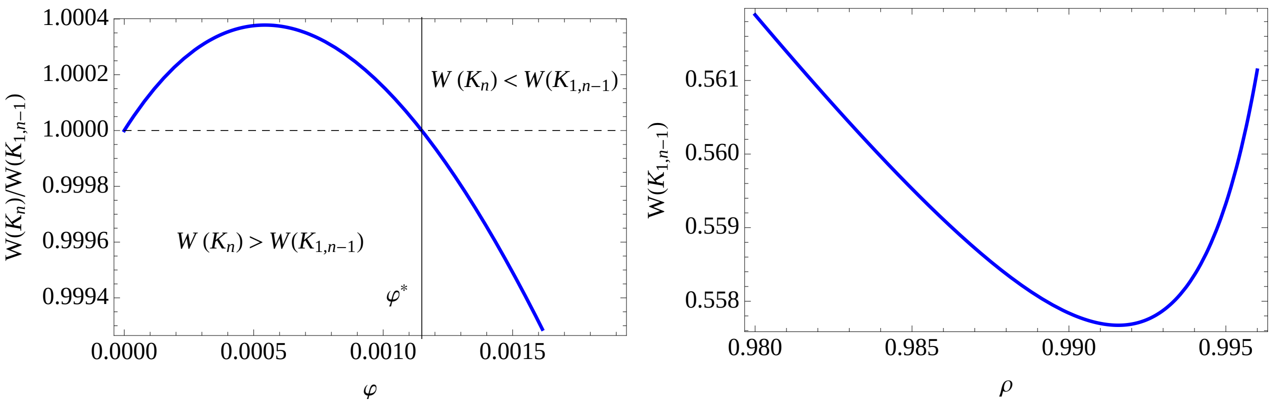


Figure G.4: (Left panel). The ratio of welfare in the complete graph, K_n, and the star, $K_{1,n-1}$, for $n = 10$, $\rho = 0.981$ and varying values of φ ($< ((1-\rho)/\lambda_{\max}(K_n) = 0.002$) (Right panel) Welfare in the star, $K_{1,n-1}$, with varying values of ρ for $n = 10$ and $\varphi = 0.001$ ($< (1-\rho)/\lambda_{\max}(K_{1,n-1})$ for all values of ρ considered).

(G.30), dominate the welfare gains through maximizing total output Q.

While total output Q (and total R&D) is increasing with the degree of competition, measured by ρ (*Schumpeterian effect*; see e.g. Aghion et al. [2014]), this may not necessarily hold for welfare. This is illustrated in the right panel in Figure G.4 where welfare for the star is shown for varying values of ρ. The presence of externalities through R&D spillovers and business stealing effects through market competition in highly centralized networks can thus give rise to a non-monotonic relationship between competition and welfare [cf. Aghion et al., 2005]. The centralization of the network structure, however, seems to be important for this result, as for example in a regular graph (such as the complete graph) welfare is decreasing monotonically with increasing ρ.[13]

H. Data

In the following appendices we give a detailed account on how we constructed our data sample. In Appendix H.1 we describe the two raw datasources we have used to obtain information on R&D collaborations between firms. In Appendix H.2 we explain how we complemented these data with information about mergers and acquisitions, while Appendix H.3 explains how we supplemented the alliance information with firms' balance sheet statements. Moreover, Appendix H.4 discusses the geographic distribution of the firms in our data sample. Finally, Appendix H.5 provides the details on how we complemented the alliance data with the firms patent portfolios and computed their technological proximities.

H.1. R&D Network

To get a comprehensive picture of alliances we use data on interfirm R&D collaborations stemming from two sources which have been widely used in the literature [cf. Schilling, 2009]. The first is the Cooperative Agreements and Technology Indicators (CATI) database [cf. Hagedoorn, 2002]. The database only records agreements for which a combined innovative activity or an exchange of technology is at least part of the agreement. Moreover, only agreements that have at least two industrial partners are included in the database, thus agreements involving only universities or government labs, or one company with a university or lab, are disregarded. The second is the Thomson Securities Data Company (SDC) alliance database. SDC collects

[13]Decreasing welfare with increasing competition is a feature not only of the standard Cournot model (without externalities) but also of many traditional models in the literature including Aghion and Howitt [1992], and Grossman and Helpman [1991].

data from the U. S. Securities and Exchange Commission (SEC) filings (and their international counterparts), trade publications, wires, and news sources. We include only alliances from SDC which are classified explicitly as research and development collaborations. A comparative analysis of these two databases (and other alternative databases) can be found in Schilling [2009].

We then merged the CATI database with the Thomson SDC alliance database. For the matching of firms across datasets we adopted the name matching algorithm developed as part of the NBER patent data project [Trajtenberg et al., 2009] and developed further by Atalay et al. [2011].[14] From the firms in the CATI database and the firms in the SDC database we could match 21% of the firms appearing in both databases. Considering only firms without missing observations on sales, output and R&D expenditures (see also Appendix H.3 below on how we obtained balance sheet and income statement information), gives us a sample of $1,186$ firms and a total of 1010 collaborations over the years 1967 to 2006.[15] The average degree of the firms in this sample is 1.68 with a standard deviation of 4.83 and the maximum degree is 63 attained by *Motorola Inc.*. Figure H.5 shows the largest connected component of the R&D collaboration network with all links accumulated up to the year 2005 (see Appendix B.1). The figure indicates two clusters appearing which are related to the different industries in which firms are operating. This may indicate specialization in R&D alliance partnerships.

Figure H.6 shows the average clustering coefficient, C, the relative size of the largest connected component, $\max_{\{H \subseteq G\}} |H|/n$, the average path length, ℓ, and the eigenvector centralization C_v (relative to a star network of the same size) over the years 1990 to 2005 (see Wasserman and Faust [1994] and Appendix B.1 for the definitions). We observe that the network shows the highest degree of clustering in the year 1990 and the largest connected component around the year 1997, an average path length of around 5, and a centralization index C_v between 0.3 and 0.7. Moreover, comparing our subsample and the original network (where firms have not been dropped because of missing accounting information) we find that both exhibit similar trends over time. This seems to suggest that the patterns found in the subsample are representative for the overall patterns in the data (see also Section J.5). Further, the clustering coefficient and the size of the largest connected component exhibit a similar trend as the number of firms and the average number of collaborations that we have seen already in Figure 2.

Figure H.7 shows the degree distribution, $P(d)$, the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, the clustering degree distribution, $C(d)$, and the component size distribution, $P(s)$ across different years of observation [cf. e.g. König, 2016]. The degree distribution decays as a power law, the average nearest neighbor degree is weakly increasing with the degree, indicating a weakly assortative network, the clustering degree distribution is decreasing with the degree and the component size distribution indicates a large connected component (see also Figure H.5) with smaller components decaying as a power law.

Figure H.8 and Tables H.1 and H.2 illustrate the industrial composition of our sample of R&D collaborating firms at the main 2-digit and 4-digit standard industry classification (SIC) levels, respectively. At the 2-digit level, the chemicals and allied products sectors make up for the largest fraction (22.43%) of firms in our data, followed by business services and electronic equipment. This sectoral composition is similar to the one provided in Schilling [2009], who identifies the biotech and information technology sectors as the most prominent in the CATI and SDC R&D collaboration databases.

[14]See `https://sites.google.com/site/patentdataproject`. We would like to thank Enghin Atalay and Ali Hortacsu for sharing their name matching algorithm with us.

[15]This is the sample that we have used for our empirical analysis in Section 6.

