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]).

□

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

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

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* $\lambda_{\text{PF}}(G)$ is the *largest real eigenvalue* of $\mathbf{A}$ associated with G, i.e. all eigenvalues λ_i of $\mathbf{A}$ satisfy $|\lambda_i| \leq \lambda_{\text{PF}}(G)$ for $i = 1, \ldots, n$ and there exists an associated nonnegative eigenvector $\mathbf{v}_{\text{PF}} \geq 0$ such that $\mathbf{A}\mathbf{v}_{\text{PF}} = \lambda_{\text{PF}}(G)\mathbf{v}_{\text{PF}}$. For a connected graph G the adjacency matrix $\mathbf{A}$ has a unique largest real eigenvalue $\lambda_{\max}(G)$ and a positive associated eigenvector $\mathbf{v}_{\text{PF}} > 0$. 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 $\mathbf{v}_{\text{PF}}$ 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 $\text{tr}\left(\mathbf{A}^k\right) = \sum_{i=1}^n \left(\mathbf{A}^k\right)_{ii} = \sum_{i=1}^n \lambda_i^k$. We further have that $\text{tr}\left(\mathbf{A}\right) = 0$, $\text{tr}\left(\mathbf{A}^2\right)$ gives twice the number of links in G and $\text{tr}\left(\mathbf{A}^3\right)$ 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 $\text{cor}_i \leq d_i$. 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 $\boldsymbol{\Lambda} \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ 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

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

which gives

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

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 $c_{\mathbf{A}}(\phi) \equiv \det(\mathbf{A} - \phi \mathbf{I}_n)$ 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].

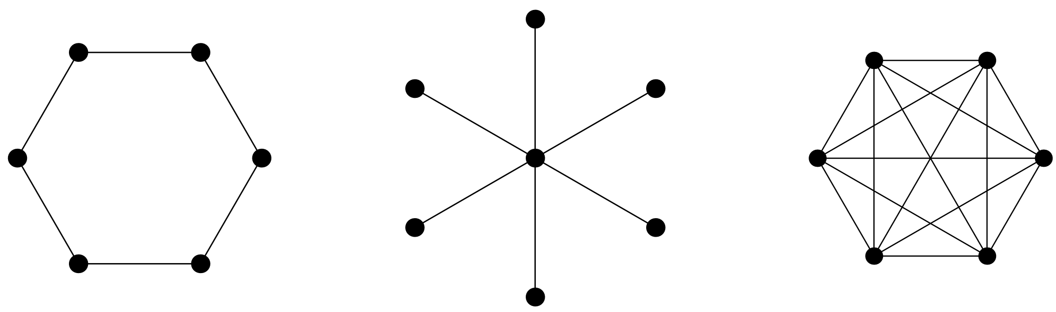

Figure B.1: Illustration of a cycle C_6, a star $K_{1,6}$ and a complete graph, K_6.

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 $\mathbf{M} = \phi\mathbf{AM} + \mathbf{I}$. Rearranging yields $\mathbf{M} = (\mathbf{I}-\phi\mathbf{A})^{-1} = \sum_{k=0}^\infty \phi^k\mathbf{A}^k$, assuming that $\phi < 1/\lambda_{\text{PF}}$. 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 $\text{cor}_i = k$ 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 $\lambda_{\min}\left(\mathbf{B}\right) > -1/\rho$. Since $\lambda_{\min}\left(\mathbf{B}\right) = -1$, 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

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

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.

