Endogenous Technology Spillovers in R&D Collaboration Networks

Chih-Sheng Hsieh
Department of Economics and Center for Research in Econometric Theory and Applications (CRETA), National Taiwan University, Taipei 10617, Taiwan.

Michael D. König
CEPR, KOF-ETH Zurich, Tinbergen Institute and Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.

Xiaodong Liu
Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309, United States.

Abstract
We introduce a stochastic R&D network formation model where firms choose both R&D efforts and collaboration partners. Neighbors in the network benefit from each other's R&D efforts through local technology spillovers, and there exists a global competition effect reflecting strategic substitutability in R&D efforts. We provide a complete equilibrium characterization of the network formation model and show that the model is consistent with empirically observed R&D networks. Based on the equilibrium characterization, we propose an estimation method that is computationally feasible even for large networks. With the estimated model we then conduct an analysis of R&D collaboration subsidies to demonstrate the policy relevance of this model. We find that a subsidy scheme targeting specific R&D collaborations in the network can be much more effective than a uniform subsidy, with a welfare gain up to five times larger than the cost of the subsidy. (JEL: C11, C63, C73, D83, L22)

Keywords: R&D networks, network formation, peer effects, technology spillovers, subsidies.

1. Introduction

Research and development (R&D) collaborations facilitate technology spillovers between firms, lead to higher R&D investment levels, and improve the economic performance of firms (e.g., Belderbos et al. 2004). Fostering R&D collaborations has therefore been widely used by policymakers as a potent means to boost economic growth, improve industrial competitiveness, and generate welfare in society (OECD 2001, 2008). Since R&D collaborations not only affect but also are affected by

---

Acknowledgments: We would like to thank the editor Allan Collard-Wexler and three anonymous referees for their valuable suggestions. We would also like to thank Anton Badev, Eric Bartelsman, Filomena Garcia, David Hemous, Timo Hiller, Matt Jackson, Terence Johnson, Alexey Kushnir, Lung-Fei Lee, Elena Manresa, Andrea Montanari, José Moraga-González, Nick Netzer, Paolo Pin, Onur Özgür, Agnieszka Rusinowska, María Sáez-Martí, Armin Schmutzler, Martin Summer, Fernando Vega-Redondo, Joachim Voth, Timothy Van Zandt, Yves Zenou, Fabrizio Zilibotti, and seminar participants at the Workshop on Networks: Dynamics, Information, Centrality and Games at Sorbonne, the NSF Conference on Network Science in Economics, the Econometric Society World Congress, the congress of the European Economic Association, Stony Brook International Conference on Game Theory, the Public Economic Theory Conference, and the Universities of Zurich, Bocconi, Amsterdam, Bonn, Geneva, Cologne, Bern, Konstanz, National Taiwan University, Rice University, Stanford University, and the Austrian National Bank for the helpful comments and discussions at various stages of this paper. We further thank Christian Helmers for data sharing and Sebastian Ottinger for the excellent research assistance. Michael König acknowledges financial support from Swiss National Science Foundation through research grants PBEZP1–131169 and 100018_140266, and thanks SIEPR and the Department of Economics at Stanford University for their hospitality.

E-mail: cshsieh@ntu.edu.tw (Hsieh); m.d.konig@vu.nl (König); xiaodong.liu@colorado.edu (Liu)