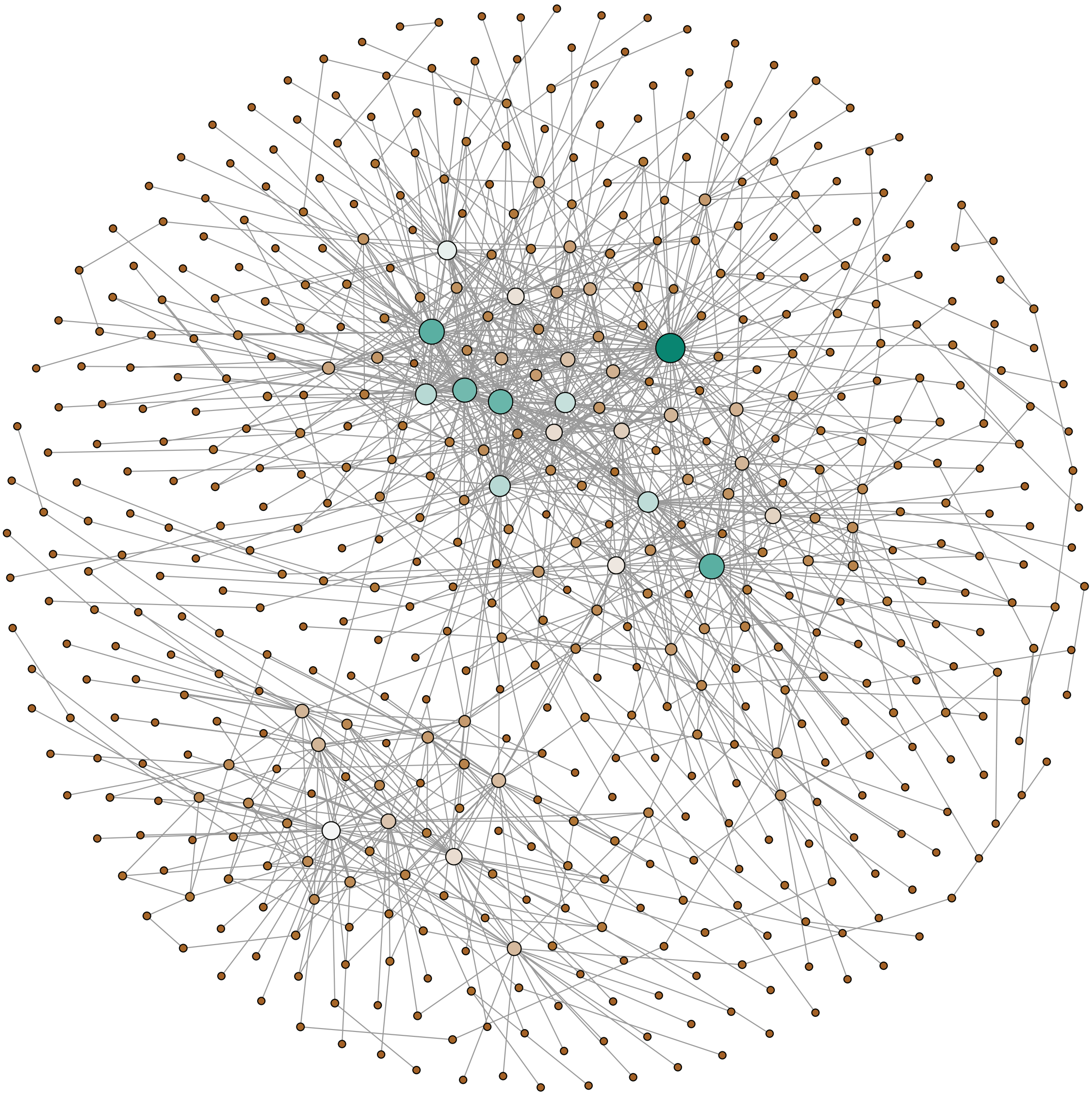


Figure H.5: The largest connected component of the R&D collaboration network with all links accumulated until the year 2005. The nodes' colors indicate sectors according to 4-digit SIC codes while the nodes' sizes indicate the number of collaborations of a firm.

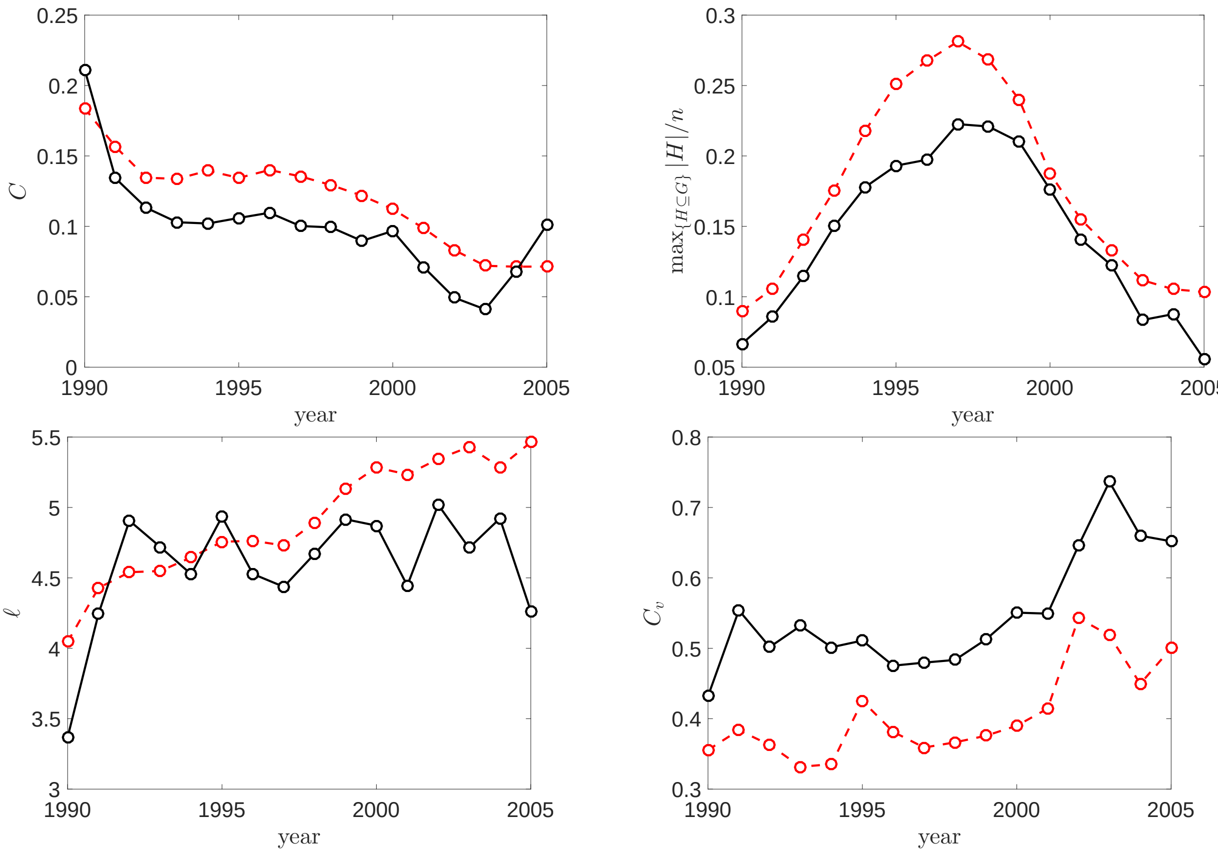


Figure H.6: The average clustering coefficient, C, the relative size of the largest connected component, $\max_{\{H \subseteq G\}} |H|/n$, the average path length, ℓ, and the eigenvector centralization C_v (relative to a star network of the same size) over the years 1990 to 2005 (see Appendix B.1). Dashed lines indicate the corresponding quantities for the original network (where firms have not been dropped because of missing accounting information), while solid lines indicate the subsample with 1,186 firms that we have used in the empirical Section 6.

Table H.1: The 20 largest sectors at the 2-digit SIC level.

Sector	2-dig SIC	# firms	% of tot.	Rank
Chemical and Allied Products	28	266	22.43	1
Business Services	73	198	16.69	2
Electronic and Other Electric Equipment	36	187	15.77	3
Instruments and Related Products	38	154	12.98	4
Industrial Machinery and Equipment	35	150	12.65	5
Transportation Equipment	37	47	3.96	6
Engineering and Management Services	87	25	2.11	7
Primary Metal Industries	33	18	1.52	8
Fabricated Metal Products	34	15	1.26	9
Oil and Gas Extraction	13	14	1.18	10
Communications	48	14	1.18	11
Rubber and Miscellaneous Plastics Products	30	10	0.84	12
Paper and Allied Products	26	9	0.76	13
Petroleum and Coal Products	29	9	0.76	14
Health Services	80	9	0.76	15
Food and Kindred Products	20	8	0.67	16
Miscellaneous Manufacturing Industries	39	7	0.59	17
Electric Gas and Sanitary Services	49	6	0.51	18
Textile Mill Products	22	5	0.42	19
Stone Clay and Glass Products	32	5	0.42	20