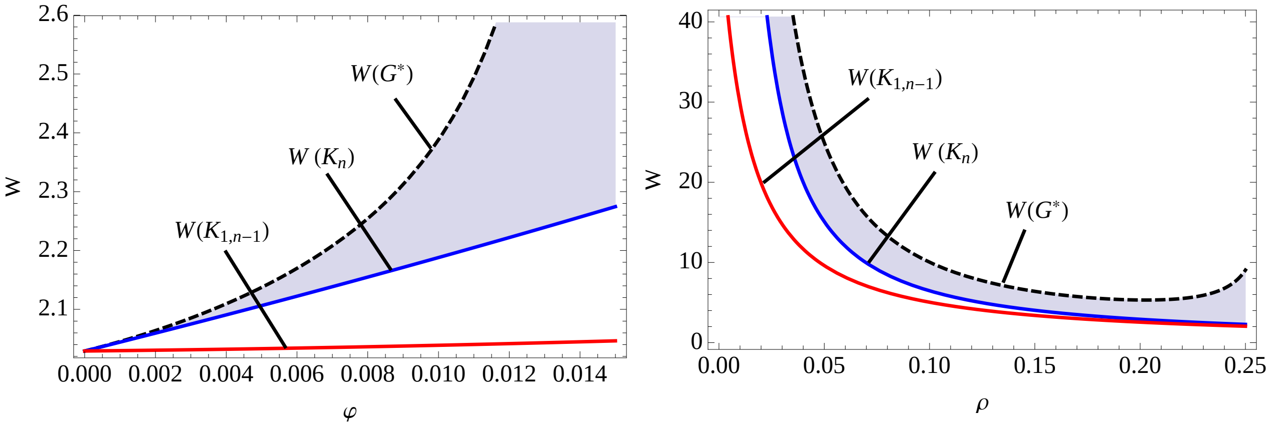

Figure G.2: (Left panel) The upper and lower bounds of Equation (G.31) with $n = 50$, $\rho = 0.25$ for varying values of φ. (Right panel) The upper and lower bounds of Equation (G.31) with $n = 50$, $\varphi = 0.015$ for varying values of ρ.

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]

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

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 $K_n = \operatorname{argmax}_{G\in\mathcal{G}^n} W(G)$.

(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 $\lambda_1 \geq \left(\frac{N_k(G)}{n}\right)^{\frac{1}{k}}$ [cf. Van Mieghem, 2011, p. 47], so that $\frac{d}{d\varphi}(\varphi N_G(\varphi)) = \sum_{k=0}^\infty \varphi^k(k+1)N_k(\varphi) \leq n\sum_{k=0}^\infty(\lambda_1\varphi)^k(k+1) = n\sum_{k=0}^\infty(\lambda_1\varphi)^k + n\sum_{k=0}^\infty k(\lambda_1\varphi)^k = n\left(\frac{1}{1+\varphi\lambda_1}+\frac{\varphi\lambda_1}{(1+\varphi\lambda_1)^2}\right) = \frac{n}{(1+\varphi\lambda_1)^2}$.

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

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

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

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

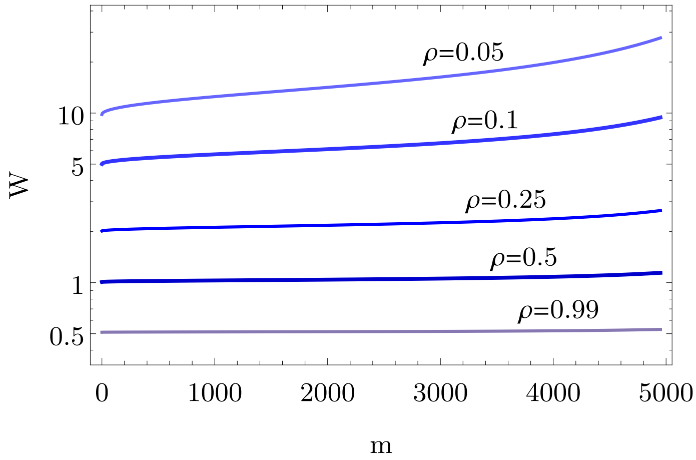

Figure G.3: The RHS in Equation (G.35) with varying values of $m \in \{0, 1, \ldots, n(n-1)/2\}$ for $n = 100$, $\varphi = 0.9(1-\rho)/n$ and $\rho \in \{0.05, 0.1, 0.25, 0.5, 0.99\}$.