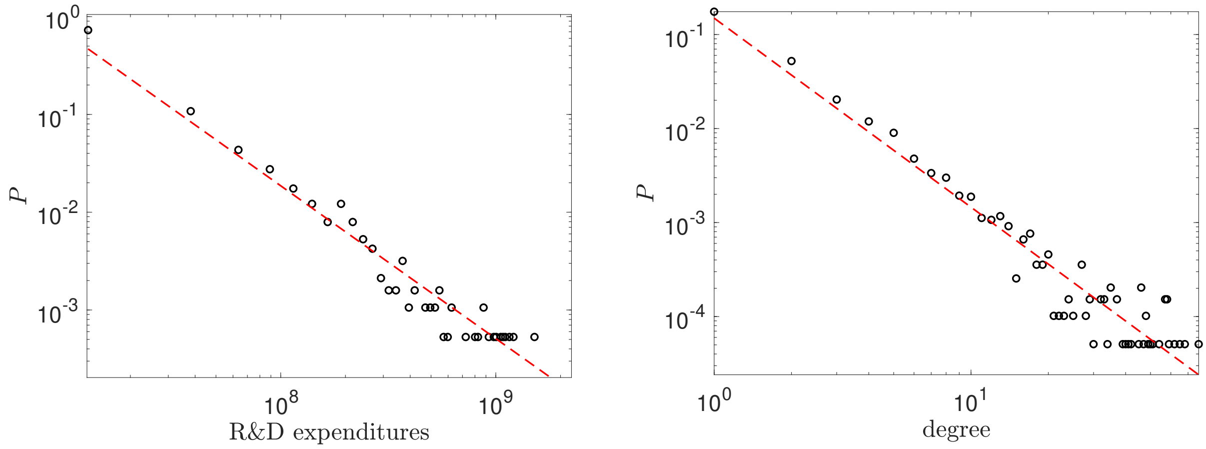


Figure 1. The left panel shows the empirical R&D expenditures distribution and the right panel shows the degree (number of R&D collaborations) distribution. The dashed lines indicate a power law decay in both distributions, which is a functional relationship of the form: $\log P(y) \propto -\alpha \log y$ for some $\alpha > 0$. A detailed description of the data used for the analysis in this paper can be found in Supplementary Appendix F.

firms' R&D investment decisions, to understand how R&D collaborations can facilitate technology spillovers and contribute to social welfare, it is crucial to account for the joint endogeneity of firms' R&D collaboration and investment decisions. Ignoring the interdependence of these two decisions might lead to biased estimation results and erroneous policy conclusions. In this paper, we introduce a tractable model to account for the co-evolution of R&D networks and firms' R&D expenditures, which can be applied to real-world network data and used for policy analysis.

Any suitable R&D network model should account for a number of empirical regularities that are characteristic of R&D networks. Similar to R&D expenditures, the distribution of the number of R&D collaborations (degree) is highly unequal and can be approximated with a power law over a wide range of the distribution. Such a power law characterizes a network in which most nodes have only a few links while a few nodes (hubs) accumulate a large fraction of all the links in the network. An illustration can be seen in Figure 1. Moreover, R&D networks exhibit a complex architecture composed of densely connected clusters dividing firms along geographic, technological, and sectoral domains. The largest connected component (i.e., a set of nodes that are connected via a sequence of links to each other) of the R&D network considered in the empirical analysis of this paper is shown in Figure 2. The figure indicates two clusters representing the manufacturing and pharmaceutical sectors, respectively, with most of the collaborations within a sector and a few collaborations across sectors. The adjacency matrices (with zero/one elements indicating whether pairs of firms have R&D collaborations) of these sectors are shown in Figure 3. These matrices are highly non-random and indicate a specific hierarchical structure of the underlying network that can be measured with the so-called "nestedness" coefficient (Atmar and Patterson 1993). A binary matrix is nested if