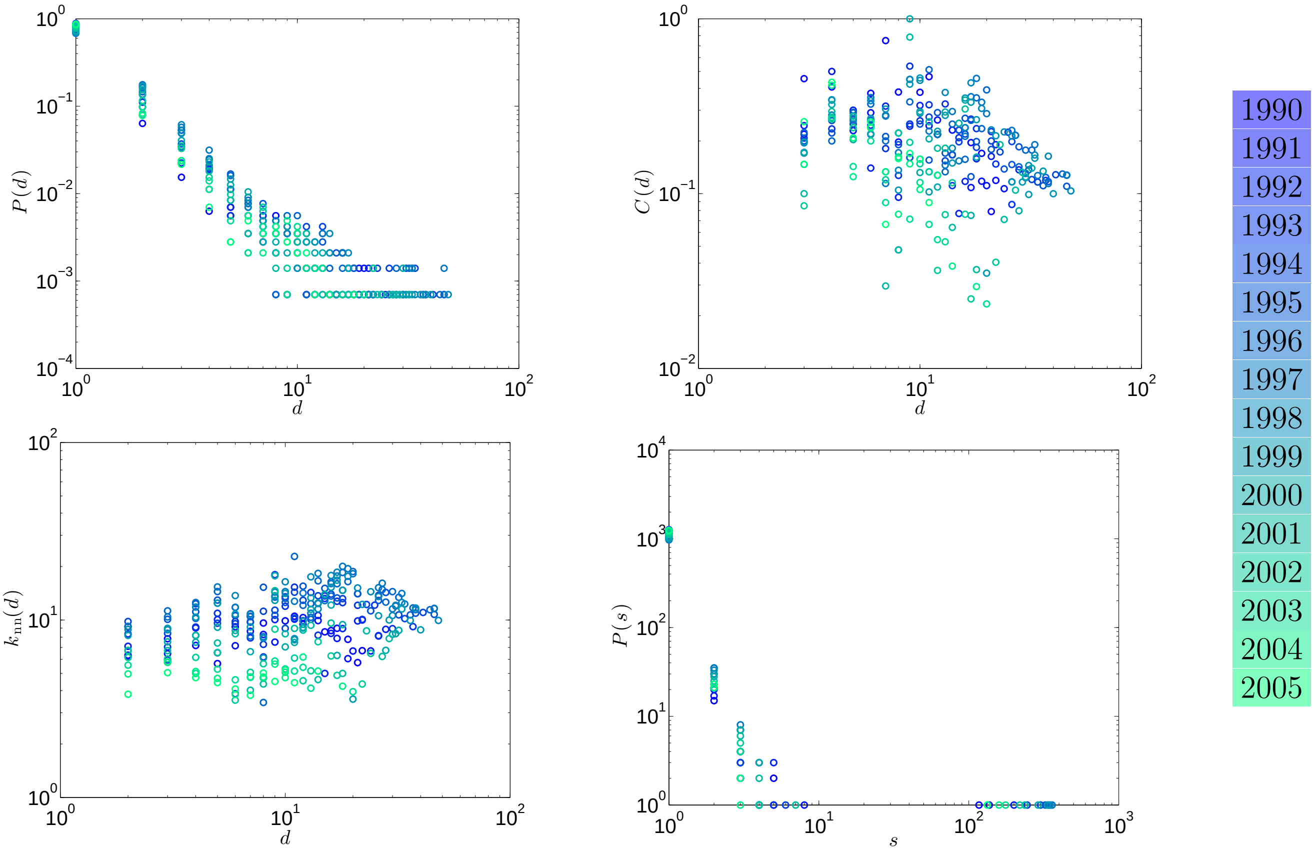


Figure H.7: The degree distribution, $P(d)$, the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, the clustering degree distribution, $C(d)$, and the component size distribution, $P(s)$.

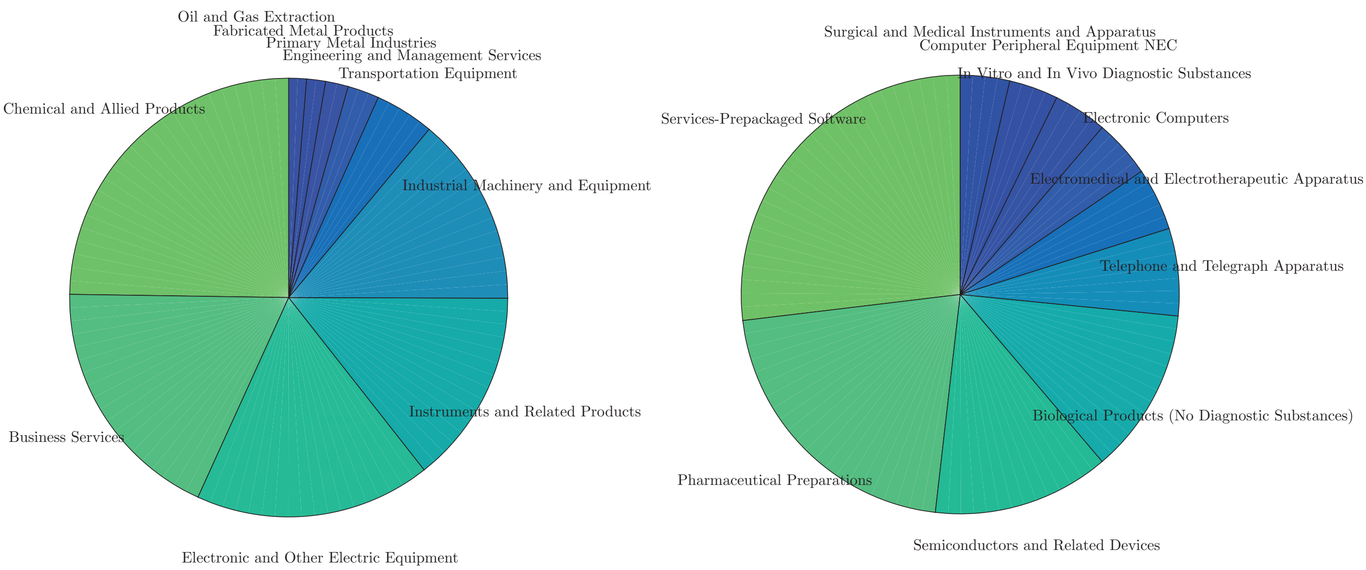


Figure H.8: The shares of the ten largest sectors at the 2-digit (left panel) and 4-digit (right panel) SIC levels.

Table H.2: The 20 largest sectors at the 4-digit SIC level.

Sector	4-dig SIC	# firms	% of tot.	Rank
Services-Prepackaged Software	7372	163	13.74	1
Pharmaceutical Preparations	2834	129	10.88	2
Semiconductors and Related Devices	3674	79	6.66	3
Biological Products (No Diagnostic Substances)	2836	74	6.24	4
Telephone and Telegraph Apparatus	3661	39	3.29	5
Electromedical and Electrotherapeutic Apparatus	3845	28	2.36	6
Electronic Computers	3571	26	2.19	7
In Vitro and In Vivo Diagnostic Substances	2835	24	2.02	8
Computer Peripheral Equipment NEC	3577	22	1.85	9
Surgical and Medical Instruments and Apparatus	3841	22	1.85	10
Special Industry Machinery NEC	3559	21	1.77	11
Laboratory Analytical Instruments	3826	20	1.69	12
Services-Computer Integrated Systems Design	7373	20	1.69	13
Radio and TV Broadcasting and Communications Equipment	3663	18	1.52	14
Motor Vehicle Parts and Accessories	3714	18	1.52	15
Instruments For Meas and Testing of Electricity and Elec Signals	3825	17	1.43	16
Computer Storage Devices	3572	15	1.26	17
Computer Communications Equipment	3576	14	1.18	18
Search Detection Navigation Guidance Aeronautical Sys	3812	14	1.18	19
Services-Commercial Physical and Biological Research	8731	14	1.18	20

H.2. Mergers and Acquisitions

Some firms might be acquired by other firms due to mergers and acquisitions (M&A) over time, and this will impact the R&D collaboration network [cf. Hanaki et al., 2010].

To get a comprehensive picture of the M&A activities of the firms in our dataset, we use two extensive datasources to obtain information about M&As. The first is the Thomson Reuters' Securities Data Company (SDC) M&A database, which has historically been the most widely used database for empirical research in the field of M&As. Data in SDC dates back to 1965 with a slightly more complete coverage of deals starting in the early 1980s. The second database with information about M&As is Bureau van Dijk's (BvD) Zephyr database, which is a recent alternative to the SDC M&As database. The history of deals recorded in Zephyr goes back to 1997. In 1997 and 1998 only European deals are recorded, while international deals are included starting from 1999. According to Huyghebaert and Luypaert [2010], Zephyr "covers deals of smaller value and has a better coverage of European transactions". A comparison and more detailed discussion of the two databases can be found in Bollaert and Delanghe [2015] and Bena et al. [2008].

We merged the SDC and Zephyr databases (with the above mentioned name matching algorithm; see also Atalay et al. [2011]; Trajtenberg et al. [2009]) to obtain information on M&As of $116,641$ unique firms. Using the same name matching algorithm we could identify 43.08% of the firms in the combined CATI-SDC alliance database that also appear in the combined SDC-Zephyr M&As database. We then account for the M&A activities of these matched firms when constructing the R&D collaboration network by assuming that an acquiring firm in a M&A inherits all the R&D collaborations of the target firm, and we remove the target firm form from the network.

H.3. Balance Sheet Statements

The combined CATI-SDC alliance database provides the names for each firm in an alliance, but it does not contain information about the firms' output levels or R&D expenses. We therefore matched the firms' names in the combined CATI-SDC database with the firms' names in Standard & Poor's Compustat U.S. fundamentals annual database and Bureau van Dijk (BvD)'s Osiris database, to obtain information about their balance sheets and income statements.[16] These databases contain only firms listed on the stock market, so they typically exclude smaller private firms, but this is inevitable if one is going to use market value data. Nevertheless, R&D is concentrated in publicly listed firms, and our data sources thus cover most of the R&D activities in the economy [cf. e.g. Bloom et al., 2013]. Compustat contains financial data extracted from company filings.