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

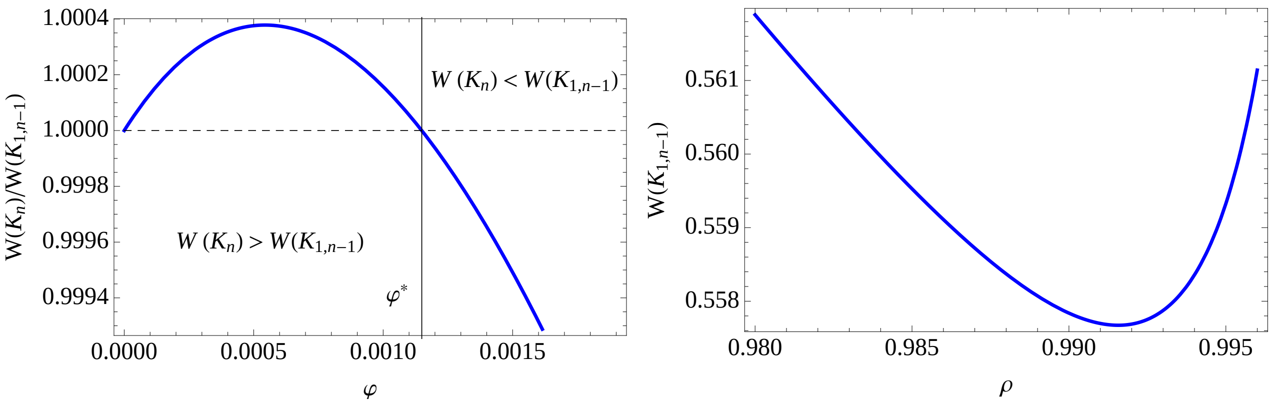

Figure G.4: (Left panel). The ratio of welfare in the complete graph, K_n, and the star, $K_{1,n-1}$, for $n = 10$, $\rho = 0.981$ and varying values of φ ($< ((1-\rho)/\lambda_{\max}(K_n) = 0.002$) (Right panel) Welfare in the star, $K_{1,n-1}$, with varying values of ρ for $n = 10$ and $\varphi = 0.001$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