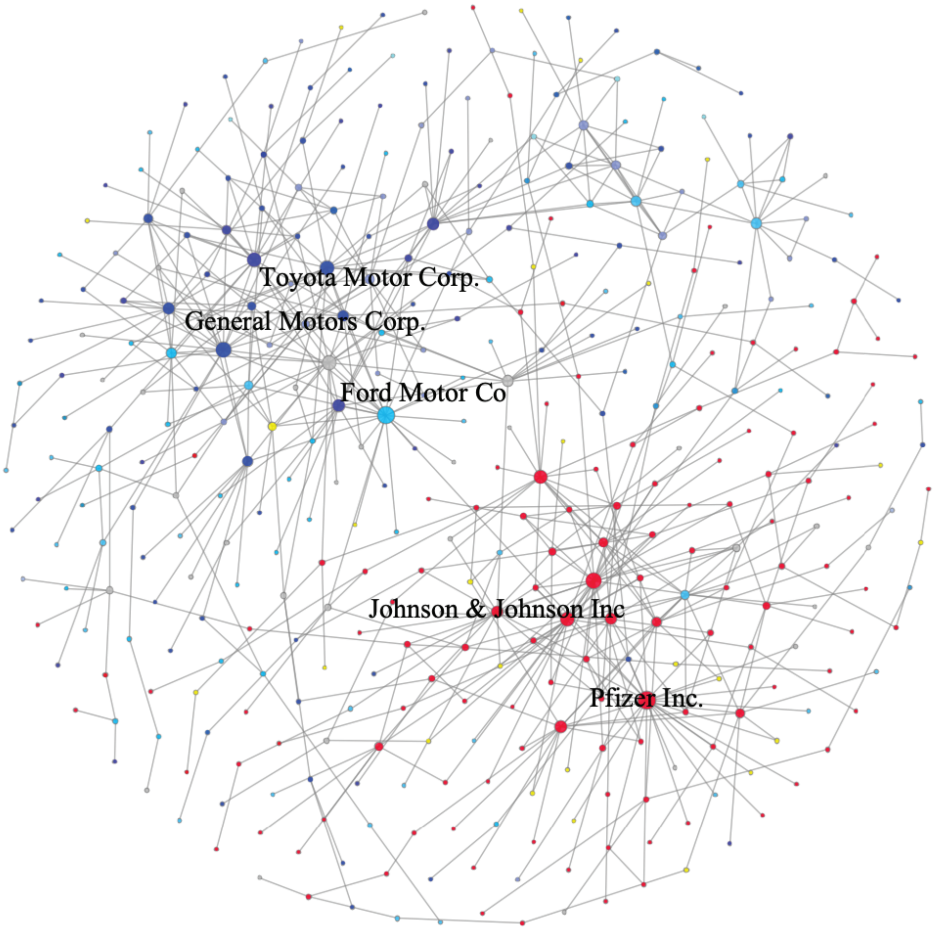


FIGURE 2. The largest connected component in the observed network of R&D collaborations in the year 2006 for the firms without missing observations on R&D expenditures and industry classifications. The color of a node indicates the sector of a firm (at the 2-digit SIC level) and its size scales with the number of R&D collaborations. The names of the five largest firms in terms of their R&D expenditures are shown in the graph. The number of firms in the largest connected component is 431.

most of its ones are concentrated in the top left corner of the matrix (in contrast to a non-nested matrix with evenly distributed ones). Moreover, in a network with a nested adjacency matrix there exists a hierarchy of connections such that nodes tend to be linked to other nodes with a higher degree (Mahadev and Peled 1995). In Figure 3 we observe that the nestedness coefficients of the adjacency matrices (which measure the degree of concentration of ones in the top left corner) in the manufacturing and pharmaceutical sectors are very high (as compared to a random matrix of the same size and density). This shows that the R&D collaborations within sectors follow a specific hierarchical organization that is indicated with a nested adjacency matrix.[1]

Overview of the results and contributions. Our paper makes three interrelated contributions in (i) network theory, (ii) econometrics, and (iii) policy analysis. Albeit our framework has a broad

1. A more detailed analysis of nestedness in the R&D network and its sub-networks can be found in Supplementary Appendix F.5.

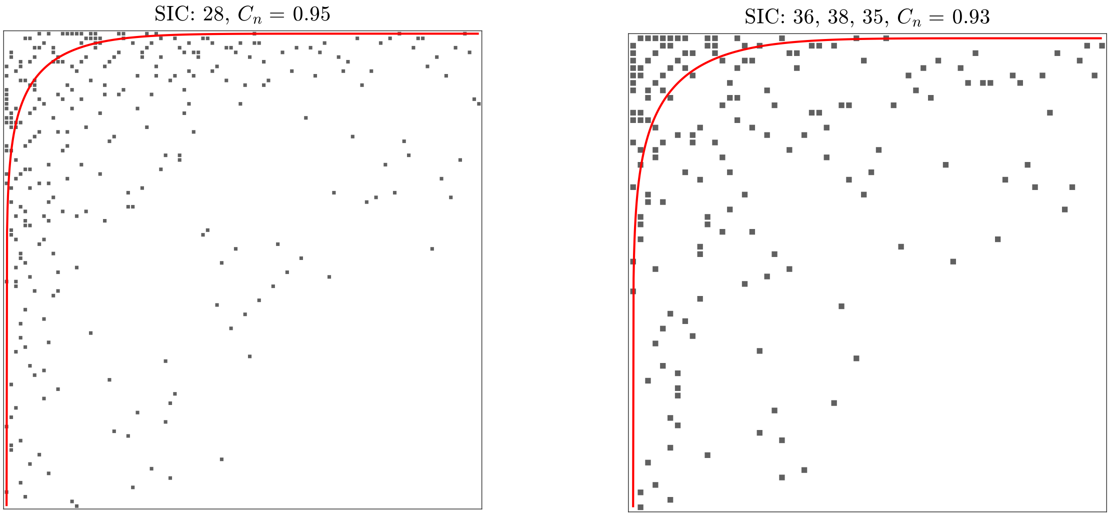


FIGURE 3. The left panel shows the adjacency matrix (with ones indicated with black squares and zeros indicated with white squares) of the largest connected component of firms in the chemicals and allied products sector (SIC code 28). The right panel shows the adjacency matrix of the largest connected component of firms in the three largest manufacturing sectors (SIC codes 35, 36, and 38). The degree of nestedness of an adjacency matrix is measured by the *nestedness coefficient*, C_n, which ranges from zero for a totally random (non-nested) network to one for a totally nested network (Atmar and Patterson 1993). The empirically observed nestedness coefficients are very high: for the left matrix (chemicals and allied products) $C_n = 0.95$ while for the right matrix (manufacturing) $C_n = 0.93$. The solid red line is the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectance). Both matrices are statistically significantly nested at the 5% level. Statistical significance is computed via comparison to a null model generated from 100 random networks with the same link density as the observed network.