Compustat North America is a database of U.S. and Canadian fundamental and market information on active and inactive publicly held companies. It provides more than 300 annual and 100 quarterly income statements, balance sheets and statement of cash flows. The Compustat database covers 99% of the total market capitalization with annual company data history available back to 1950.

Osiris is owned by Bureau van Dijk (BvD) and it contains a wide range of accounting and other items for firms from over 120 countries. Osiris contains financial information on globally listed public companies with coverage for up to 20 years on over $62,191$ companies by major international industry classifications. It claims to cover all publicly listed companies worldwide. In addition, it covers major non-listed companies when they are primary subsidiaries of publicly listed companies, or in certain cases, when clients request information from a particular company.

For a detailed comparison and discussion of the Compustat and Osiris databases see Dai [2012] and Papadopoulos [2012].

For the matching of firms across datasets we adopted the name matching algorithm developed as part of the NBER patent data project [Atalay et al., 2011; Trajtenberg et al., 2009]. We could match 25.53% of the firms in the combined CATI-SDC database with the combined Compustat-Osiris database (where accounting information was available). For the matched firms we obtained their sales and R&D expenditures. We adjusted for inflation using the consumer price index of the Bureau of Labor Statistics (BLS), averaged annually, with 1983 as the base year. Individual firms' output levels are computed from deflated sales using 2-SIC digit industry-year specific price deflators from the OECD-STAN database [cf. Gal, 2013]. We then dropped all firms with missing information on sales, output and R&D expenditures. This pruning procedure left us with a subsample of $1,186$, on which the empirical analysis in Section 6 is based.[17]

The empirical distributions for sales, $P(s)$, output, $P(q)$, R&D expenditures, $P(e)$, and the patent stocks, $P(k)$, across different years ranging from 1990 to 2005 (using a logarithmic binning of the data with 100 bins [cf. McManus et al., 1987]) are shown in Figure H.9. All distributions are highly skewed, indicating a large degree of inequality in firms' sizes and patent activities.

H.4. Geographic Location and Distance

In order to determine the locations of the firms in our data we have added the longitude and latitude coordinates associated with the city of residence of each firm in our data. Among the matched cities in our dataset 93.67% could be geo-localized using ArcGIS [cf. e.g. Dell,

[16] We chose to use two alternative database for firm level accounting data to get as much information as possible about balance sheets and income statements for the firms in the R&D collaboration database. The accounting databases used here are complementary, as Compustat features a greater coverage of large companies, while BvD Osiris contains a higher number of small firms and tends to have a better coverage of European firms [cf. Dai, 2012].

[17] Section J.5 discusses how sensitive our empirical results are with respect to subsampling (i.e. missing data).

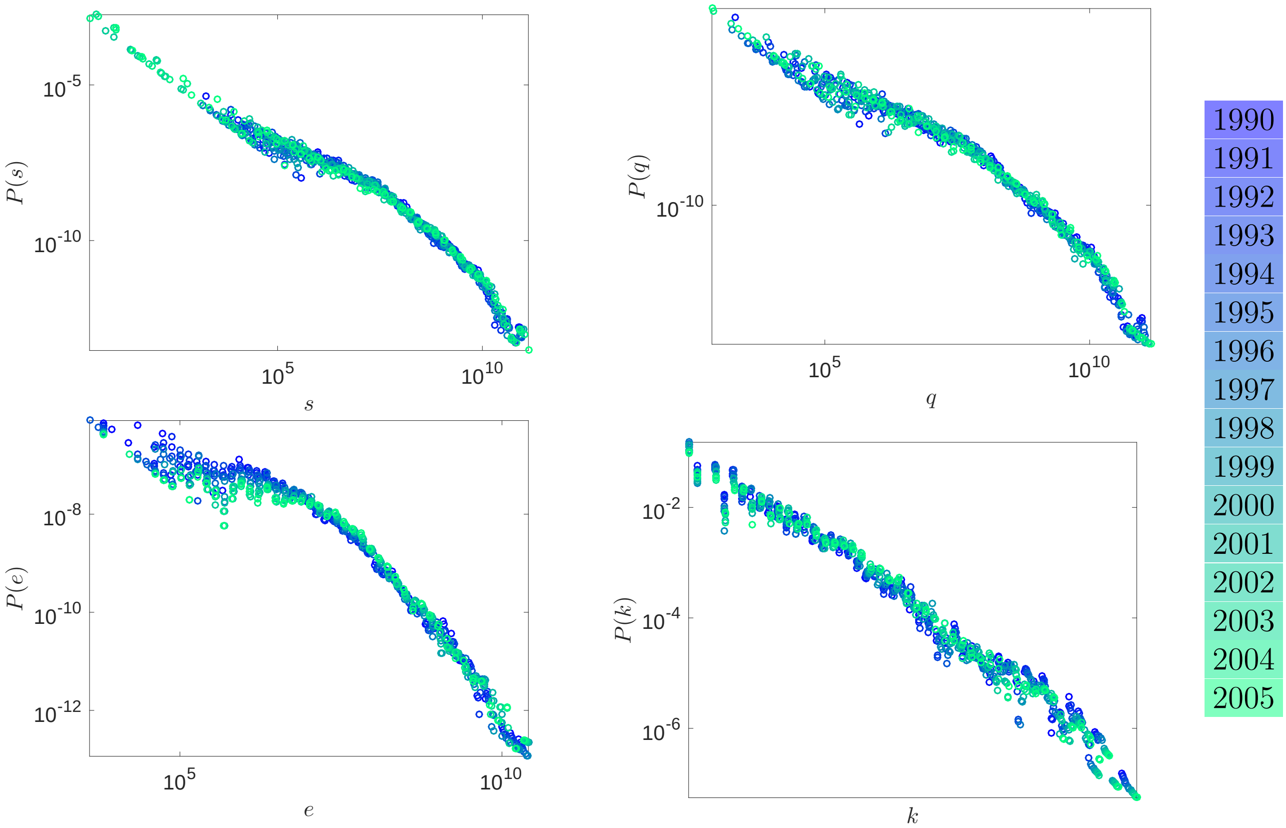


Figure H.9: The sales distribution, $P(s)$, the output distribution, $P(q)$, the R&D expenditures distribution, $P(e)$, and the patent stock distribution, $P(k)$, across different years ranging from 1990 to 2005 using a logarithmic binning of the data [McManus et al., 1987].

2009] and the Google Maps Geocoding API.[18] We then used Vincenty's algorithm to compute the distances between pairs of geo-localized firms [cf. Vincenty, 1975]. The mean distance, $\overline{d}$, and the distance distribution, $P(d)$, across collaborating firms are shown in Figure I.11, while Figure H.10 shows the locations (at the city level) of firms in the database and the collaborations between them. The largest distance between collaborating firms appears around the turn of the millennium, while the distance distribution is heavily skewed. We find that R&D collaborations tend to be more likely between firms that are close, showing that geography matters for R&D collaborations and spillovers, in line with previous empirical studies [cf. Lychagin et al., 2010].

H.5. Patents

We identified the patent portfolios of the firms in our dataset using the EPO Worldwide Patent Statistical Database (PATSTAT) [Hall et al., 2001; Jaffe and Trajtenberg, 2002]. The creation of this worldwide statistical patent database was initiated by the OECD task force on patent statistics. It includes bibliographic details on patents filed to 80 patent offices worldwide, covering more than 60 million documents. Hence filings in all major countries and at the World International Patent Office are covered. We matched the firms in our data with the assignees in the PATSTAT database using the above mentioned name matching algorithm [Atalay et al., 2011; Trajtenberg et al., 2009]. We only consider granted patents (or successful patents), as opposed to patents applied for, as they are the main drivers of revenue derived from R&D expenditures [cf. Copeland and Fixler, 2012]. Using our name matching algorithm we obtained

[18]See https://developers.google.com/maps/documentation/geocoding/intro.