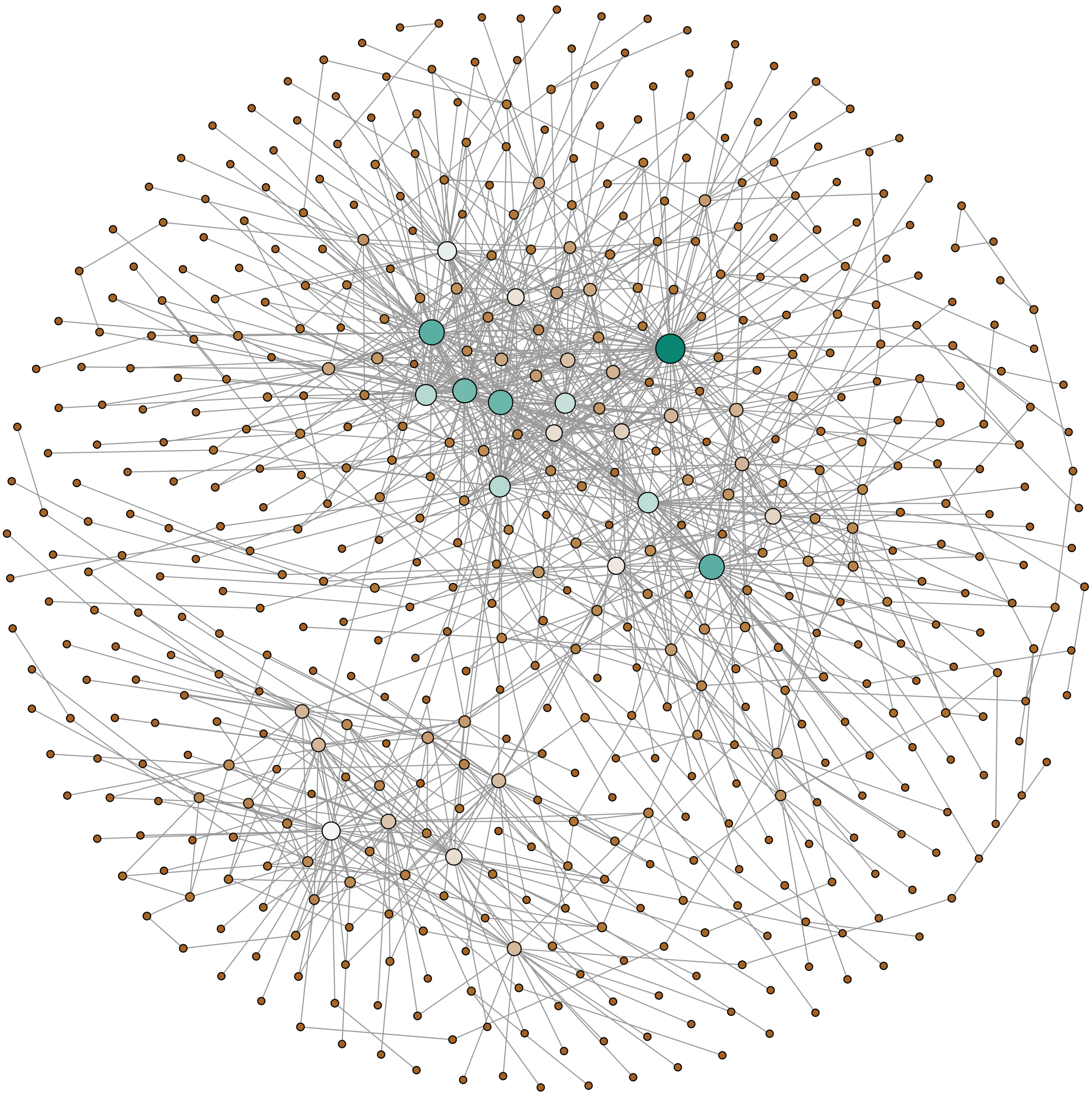

Figure H.5: The largest connected component of the R&D collaboration network with all links accumulated until the year 2005. The nodes' colors indicate sectors according to 4-digit SIC codes while the nodes' sizes indicate the number of collaborations of a firm.

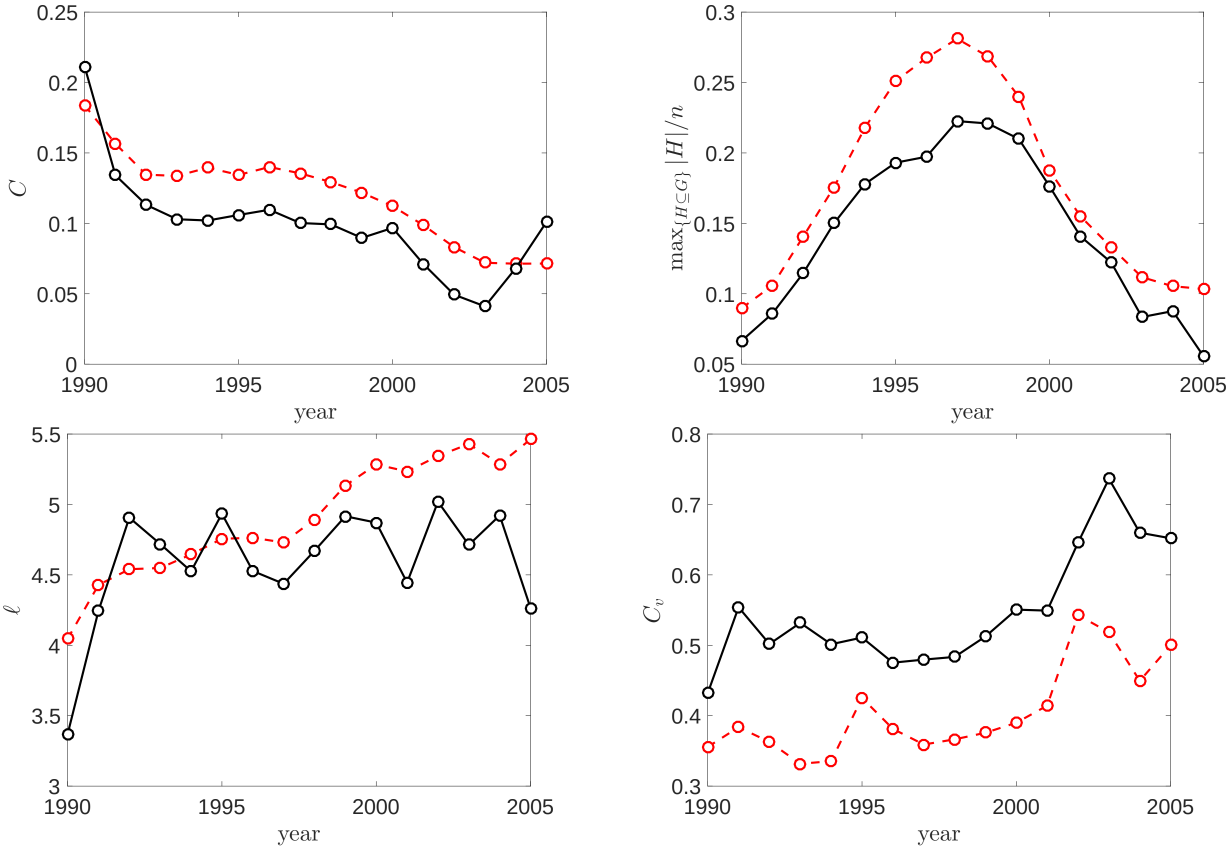

Figure H.6: The average clustering coefficient, C, the relative size of the largest connected component, $\max_{\{H \subseteq G\}} |H|/n$, the average path length, ℓ, and the eigenvector centralization C_v (relative to a star network of the same size) over the years 1990 to 2005 (see Appendix B.1). Dashed 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