range of applications in various fields (Jackson and Zenou 2015), in the following we will outline our contributions in the context of firms forming R&D collaborations to benefit from technology spillovers while, at the same time, competing in the product market (D'Aspremont and Jacquemin 1988).

First, we provide an analytic characterization of both R&D collaboration and investment choices in equilibrium by endogenizing the network structure in the model introduced by Ballester et al. (2006). We show that firms' R&D expenditures and the number of collaborations follow a power law distribution, consistent with the data illustrated in Figure 1. Moreover, we show that equilibrium networks are "nested split graphs" (Mahadev and Peled 1995), providing an explanation for why nestedness has been observed in empirical R&D networks (cf. Figure 3). Nested graphs have a core-periphery structure which has also been documented in empirical studies on R&D networks (Kitsak et al. 2010; Rosenkopf and Schilling 2007). Furthermore, Kitsak et al. (2010) find that firms in the core have a higher market share, consistent with the predictions of our model. We then investigate the efficient network structure that maximizes social welfare and show that efficient networks are also nested split graphs. However, we find that equilibrium networks tend to be under-connected compared to the social optimum. This suggests that policies promoting R&D collaborations can be welfare improving.

Next, we proposed an empirical estimation framework based on the analytic equilibrium characterization. The proposed composite likelihood estimation method (Lindsay 1988; Varin et al. 2011) accounts for the joint endogeneity of R&D collaborations and expenditures and remains computationally feasible even for large network datasets. We study firms' R&D collaboration and investment decisions using a comprehensive dataset on R&D collaborations matched to firms' balance sheets and patents. Consistent with the theory's predictions, our estimates show that the technology spillover effect has a positive impact on R&D investment while the competition effect has a negative impact. Technology spillovers in R&D collaborations dominate business stealing effects from competition. Moreover, our estimated model fits the data well along with various network statistics considered in the literature.

Finally, we conduct a novel counterfactual policy analysis taking into account that both the R&D collaboration network and firms' R&D investment decisions are endogenous and respond to a policy intervention. In particular, we consider several different policy scenarios subsidizing R&D collaboration costs: a uniform subsidy, a size-dependent subsidy (small vs. large firms), a sector-dependent subsidy (within vs. across sectors), and a targeted subsidy to specific R&D collaborations. To evaluate the efficiency of the policy, we compute the return on the R&D collaboration subsidy (ROS) defined as the ratio of the welfare gain induced by the subsidy relative to the cost of the subsidy. For the case of a uniform subsidy, we find that ROS of up to two can be achieved when subsidizing up to 20% of the collaboration costs. We also find that subsidizing collaborations involving large firms yields higher returns than subsidizing collaborations with small firms. Furthermore, subsidizing collaborations in more productive sectors yields higher returns than subsidizing less productive sectors, and subsidizing intra-industry collaborations tends to be more effective than cross-industry collaborations. Moreover, we find that a targeted subsidy can yield a welfare gain almost five times larger than the cost of the subsidy, which is more than two times higher than the ROS from a uniform subsidy. Subsidizing firms with higher productivity, which also tend to be larger firms in our data, yields the largest expected welfare gains and ROS as these firms have more collaborations and thus generate the highest spillover effects in the R&D network. Our finding contributes to the innovation policy debate emphasizing that it is important to subsidize the "right firms" (i.e., more productive firms) to be effective (cf. e.g. König et al. 2022). Our analysis further shows that ignoring network endogeneity in the policy analysis would underestimate the ROS substantially, which highlights the importance of taking network endogeneity into account when evaluating R&D policies. As subsidies have been increasingly used by governmental organizations to stimulate collaborative R&D activities (cf. Cohen 1994; Broekel and Graf 2012; Eureka 2021), our framework could be used to assist governmental funding agencies that typically do not take into account the technology spillovers generated within an endogenous R&D network structure.