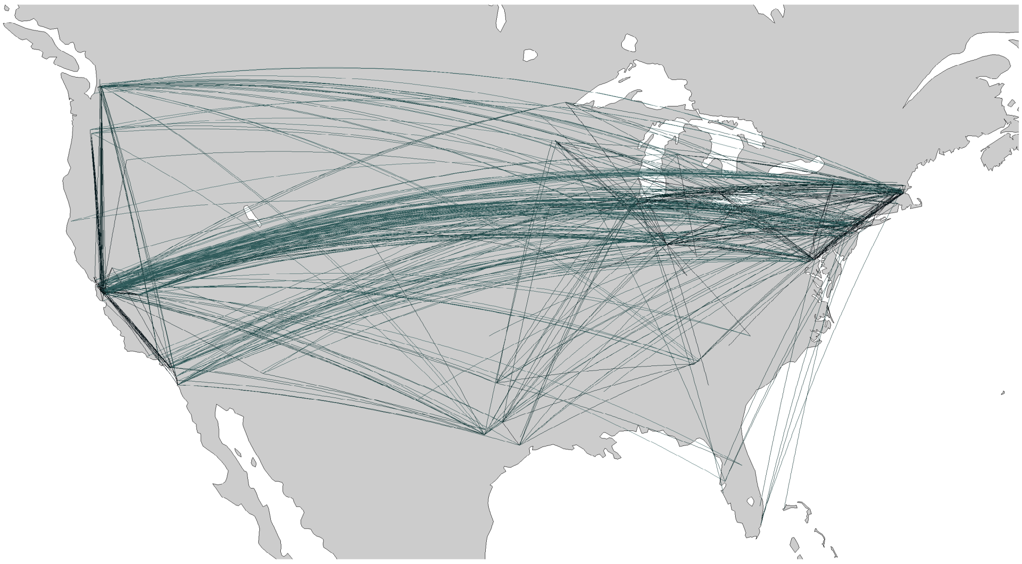


Figure H.10: The locations (at the city level) of firms and their R&D alliances in the combined CATI-SDC databases.

matches for 36.05% of the firms in our data with patent information. The distribution of the number of patents is shown in Figure H.9. The technology classes were identified using the main international patent classification (IPC) numbers at the 4-digit level.

From the firms' patents, we then computed the technological proximity of firm i and j as

$$f_{ij}^{\mathrm{J}} = \frac{\mathbf{P}_i^\top \mathbf{P}_j}{\sqrt{\mathbf{P}_i^\top \mathbf{P}_i}\sqrt{\mathbf{P}_j^\top \mathbf{P}_j}}, \tag{H.37}$$

where, for each firm i, $\mathbf{P}_i$ is a vector whose k-th component, P_{ik}, counts the number of patents firm i has in technology category k divided by the total number of technologies attributed to the firm [cf. Bloom et al., 2013; Jaffe, 1989]. Thus, $\mathbf{P}_i$ represents the patent portfolio of firm i. We use the three-digit U.S. patent classification system to identify technology categories [Hall et al., 2001]. We denote by $\mathbf{F}^{\mathrm{J}}$ the $(n \times n)$ matrix with elements $(f_{ij}^{\mathrm{J}})_{1\leq i,j\leq n}$.

We next consider the Mahalanobis technology proximity measure introduced by Bloom et al. [2013]. To construct this metric, we need to introduce some additional notation. Let N be the number of technology classes, n the number of firms, and let $\mathbf{T}$ be the $(N \times n)$ patent shares matrix with elements

$$T_{ji} = \frac{1}{\sum_{k=1}^n P_{ki}} P_{ji},$$

for all $1 \leq i \leq n$ and $1 \leq j \leq N$. Further, we construct the $(N \times n)$ normalized patent shares matrix $\tilde{\mathbf{T}}$ with elements

$$\tilde{T}_{ji} = \frac{1}{\sqrt{\sum_{k=1}^N T_{ki}^2}} T_{ji},$$

and the $(n \times N)$ normalized patent shares matrix across firms is defined by $\tilde{\mathbf{x}}$ with elements

$$\tilde{X}_{ik} = \frac{1}{\sqrt{\sum_{i=1}^N T_{ki}^2}} T_{ki}.$$

Let $\mathbf{\Omega} = \tilde{\mathbf{x}}^\top \tilde{\mathbf{x}}$. Then the $(n \times n)$ Mahalanobis technology similarity matrix with elements $(f_{ij}^{\mathrm{M}})_{1\leq i,j\leq n}$ is defined as

$$\mathbf{F}^{\mathrm{M}} = \tilde{\mathbf{T}}^\top \mathbf{\Omega} \tilde{\mathbf{T}}. \tag{H.38}$$

Figure I.12 shows the average patent proximity across collaborating firms using the Jaffe metric

f_{ij}^{J} of Equation (H.37) or the Mahalanobis metric f_{ij}^{M} of Equation (H.38). Both are monotonic increasing over almost all years of observations. This suggests that R&D collaborating firms tend to become more similar over time.

I. Numerical Algorithm for Computing the Optimal Subsidies

The Nash equilibrium output levels, $\mathbf{q} \in [0, \bar{q}]^n$, in the presence of the subsidy, $\mathbf{s} \in [0, \bar{s}]^n$, satisfy

$$
\begin{aligned}
q_i = 0, \text{ if } & -\mu_i + q_i + \rho \sum_{j=1}^n b_{ij} q_j - \varphi \sum_{j=1}^n a_{ij} q_j - s_i - \varphi \sum_{j=1}^n a_{ij} s_j > 0, \\
q_i = \mu_i - \rho \sum_{j \neq i} b_{ij} q_j + \varphi \sum_{j=1}^n a_{ij} q_j + s_i + \varphi \sum_{j=1}^n a_{ij} s_j, \text{ if } & -\mu_i + q_i + \rho \sum_{j=1}^n b_{ij} q_j - \varphi \sum_{j=1}^n a_{ij} q_j - s_i - \varphi \sum_{j=1}^n a_{ij} s_j = 0, \\
q_i = \bar{q}, \text{ if } & -\mu_i + q_i + \rho \sum_{j=1}^n b_{ij} q_j - \varphi \sum_{j=1}^n a_{ij} q_j - s_i - \varphi \sum_{j=1}^n a_{ij} s_j < 0.
\end{aligned}
\tag{I.39}
$$

The problem of finding a vector $\mathbf{q}$ such that the conditions in (I.39) hold is known as the bounded linear complementarity problem [cf. Byong-Hun, 1983].

The bounded linear complementarity problem (LCP) of Equation (I.39) is equivalent to the Kuhn-Tucker optimality conditions of the following quadratic programming (QP) problem with box constraints

$$
\min_{\mathbf{q} \in [0, \bar{q}]^n} \left\{ -\boldsymbol{\nu}(\mathbf{s})^\top \mathbf{q} + \frac{1}{2} \mathbf{q}^\top \left(\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A} \right) \mathbf{q} \right\}, \tag{I.40}
$$

where $\boldsymbol{\nu}(\mathbf{s}) \equiv \boldsymbol{\mu} + (\mathbf{I} + \varphi \mathbf{A})\mathbf{s}$. Moreover, net welfare is given by

$$
\overline{W}(G, \mathbf{s}) = \sum_{i=1}^n \left(\frac{q_i^2}{2} + \pi_i - s_i e_i \right) = \boldsymbol{\mu}^\top \mathbf{q} - \mathbf{q}^\top \left(\frac{\rho}{2} \mathbf{B} - \varphi \mathbf{A} \right) \mathbf{q} + \varphi \mathbf{q}^\top \mathbf{A} \mathbf{s} - \frac{1}{2} \mathbf{s}^\top \mathbf{A} \mathbf{s}.
$$

Finding the optimal subsidy program $\mathbf{s}^* \in [0, \bar{s}]^n$ is then equivalent to solving the following *bilevel optimization problem* [cf. Bard, 2013]

$$
\begin{aligned}
\max_{\mathbf{s} \in [0, \bar{s}]^n} \quad & \overline{W}(G, \mathbf{s}) = \boldsymbol{\mu}^\top \mathbf{q}^*(\mathbf{s}) - \mathbf{q}^*(\mathbf{s})^\top \left(\frac{\rho}{2} \mathbf{B} - \varphi \mathbf{A} \right) \mathbf{q}^*(\mathbf{s}) + \varphi \mathbf{q}^*(\mathbf{s})^\top \mathbf{A} \mathbf{s} - \frac{1}{2} \mathbf{s}^\top \mathbf{A} \mathbf{s} \\
\text{s.t.} \quad & \mathbf{q}^*(\mathbf{s}) = \min_{\mathbf{q} \in [0, \bar{q}]^n} \left\{ -\boldsymbol{\nu}(\mathbf{s})^\top \mathbf{q} + \frac{1}{2} \mathbf{q}^\top \left(\mathbf{I} + \rho \mathbf{B} - \varphi \mathbf{A} \right) \mathbf{q} \right\}.
\end{aligned}
\tag{I.41}
$$

The bilevel optimization problem of Equation (I.41) can be implemented in `MATLAB` following a two-stage procedure. First, one computes the Nash equilibrium output levels $\mathbf{q}^*(\mathbf{s})$ as a function of the subsidies $\mathbf{s}$ by solving a quadratic programming problem, for example using the `MATLAB` function `quadprog`, or the nonconvex quadratic programming problem solver with box constraints `QuadProgBB` introduced in Chen and Burer [2012].[19] Second, one can apply an optimization routine to this function calculating the subsidies which maximize net welfare $\overline{W}(G, \mathbf{s})$, for example using `MATLAB`'s function `fminsearch` (which uses a Nelder-Mead algorithm).

This bilevel optimization problem can be formulated more efficiently as a *mathematical pro-*

[19]However, in the data that we have analyzed in this paper the quadratic programming subproblem of determining the Nash equilibrium outptut levels always turned out to be convex, and therefore we always obtained a unique Nash equilibrium.