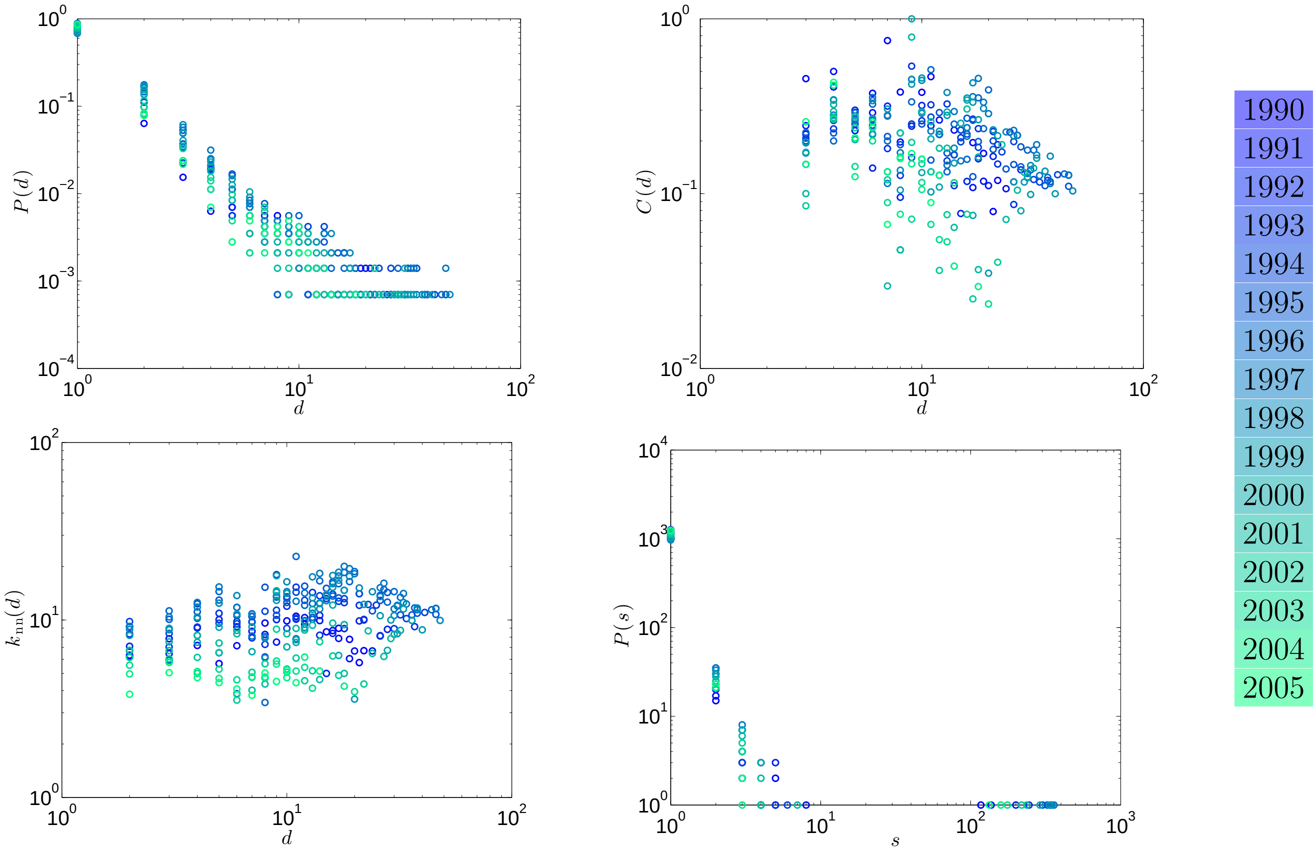

Figure H.7: The degree distribution, $P(d)$, the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, the clustering degree distribution, $C(d)$, and the component size distribution, $P(s)$.