Related literature. Our paper is related to previous studies on networks in the economics literature. The most closely related is the paper by Hsieh et al. (2022) where the authors analyze a similar econometric model for the coevolution of networks and behavior. However, different from Hsieh et al. (2022), here we also include a competition term in the firms' profit function, and we show analytically that our model generates equilibrium network structures that resemble real-world networks (i.e., the stochastically stable networks are nested split graphs and the degrees, as well as the R&D effort levels, are power law distributed, all of which is supported by the data). Furthermore, we develop a composite likelihood estimation method that can be applied to large networks, and we conduct a policy analysis of subsidies to R&D collaboration costs.

Moreover, similar to our framework, Dawid and Hellmann (2014), Goyal and Moraga-González (2001), and Westbrock (2010) study the formation of R&D networks in which firms can form collaborations to reduce their production costs. In particular, Dawid and Hellmann (2014) study a noisy best response dynamic process as we do here and analyze the stochastically stable states. However, different from the current model, they ignore the R&D investment decision, and the technology spillovers from collaborations in these models are independent of the identity and the characteristics of the firms involved. Goyal and Moraga-González (2001) present a more general setup which relaxes this assumption, but their analysis is restricted to regular graphs or networks with four firms. In this paper, we consider general equilibrium structures with an arbitrary number of firms and make no ex-ante restrictions on the potential collaboration pattern between them.

Our approach is a further generalization of the econometric network formation models proposed in Snijders (2001), Chandrasekhar and Jackson (2012), and Mele (2017), which also use a potential function (Monderer and Shapley 1996) to characterize the stationary states. But different from these papers, both action choices and linking decisions are fully endogenized in our model. Moreover, we provide microfoundations (from a Cournot competition model with externalities) for the potential function. In a recent paper by Badev (2021), a potential function is used to analyze the formation of networks in which agents not only form links but also make binary choices of adopting a certain behavior depending on the choices of their neighbors. Different from Badev (2021), we consider a continuum of choices, provide an explicit equilibrium characterization, and propose an alternative estimation method that can also be applied to large networks. Relatedly, Hsieh et al. (2020) also apply a potential function to an empirical model of joint network formation and action choices. However, their potential function is based on a transferable utility function so that linking decisions are based on maximizing aggregate payoffs, while here we consider decentralized link formation decisions between payoff maximizing agents.

Finally, in a related paper by König et al. (2019), the authors characterize the Nash equilibrium of R&D investment decisions taking the R&D network as given. In contrast, in the current paper, we explicitly model the joint endogeneity of R&D investment and collaboration decisions. Moreover, the R&D policy considered in König et al. (2019) subsidizes the total R&D expenditures of firms, while the R&D policy considered in this paper aims at subsidizing the R&D collaboration costs between firms. The latter is more relevant for large-scale R&D collaboration funding programs such as Eureka (Eureka 2021).

Organization of the paper. The theoretical model is outlined in Section 2. Section 2.1 introduces the profit function of the firm. Section 2.2 defines the stochastic network formation process, with its stationary distribution and complete equilibrium characterization given in Sections 2.3 and 2.4, respectively. In Section 2.5 the welfare maximizing networks are derived, and compared to the equilibrium outcome. Section 2.6 discusses several extensions of the model that allow for firm heterogeneity. Section 3 provides an empirical analysis of the model. Section 3.1 describes the data used for the empirical analysis, Section 3.2 introduces the empirical model, and Section 3.3 explains the estimation strategy. The estimation results are reported in Section 3.4. Section 3.5 analyzes the goodness-of-fit of our model to the data. Section 4 uses the estimated model to analyze the effectiveness of R&D collaboration subsidies. We consider four different cases, namely, a uniform subsidy (Section 4.1), a firm-size-specific subsidy (Section 4.2), a sector-specific subsidy (Section 4.3), and a subsidy targeting specific collaborations (Section 4.4). Finally, Section 5 concludes. All proofs are relegated to Appendix A.

Additional relevant material can be found in the supplementary appendices. Supplementary Appendix B provides basic definitions and characterizations of networks. Microfoundations for the profit function used in this paper from a Cournot oligopoly model with cost-reducing R&D collaborations can be found in Supplementary Appendix C. Supplementary Appendix D provides a complete equilibrium characterization in the case of homogenous firms. Supplementary Appendix E explains in detail the extensions mentioned in the main text. Supplementary Appendix F provides a detailed description of the data used for our empirical analysis. Supplementary Appendix G provides further details for the implementation of the estimation algorithm. Supplementary Appendix H provides a simulation study to examine the small sample properties of our estimation algorithm, as well as the impact of unobserved heterogeneity on the estimation results. Finally, Supplementary Appendix I provides additional empirical results and robustness checks.