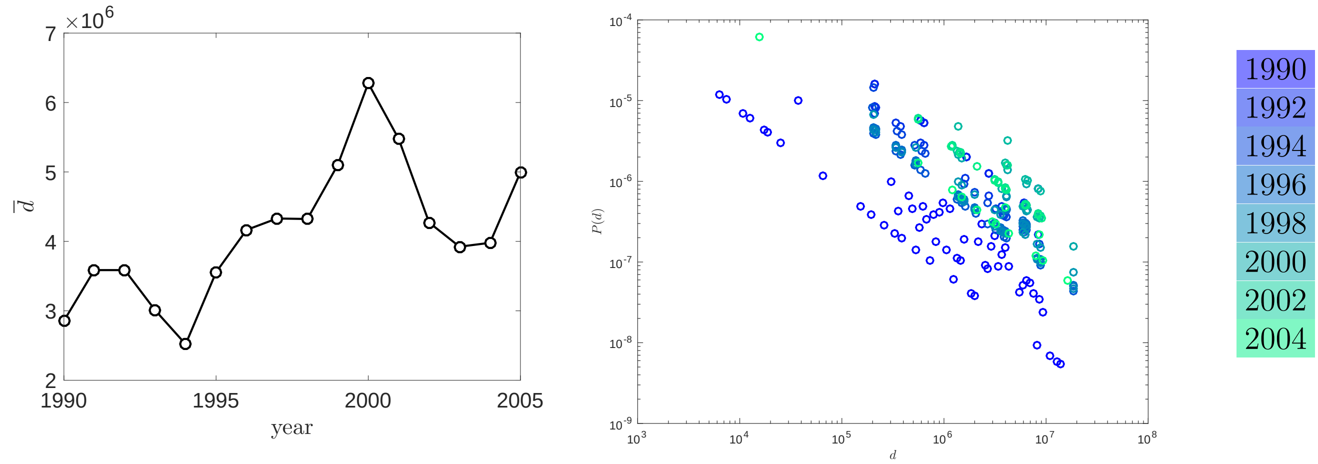


Figure I.11: The mean distance, $\overline{d}$, and the distance distribution, $P(d)$, across collaborating firms in the combined CATI-SDC database.

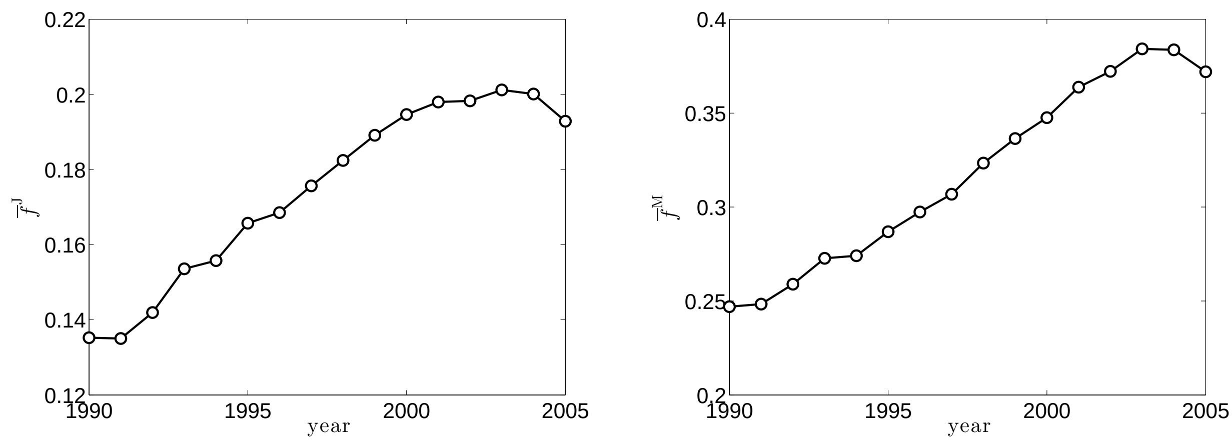


Figure I.12: The mean patent proximity across collaborating firms using the Jaffe metric f_{ij}^{J} of Equation (H.37) or the Mahalanobis metric f_{ij}^{M} of Equation (H.38).

gramming problem with equilibrium constraints (MPEC; see also Luo et al. [1996]). While in the above procedure the `quadprog` algorithm solves the quadratic problem with high accuracy for each iteration of the `fminsearch` routine, MPEC circumvents this problem by treating the equilibrium conditions as constraints. This method has recently been proposed to structural estimation problems following the seminal paper by Su and Judd [2012]. The MPEC approach can be implemented in `MATLAB` using a constrained optimization solver such as `fmincon`.[20]

Finally, to initialize the optimiziation algorithm we can use the theoretical optimal subsidies from Propositions 2 and 3, by setting the output levels of the firms which would produce at negative quantities under these policies to zero (if there are any), and then apply a bounded quadratic programming algorithm to determine the Nash equilibrium quantities under these subsidy policies.

[20] Su and Judd [2012] further recommend to use the `KNITRO` version of `MATLAB`'s `fmincon` function to improve speed and accuracy.

Table J.3: Parameter estimates from a panel regression of Equation (26) with both firm and time fixed effects. The duration of an alliance ranges from 3 to 7 years. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

alliance duration	3 years	4 years	5 years	6 years	7 years
φ	0.0131**	0.0119**	0.0106**	0.0089*	0.0077*
	(0.0055)	(0.0053)	(0.0051)	(0.0047)	(0.0044)
ρ	0.0188***	0.0188***	0.0189***	0.0189***	0.0189***
	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
β	0.0027***	0.0027***	0.0027***	0.0027***	0.0027***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
# firms	1186	1186	1186	1186	1186
# observations	16924	16924	16924	16924	16924
Cragg-Donald Wald F stat.	7064.104	7071.522	7078.856	7084.185	7096.780
firm fixed effects	yes	yes	yes	yes	yes
time fixed effects	yes	yes	yes	yes	yes

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

J. Additional Robustness Checks

In the following sections we perform some additional robustness checks related to the duration of an alliance (Appendix J.1), heterogeneous competition and spillover effects across different sectors (Appendix J.2), input-supplier effects (Appendix J.3), alternative specifications of the competition matrix based on the product mix of the firms (Appendix J.4) and the impact of missing data on our estimates (Appendix J.5).

J.1. Time Span of Alliances

In Section 6.3, we assume the duration of a R&D alliance is 5 years. Here, we analyze the impact of different durations of an R&D alliance on the estimated spillover effect. The estimation results for alliance durations ranging from 3 to 7 years are shown in Table J.3. We find that the estimates are robust over the different durations considered.

However, our assumption that the duration is the same for all alliances may seem restrictive. As a further robustness check, we randomly draw a life span for each alliance from an exponential distribution with the mean ranging from 3 to 7 years. The estimation results are shown in Table J.4. We find that the estimates are still robust.

J.2. Heterogeneous Spillover and Competition Effects

In keeping with the literature such as Bloom et al. [2013], the spillover effect and competition coefficients are assumed to be identical across markets in Equation (25). Here, we conduct a robustness analysis using two major divisions in our data, namely the manufacturing and services sectors that cover, respectively, 76.8% and 19.3% firms in our sample, in order to re-estimate Equation (25). The estimation results are reported in Table J.5. The estimated spillover and competition parameters for these two sectors are largely the same, supporting the assumption of homogeneous spillover and competition effects as in the benchmark specifciation.

Table J.4: Parameter estimates from a panel regression of Equation (26) with both firm and time fixed effects. The duration of an alliance follows an exponential distribution with the mean ranging from 3 to 7 years. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

average alliance duration	3 years	4 years	5 years	6 years	7 years
φ	0.0106**	0.0139***	0.0113**	0.0140**	0.0074
	(0.0046)	(0.0046)	(0.0052)	(0.0057)	(0.0048)
ρ	0.0186***	0.0188***	0.0187***	0.0188***	0.0187***
	(0.0028)	(0.0028)	(0.0028)	(0.0028)	(0.0028)
β	0.0027***	0.0027***	0.0027***	0.0027***	0.0027***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
# firms	1186	1186	1186	1186	1186
# observations	16924	16924	16924	16924	16924
Cragg-Donald Wald F stat.	7046.331	7063.207	7081.713	7080.294	7045.043
firm fixed effects	yes	yes	yes	yes	yes
time fixed effects	yes	yes	yes	yes	yes

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

Table J.5: Parameter estimates from a panel regression of Equation (25) for the manufacturing and services sectors with both firm and time fixed effects. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

	Manufacturing		Services	
φ	0.0111*	(0.0061)	0.0099**	(0.0040)
ρ	0.0178***	(0.0030)	0.0164***	(0.0040)
β	0.0027***	(0.0002)	0.0027***	(0.0002)
# firms	911		229	
# observations	14352		2073	
Cragg-Donald Wald F stat.	6817.740		2196.649	
firm fixed effects	yes		yes	
time fixed effects	yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

J.3. Input-output Linkages

If a firm is an input supplier of another firm, then their output levels are likely to be correlated. Here, we conduct a robustness analysis by directly controlling for potential input-supplier effects. More specifically, we estimate an extended version of Equation (25) given by

$$q_{it} = \varphi \sum_{j=1}^{n} a_{ij,t} q_{jt} + \lambda \sum_{j=1}^{n} c_{ij,t} q_{jt} - \rho \sum_{j=1}^{n} b_{ij} q_{jt} + \beta x_{it} + \eta_i + \kappa_t + \epsilon_{it}, \tag{J.42}$$

where $c_{ij,t}$ are indicator variables such that $c_{ij,t} = 1$ if firm j is an input supplier of firm i in period t and $c_{ij,t} = 0$ otherwise.