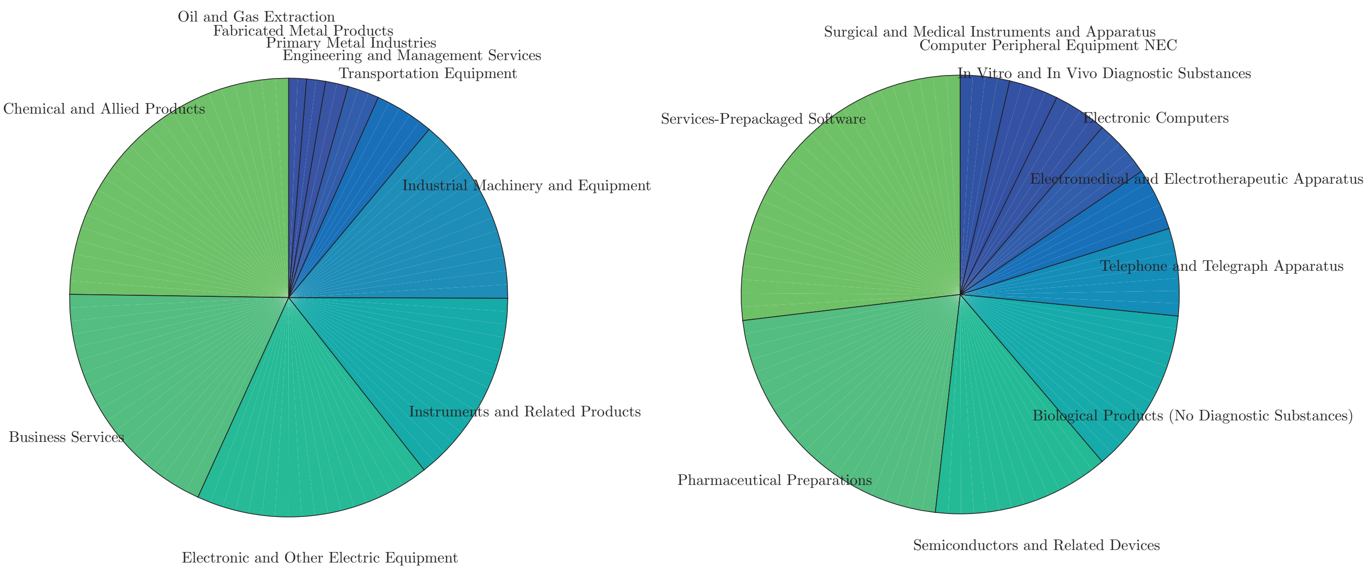

Figure H.8: The shares of the ten largest sectors at the 2-digit (left panel) and 4-digit (right panel) SIC levels.

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).