2. Theoretical Framework

2.1. Profits from R&D Collaborations

Consider a set $\mathcal{N} = \{1, \ldots, n\}$ of firms and a collaboration network $G \in \mathcal{G}^n$, where $\mathcal{G}^n$ denotes the set of all graphs/networks with n nodes.[2] Let the firms' R&D effort levels be given by $\mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathcal{Y}^n$. Firm $i \in \mathcal{N}$ sets its R&D effort $y_i \in \mathcal{Y}$ and earns a profit $\pi_i \colon \mathcal{Y}^n \times \mathcal{G}^n \to \mathbb{R}$ given by the following linear-quadratic function:[3]

$$\pi_i(\mathbf{y}, G) = \eta_i y_i - \frac{1}{2} y_i^2 + \rho y_i \sum_{j=1}^{n} a_{ij} y_j - \lambda y_i \sum_{j \neq i}^{n} y_j - \zeta d_i. \tag{1}$$

The indicator variables $a_{ij} \in \{0, 1\}$ in Equation (1) indicate whether firms i and j are collaborating (or not), and can be represented by the symmetric adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ (with a zero diagonal, $a_{ii} = 0$). The degree, $d_i = \sum_{j=1}^{n} a_{ij}$, counts the number of collaborations of i, with its coefficient $\zeta > 0$ representing the linking cost. Net of linking costs, Equation (1) is identical to the payoff function analyzed in Ballester et al. (2006). This payoff function is characterized by a linear-quadratic term, $\eta_i y_i - \frac{1}{2} y_i^2$, representing direct gains from a firm's own R&D effort with η_i capturing firm heterogeneity in productivity and $-\frac{1}{2} y_i^2$ indicating decreasing returns in R&D effort; a local complementarity term, $\rho y_i \sum_{j=1}^{n} a_{ij} y_j$, reflecting technology spillover effects from R&D collaborations with the spillover parameter $\rho > 0$; and a global substitutability term, $\lambda y_i \sum_{j \neq i}^{n} y_j$, reflecting market stealing effects from the competition with the competition parameter $\lambda > 0$.

2. A summary of basic network definitions can be found in Supplementary Appendix B.

3. Microfoundations for the profit function in Equation (1) derived from a Cournot oligopoly model with cost reducing R&D collaborations can be found in Supplementary Appendix C.

An important feature of the profit function introduced in Equation (1) is that it admits a potential function (Monderer and Shapley 1996) as shown in the following proposition.

Proposition 1. *The profit function of Equation (1) admits a potential game where firms choose both R&D efforts and links with a potential function* $\Phi\colon \mathcal{Y}^n \times \mathcal{G}^n \to \mathbb{R}$ *given by*

$$\Phi(\mathbf{y}, G) = \sum_{i=1}^{n} \eta_i y_i - \frac{1}{2}\sum_{i=1}^{n} y_i^2 + \frac{\rho}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} y_i y_j - \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} y_i y_j - \zeta m, \tag{2}$$

for any $\mathbf{y} \in \mathcal{Y}^n$ *and* $G \in \mathcal{G}^n$*, where* $m = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$ *denotes the number of links in* G*.*

The potential function has the property that the marginal profit of a firm from adding or removing a link is equivalent to the change in the potential function of adding or removing a link. Similarly, the marginal profit of a firm from changing its R&D effort level is equivalent to the corresponding change of the potential function.[4] The potential function thus allows us to aggregate the incentives of the firms to either change their links or adjust their effort levels in a single global function. This feature of the potential function is crucial for the equilibrium characterization of the network formation process that will be introduced in the following section.

2.2. R&D Network Formation Process

In the following, we introduce a network formation process where firms endogenously choose both R&D efforts and collaboration partners based on the profit function given by Equation (1). We follow the *best response* dynamics dating back to Cournot (1838), where the opportunities for change arrive as a Poisson process (similar to the pricing model in Calvo 1983) and the firm that receives the opportunity maximizes profit by taking the R&D effort levels and collaborations of the other firms as given.[5] To capture the fact that R&D projects and collaborations are fraught with ambiguity