We obtain information about firms' buyer-supplier relationships from two data sources. The first is the Compustat Segments database [cf. e.g. Atalay et al., 2011; Barrot and Sauvagnat, 2016]. Compustat Segments provides business details, product information and customer data for over 70% of the companies in the Compustat North American database, with firms coverage starting in the year 1976. However, this dataset suffers from a truncation bias as firms only report customers which make up more than 10% of their total sales. We therefore use as a second datasource the Capital IQ Business Relationships database [Barrot and Sauvagnat, 2016; Lim, 2016; Mizuno et al., 2014]. The Capital IQ data includes any customers/suppliers that are mentioned in the firms' annual reports, news, websites surveys etc, with firms coverage starting in the year 1990.[21] We then merged these two datasources to obtain a more complete picture of the potential buyer-supplier linkages between the firms in our R&D network.[22] Aggregated over all years we obtained a total of $2,573$ buyer-supplier relationships for the firms matched with our R&D network dataset.

As the data on the input-output linkages is only available in more recent years, the estimation is based on years from 1980 to 2006. The estimation results are reported in Table J.6. We find that, after controlling for input-supplier effects, the spillover and competition effects remain statistically significant with the expected signs.

Furthermore, having a firm as an input supplier might increase the probability to form an R&D alliance. We use the information on input-output linkages as an additional predictor in the link formation regression of Equation (29), and use the predicted link-formation probability to construct IVs as explained in Section 6.2.4. The estimation results of the link formation regression Equations (29) and (25) are reported in Tables J.7 and J.8, respectively. As expected, having an input-output linkage increases the likelihood of forming an R&D collaboration. Moreover, controlling for input-output linkages gives qualitatively the same result as in the baseline specification.

J.4. Alternative Specifications of the Competition Matrix

In the empirical model estimated in Section 6.3, the entries of the competition matrix, $\mathbf{B} = [b_{ij}]$, are specified as indicator variables such that $b_{ij} = 1$ if firms i and j are the same industry (measured by the industry SIC codes at the 4-digit level) and $b_{ij} = 0$ otherwise. Here, we consider three alternative specifications of the competition matrix based on the primary and secondary industry classification codes that can be found in the Compustat Segments and

[21] About 23.37% of the observations come with information about the date of the relationship in Capital IQ. This gives a total of $38,513$ potential links.

[22] Note that it is possible to merge the firms in the Compustat Segments database with the Capital IQ database using common firm identifiers (there exists a correspondence table for Capital IQ firm id's with Compustat's gvkeys).

Table J.6: Parameter estimates from a panel regression of Equation (J.42) with both firm and time fixed effects. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1980–2006.

φ	0.0126***	(0.0048)
λ	0.6933***	(0.1172)
ρ	0.0146***	(0.0021)
β	0.0022***	(0.0002)
# firms	1251	
# observations	15463	
Cragg-Donald Wald F stat.	2668.988	
firm fixed effects	yes	
time fixed effects	yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

Table J.7: Link formation regression results with input-output linkage information. Technological similarity, f_{ij}, is measured using either the Jaffe or the Mahalanobis patent similarity measures. The dependent variable $a_{ij,t}$ indicates if an R&D alliance exists between firms i and j at time t. The estimation is based on the observed alliances in the years 1980–2006.

technological similarity	Jaffe	Mahalanobis
Past collaboration	0.5715***	0.5682***
	(0.0144)	(0.0143)
Past common collaborator	0.1753***	0.1779***
	(0.0216)	(0.0214)
Input supplier	4.0606***	4.0215***
	(0.1370)	(0.1374)
$f_{ij,t-s-1}$	10.4884***	4.3003***
	(0.6798)	(0.3212)
$f^2_{ij,t-s-1}$	-15.5768***	-2.4457***
	(1.6995)	(0.4379)
$city_{ij}$	1.0794***	1.0922***
	(0.1030)	(0.1030)
$market_{ij}$	0.9417***	0.9501***
	(0.0421)	(0.0419)
# observations	2,776,488	2,776,488
McFadden's R^2	0.0856	0.0854

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

Table J.8: Parameter estimates from a panel regression of Equation (26) with endogenous R&D alliance matrix. The IVs are based on the predicted links from the logistic regression reported in Table J.7, where technological similarity is measured using either the Jaffe or the Mahalanobis patent similarity measures. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1980–2006.

technological similarity	Jaffe		Mahalanobis	
φ	0.0317**	(0.0148)	0.0323**	(0.0148)
ρ	0.0200***	(0.0028)	0.0201***	(0.0028)
β	0.0026***	(0.0002)	0.0026***	(0.0002)
# firms	1245		1245	
# observations	15296		15296	
Cragg-Donald Wald F stat.	191.866		192.407	
firm fixed effects	yes		yes	
time fixed effects	yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

Orbis databases [cf. Bloom et al., 2013],[23] or the Hoberg-Phillips product similarity measures [cf. Hoberg and Phillips , 2016].[24]

The estimation results of Equation (26) with alternative specifications of the competition matrix are reported in Table J.9. The estimated technology spillover effect is positively significant, with the magnitude similar to that reported in Table 2, suggesting that the estimation of the spillover effect is robust with respect to different specifications of the competition matrix. The magnitude of the product rivalry effect reported in Table J.9, on the other hand, is more difficult to compare with that reported in Table 2, as they are based on different competition matrices. Nevertheless, the estimated product rivalry effect with alternative specifications of the competition matrix remains statistically significant with the expected sign.

J.5. Sampled Networks

The balance sheet data we used for the empirical analysis covers only publicly listed firms. It is now well known that the estimation with sampled network data could lead to biased estimates [see, e.g. Chandrasekhar and Lewis, 2011]. To investigate the direction and magnitude of the bias due to the sampled network data, we conduct a limited simulation experiment. In the experiment, we randomly drop 10%, 20%, and 30% of the firms (and the R&D alliances associated with the dropped firms) in our data (corresponding to the sampling rate of 90%, 80%, and 70%). For each sampling rate, we randomly draw 500 subsamples and re-estimate Equation (26) for each subsample. We report the empirical mean and standard deviation of the estimates for each sampling rate in Table J.10. As the sampling rate reduces, the standard deviation of the estimates increases while the mean remains roughly the same. This simulation result alleviates the concern on the estimation bias due to sampling (i.e. missing data).

[23] Our definition of the pairwise competition intensity is calculated as the Jaffe similarity score of the combined vectors of primary and secondary industry codes (see also Footnote 28), and follows the product market proximity index suggested in Bloom et al. [2013].

[24] The Hoberg-Phillips product similarity measures are based on firm pairwise similarity scores from text analysis of the firms' 10K product descriptions. See Hoberg and Phillips [2016] for further details and explanation.

Table J.9: Parameter estimates from a panel regression of Equation (26) with both firm and time fixed effects. The competition matrix is based on the Compustat Segments, Orbis or Hoberg-Phillips industry/product similarity measures. The dependent variable is output obtained from deflated sales. Standard errors (in parentheses) are robust to arbitrary heteroskedasticity and allow for first-order serial correlation using the Newey-West procedure. The estimation is based on the observed alliances in the years 1967–2006.

competition matrix	Compustat		Orbis		Hoberg-Phillips	
φ	0.0089*	(0.0049)	0.0110**	(0.0051)	0.0096**	(0.0048)
ρ	0.0526***	(0.0088)	0.0438***	(0.0077)	0.4753***	(0.0761)
β	0.0029***	(0.0002)	0.0027***	(0.0002)	0.0026***	(0.0002)
# firms	1199		1199		1199	
# observations	17433		17433		17433	
Cragg-Donald Wald F stat.	3638.903		3079.453		1.1×10^4	
firm fixed effects	yes		yes		yes	
time fixed effects	yes		yes		yes	

*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

Table J.10: Parameter estimates from a panel regression of Equation (26) with both firm and time fixed effects using a random subsample of the firms under different sampling rates. The dependent variable is output obtained from deflated sales. The empirical mean and standard deviation (in parentheses) of the estimates from 500 random subsamples are reported. The estimation is based on the observed alliances in the years 1967–2006.

sampling rate	90%	80%	70%
φ	0.0109	0.0114	0.0113
	(0.0035)	(0.0059)	(0.0084)
ρ	0.0185	0.0187	0.0191
	(0.0021)	(0.0031)	(0.0043)
β	0.0027	0.0027	0.0027
	(0.0001)	(0.0002)	(0.0002)
firm fixed effects	yes	yes	yes
time fixed effects	yes	yes	yes

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R. and P. Howitt (2005). Competition and innovation: An inverted-U relationship. *Quarterly Journal of Economics* 120(2), 701–728.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica* 60(2), 323–351.