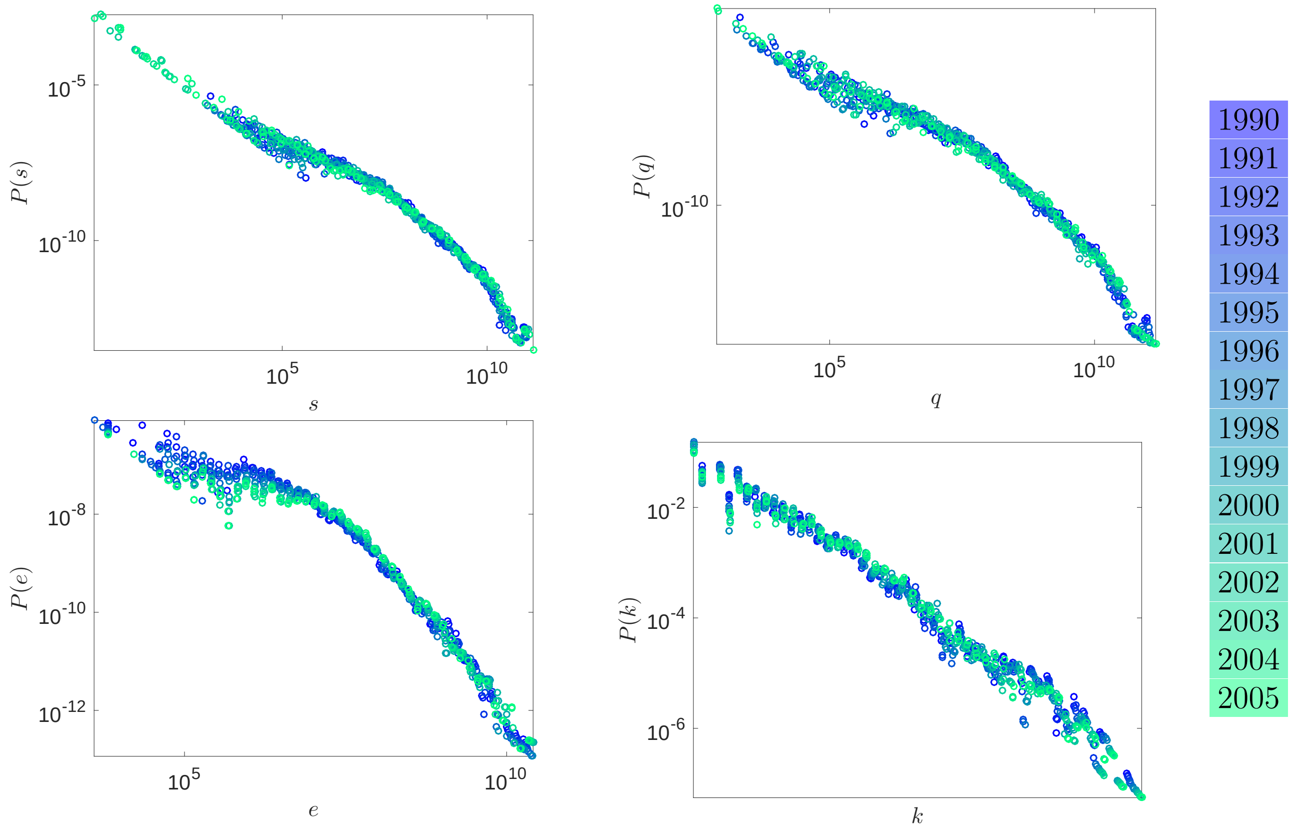

Figure H.9: The sales distribution, $P(s)$, the output distribution, $P(q)$, the R&D expenditures distribution, $P(e)$, and the patent stock distribution, $P(k)$, across different years ranging from 1990 to 2005 using a logarithmic binning of the data [McManus et al., 1987].

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, $\bar{d}$, 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.

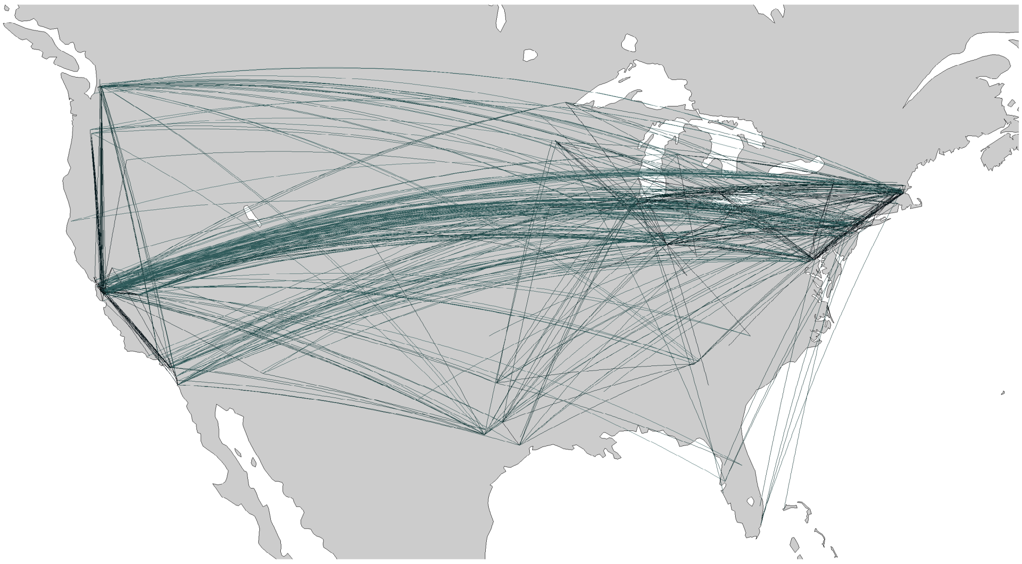

Figure H.10: The locations (at the city level) of firms and their R&D alliances in the combined CATI-SDC databases.