4. More formally, the potential Φ has the property that for any $\mathbf{y} \in \mathcal{Y}^n$ and $G, G' \in \mathcal{G}^n$ with $G' = G \oplus (i,j)$ or $G' = G \ominus (i,j)$ we have that $\Phi(\mathbf{y}, G') - \Phi(\mathbf{y}, G) = \pi_i(\mathbf{y}, G') - \pi_i(\mathbf{y}, G)$, where $G \oplus (i,j)$ ($G \ominus (i,j)$) denotes the network obtained from G by adding (removing) the link (i,j). Moreover, for $y_i, y_i' \in \mathcal{Y}$, $\mathbf{y}_{-i} \in \mathcal{Y}^{n-1}$ and $G \in \mathcal{G}^n$ we have that $\Phi(y_i', \mathbf{y}_{-i}, G) - \Phi(y_i, \mathbf{y}_{-i}, G) = \pi_i(y_i', \mathbf{y}_{-i}, G) - \pi_i(y_i, \mathbf{y}_{-i}, G)$.

5. Cournot (1838) analyzed a dynamic process in which firms myopically best respond in the current period to the existing output levels of all rivals (cf. Daughety 2008). For similar network formation models with myopic agents, see, e.g., Watts (2001) and Jackson and Watts (2002). The assumption of myopic behavior is common in the complex strategic environment that R&D networks represent. For example, Jackson and Watts (2002) state that "...in larger networks and

and uncertainty (Czarnitzki et al. 2015), we will introduce some noise in this decision process. The precise definition of the R&D effort adjustment and network evolution process is given as follows.

Definition 1 (Cournot Best Response Dynamics). *The network formation process is characterized by a sequence of states* $(\boldsymbol{\omega}_t)_{t\in\mathbb{R}_+}$, $\boldsymbol{\omega}_t \in \Omega = \mathcal{Y}^n \times \mathcal{G}^n$, *where each state* $\boldsymbol{\omega}_t = (\mathbf{y}_t, G_t)$ *consists of a vector of firms' R&D effort levels,* $\mathbf{y}_t \in \mathcal{Y}^n$, *and a network of collaborations,* $G_t \in \mathcal{G}^n$. *In a short time interval* $[t, t+\Delta t)$, $t \in \mathbb{R}_+$, *one of the following events happens:*

Effort adjustment: *At rate* $\chi \geq 0$, *a firm* $i \in \mathcal{N}$ *receives an R&D effort adjustment opportunity. We follow the random utility framework by assuming that the profit of firm i from choosing an R&D effort level* $y \in \mathcal{Y}$ *is given by* $\pi_i(y, \mathbf{y}_{-it}, G) + \varepsilon_{it}$, *where* $\pi_i(\mathbf{y}_t, G_t)$ *is defined in Equation (1) and* ε_{it} *is a random noise. Under the assumption that* ε_{it} *is identically and independently type-I extreme value distributed, the probability of choosing a certain R&D effort level is given by a continuous analog of the standard multinomial logit function (Anderson et al. 1992). More specifically, taking the R&D effort levels of all the other firms,* $\mathbf{y}_{-it}$, *and the network,* G_t, *as given, the probability that firm* i *adjusts the R&D effort level to* y *at time* t *is*

$$\mathbb{P}\left(\boldsymbol{\omega}_{t+\Delta t} = (y, \mathbf{y}_{-it}, G_t) | \boldsymbol{\omega}_t = (\mathbf{y}_t, G_t)\right) = \chi \frac{e^{\vartheta \pi_i(y, \mathbf{y}_{-it}, G_t)}}{\int_{\mathcal{Y}} e^{\vartheta \pi_i(y', \mathbf{y}_{-it}, G_t)} dy'} \Delta t + o(\Delta t), \tag{3}$$

where ϑ *captures the level of noise in the decision process.*

Link adjustment: *With rate* $\xi \geq 0$, *a pair of firms* (i,j) *receives the opportunity to form a link if they are not connected, or to remove the link if they are connected. Let* $G' = G \oplus (i,j)$ *(or* $G' = G \ominus (i,j)$*) denote the network obtained from* G *by adding (or removing) link* (i,j). *As* $\pi_i(\mathbf{y}, G') - \pi_i(\mathbf{y}, G) = \pi_j(\mathbf{y}, G') - \pi_j(\mathbf{y}, G) = \Phi(\mathbf{y}, G') - \Phi(\mathbf{y}, G)$, *where* $\pi_i(\mathbf{y}, G)$ *and* $\Phi(\mathbf{y}, G)$ *are defined in Equations (1) and (2) respectively, the incentives for firms* i *and* j *to create (or remove) a link are in line with each other and are identical to the corresponding change in the potential function. Following the random utility framework, we assume link* (i,j) *is created (or removed) if and only if* $\Phi(\mathbf{y}_t, G'_t) + \varepsilon'_{ij,t} > \Phi(\mathbf{y}_t, G_t) + \varepsilon_{ij,t}$, *where* $\varepsilon_{ij,t}$ *and* $\varepsilon'_{ij,t}$ *are identically and independently type-I extreme value distributed. Then, taking the R&D effort levels of all firms,* $\mathbf{y}_t$, *and the rest of*

networks where players' information might be local and limited, or in networks where players significantly discount the future, myopic behavior is a more natural assumption".