Aghion, P., Akcigit, U., and Howitt, P. (2014). *Handbook of Economic Growth*, Volume 2B, chapter What Do We Learn From Schumpeterian Growth Theory?, pages 515–563.

Atalay, E., Hortacsu, A., Roberts, J. and C. Syverson (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13), 5199.

Ballester, C., Calvó-Armengol, A. and Y. Zenou (2006). Who's who in networks. wanted: The key player. *Econometrica* 74(5), 1403–1417.

Ballester, C. and M. Vorsatz (2013). Random walk–based segregation measures. *Review of Economics and Statistics* 96(3), 383–401.

Bard, J. F. (2013). *Practical Bilevel Optimization: Algorithms and Applications*. Berlin: Springer Science.

Barrot, J.-N. and Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics*, 131(3):1543–1592.

Belhaj, M., Bramoullé, Y. and F. Deroïan (2014). Network games under strategic complementarities. *Games and Economic Behavior* 88, 310–319.

Bell, F.K. (1992). A note on the irregularity of graphs. *Linear Algebra and its Applications* 161, 45–54.

Bena, J., Fons-Rosen, C. and P. Ondko (2008). Zephyr: Ownership changes database. London School of Economics, Working Paper.

Bloom, N., Schankerman, M. and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica* 81(4), 1347–1393.

Bollaert, H., Delanghe, M., 2015. Securities data company and Zephyr, data sources for M&A research. *Journal of Corporate Finance* 33, 85–100.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology* 92(5), 1170–1182.

Bramoullé, Y., Kranton, R. and M. D'amours (2014). Strategic interaction and networks. *American Economic Review* 104 (3), 898–930

Brualdi, R. A., Solheid, Ernie, S., 1986. On the spectral radius of connected graphs. Publications de l' Institut de Mathématique 53, 45–54.

Byong-Hun, A. (1983). Iterative methods for linear complementarity problems with upper-bounds on primary variables. *Mathematical Programming* 26(3), 295–315.

Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies* 76, 1239–1267.

Chandrasekhar, A. and R. Lewis (2011). Econometrics of sampled networks. Unpublished manuscript, Standford University.

Chen, J. and S. Burer (2012). Globally solving nonconvex quadratic programming problems via completely positive programming. *Mathematical Programming Computation* 4(1), 33–52.

Copeland, A. and D. Fixler (2012). Measuring the price of research and development output. *Review of Income and Wealth* 58(1), 166–182.

Cvetkovic, D., Doob, M. and H. Sachs (1995). *Spectra of Graphs: Theory and Applications*. Johann Ambrosius Barth.

Cvetkovic, D. and P. Rowlinson (1990). The largest eigenvalue of a graph: A survey. *Linear and Multinilear Algebra* 28(1), 3–33.

Dai, R. (2012). International accounting databases on wrds: Comparative analysis. Working paper, Wharton Research Data Services, University of Pennsylvania.

Debreu, G. and , I.N. Herstein (1953). Nonnegative square matrices. *Econometrica* 21(4), 597–607.

Dell, M. (2009). GIS analysis for applied economists. Unpublished manuscript, MIT Department of Economics.

Freeman, L., 1979. Centrality in social networks: Conceptual clarification. *Social Networks* 1(3), 215–239.

Gal, P. N., 2013. Measuring total factor productivity at the firm level using OECD-ORBIS. OECD Working Paper, ECO/WKP(2013)41.

Goyal, S. and J.L. Moraga-Gonzalez (2001). R&D networks. *RAND Journal of Economics* 32 (4), 686–707.

Grossman, G., Helpman, E., 1991. Quality ladders in the theory of growth. *Review of Economic Studies* 58(1), 43–61.

Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research Policy* 31(4), 477–492.

Hall, B. H., Jaffe, A. B., Trajtenberg, M., 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper No. 8498.

Hanaki, N., Nakajima, R., Ogura, Y., 2010. The dynamics of R&D network in the IT industry. *Research Policy* 39(3), 386–399.

Hirschman, A. O., 1964. The paternity of an index. *American Economic Review*, 761–762.

Hoberg, Gerard and Phillips, Gordon (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124(5), 1423–1465.

Horn, R. A., Johnson, C. R., 1990. *Matrix Analysis*. Cambridge University Press.

Huyghebaert, N., Luypaert, M., 2010. Antecedents of growth through mergers and acquisitions: Empirical results from belgium. *Journal of Business Research* 63(4), 392–403.

Jaffe, A.B. and M. Trajtenberg (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge: MIT Press.

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

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18(1), 39–43.

Khalil, H. K. (2002). *Nonlinear Systems*. Prentice Hall.

Kitsak, M., Riccaboni, M., Havlin, S., Pammolli, F. and H. Stanley (2010). Scale-free models for the structure of business firm networks. *Physical Review E* 81, 036117.

Kogut, B. (1988). Joint ventures: Theoretical and empirical perspectives. *Strategic Management Journal* 9(4), 319–332.

König, M., Tessone, C. and Y. Zenou (2014). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics* 9, 695–752.

König, M. D. (2016). The formation of networks with local spillovers and limited observability. *Theoretical Economics*, 11, 813–863.

Lee, G., Tam, N. and N. Yen (2005). *Quadratic Programming and Affine Variational Inequalities: A Qualitative Study*. Berlin: Springer Verlag.

Leicht, E.A., Holme, P. and M.E.J. Newman (2006). Vertex similarity in networks. *Physical Review E* 73(2), 026120.

Lim, K. (2016). Firm to firm trade in sticky production networks. *Mimeo, Princeton University*.

Luo, Z.-Q., Pang, J.-S., Ralph, D., 1996. *Mathematical programs with equilibrium constraints*. Cambridge University Press.

Lychagin, S., and Pinkse, J. and Slade, M. E. and Van Reenen, J., 2010. Spillovers in space: does geography matter? National Bureau of Economic Research Working Paper No. w16188.

Mahadev, N., Peled, U., 1995. *Threshold Graphs and Related Topics*. North Holland.

Manshadi, V., Johari, R., 2010. Supermodular network games. In: 47thAnnual Allerton Conference on Communication, Control, and Computing, 2009. IEEE, pp. 1369–1376.

McManus, O., Blatz, A. and K. Magleby (1987). Sampling, log binning, fitting, and plotting durations of open and shut intervals from single channels and the effects of noise. *Pflügers Archiv* 410 (4-5), 530–553.

Mizuno, T., Ohnishi, T., and Watanabe, T. (2014). The structure of global inter-firm networks. In *Social Informatics*, pages 334–338. Springer.

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

Nocedal, J., Wright, S., 2006. Numerical Optimization. Springer Verlag.

Papadopoulos, A., 2012. Sources of data for international business research: Availabilities and implications for researchers. In: Academy of Management Proceedings. Vol. 2012. Academy of Management, pp. 1–1.

Peled, U. N., Petreschi, R., Sterbini, A., 1999. (n, e)-graphs with maximum sum of squares of degrees. Journal of Graph Theory 31 (4), 283–295.

Rencher, A. C., Christensen, W. F., 2012. *Methods of multivariate analysis*. John Wiley & Sons.

Rosenkopf, L., Schilling, M., 2007. Comparing alliance network structure across industries: Observations and explanations. Strategic Entrepreneurship Journal 1, 191–209.

Samuelson, P., 1942. A method of determining explicitly the coefficients of the characteristic equation. *Annals of Mathematical Statistics*, 424–429.

Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal* 30(3), 233–260.

Singh, N., Vives, X., 1984. Price and quantity competition in a differentiated duopoly. *RAND Journal of Economics* 15(4), 546–554.

Su, C.-L., Judd, K. L., 2012. Constrained optimization approaches to estimation of structural models. *Econometrica* 80(5), 2213–2230.

Tirole, J. (1988). *The Theory of Industrial Organization.* Camridge: MIT Press.

Trajtenberg, M., Shiff, G. and R. Melamed (2009). The "names game": Harnessing inventors, patent data for economic research. *Annals of Economics and Statistics*, 79–108.

Van Mieghem, P. (2011). *Graph Spectra for Complex Networks.* Cambridge: Cambridge University Press.

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review* 23(176), 88–93.

Wasserman, S. and K. Faust (1994). *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press.