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

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

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

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

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]

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

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.

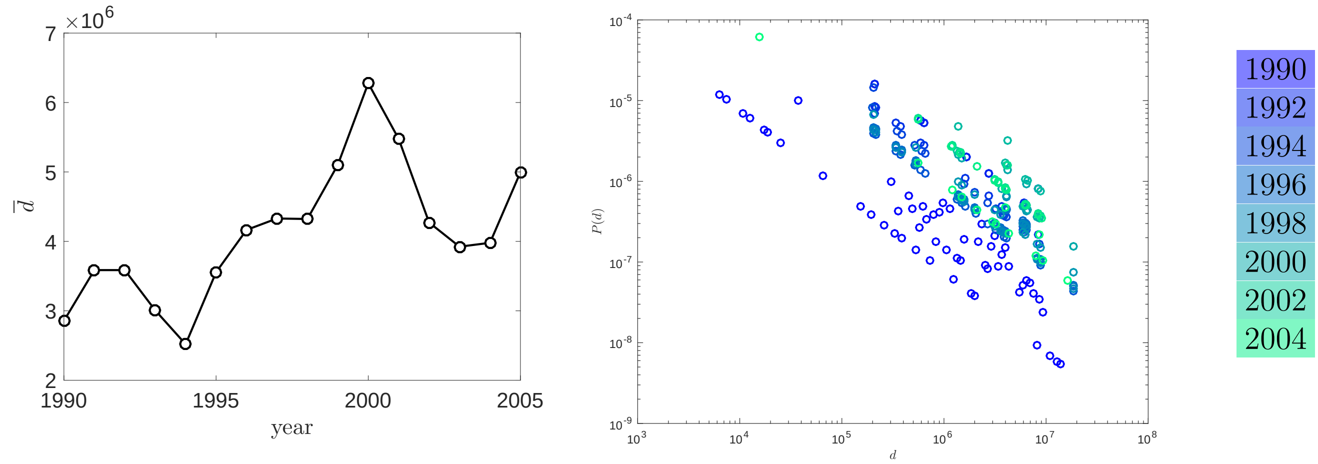

Figure I.11: The mean distance, $\overline{d}$, and the distance distribution, $P(d)$, across collaborating firms in the combined CATI-SDC database.

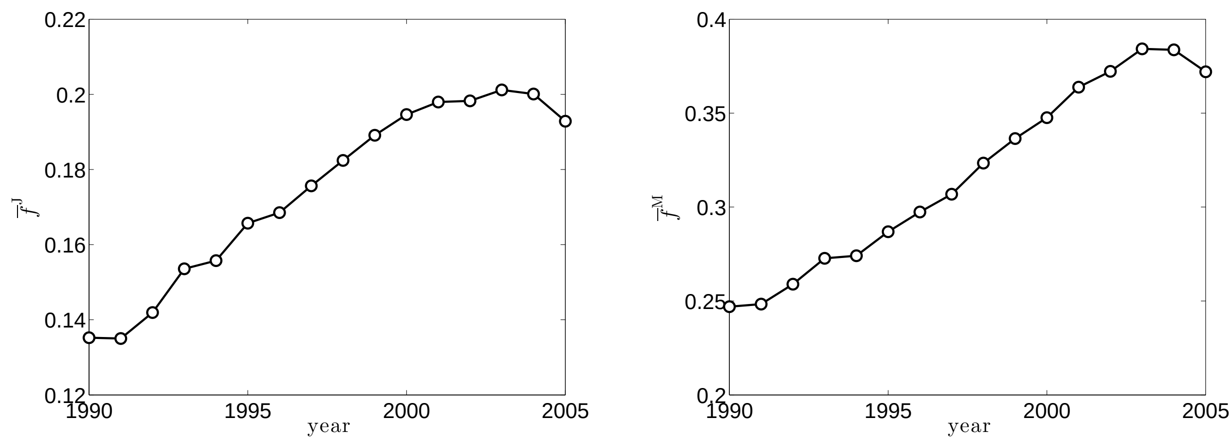

Figure I.12: The mean patent proximity across collaborating firms using the Jaffe metric f_{ij}^{J} of Equation (H.37) or the Mahalanobis metric f_{ij}^{M} of Equation (H.38).

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 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. 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.