the network as given, the probability that link (i,j) is created (or removed) is

$$\mathbb{P}\left(\boldsymbol{\omega}_{t+\Delta t} = (\mathbf{y}_t, G'_t)|\boldsymbol{\omega}_t = (\mathbf{y}_t, G_t)\right) = \xi \frac{e^{\vartheta\Phi(\mathbf{y}_t, G'_t)}}{e^{\vartheta\Phi(\mathbf{y}_t, G'_t)} + e^{\vartheta\Phi(\mathbf{y}_t, G_t)}} \Delta t + o(\Delta t), \tag{4}$$

where ϑ captures the level of noise in the decision process.

The parameter ϑ is inversely related to the level of noise. In the limit of $\vartheta \to \infty$ the noise vanishes and firms choose the R&D effort level that maximizes $\pi_i(\mathbf{y}, G)$, while in the limit of $\vartheta \to 0$ the noise term dominates and R&D effort adjustments in Equation (3) become totally random. The same holds for the link formation and removal decisions in Equation (4).

Note that we can numerically simulate the stochastic process introduced in Definition 1 using the "next reaction method" for simulating a continuous time Markov chain (Gibson and Bruck 2000). We use this method throughout the paper to illustrate our theoretical predictions for various network statistics (for example in Figure 5). However, this method is computationally infeasible for large networks, where we have to rely on our theoretical equilibrium characterization and alternative simulation methods.

2.3. Stationary Distribution

With the potential function $\Phi(\mathbf{y}, G)$ defined in Proposition 1, we can derive the *stationary distribution* of the network formation process in the form of an exponential family distribution called the *Gibbs measure* (cf. Grimmett 2010).

Theorem 1. *The stochastic process $(\boldsymbol{\omega}_t)_{t\in\mathbb{R}_+}$ with states $\boldsymbol{\omega}_t \in \Omega = \mathcal{Y}^n \times \mathcal{G}^n$ is an ergodic Markov chain with a unique stationary distribution $\mu^\vartheta(\mathbf{y}, G) \colon \Omega \to [0,1]$ such that $\lim_{t\to\infty} \mathbb{P}(\boldsymbol{\omega}_t = (\mathbf{y}, G)|\boldsymbol{\omega}_0 = (\mathbf{y}_0, G_0)) = \mu^\vartheta(\mathbf{y}, G)$. The distribution $\mu^\vartheta(\mathbf{y}, G)$ is given by the Gibbs measure*

$$\mu^\vartheta(\mathbf{y}, G) = \frac{e^{\vartheta\Phi(\mathbf{y}, G)}}{\sum_{G' \in \mathcal{G}^n} \int_{\mathcal{Y}^n} e^{\vartheta\Phi(\mathbf{y}', G')} d\mathbf{y}'}, \tag{5}$$

for any $\mathbf{y} \in \mathcal{Y}^n$ and $G \in \mathcal{G}^n$.

From Theorem 1 we know that the Markov chain is ergodic, so that the Ergodic Theorem applies (Norris 1998), which states that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{1}_{\{\omega_s \in \mathcal{A}\}} ds = \mu^\vartheta(\mathcal{A}), \quad \mathbb{P}\text{-a.s.},$$

for any measurable set $\mathcal{A} \in \Omega$, and long-run averages of sample paths converge to the invariant distribution. Moreover, for any measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ in $L^1(\mathbb{P})$ we have that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(\omega_s)\, ds = \mathbb{E}_{\mu^\vartheta}(f), \quad \mathbb{P}\text{-a.s.},$$

where $\mathbb{E}_{\mu^\vartheta}(f)$ is the expected value of f under the invariant probability measure $\mu^\vartheta(\mathbf{y}, G)$. Note that since the stationary distribution $\mu^\vartheta(\mathbf{y}, G)$ characterizes the long run outcome, Equation (5) does not depend on the R&D effort adjustment rate χ nor the link adjustment rate ξ.

In the empirical analysis of the model, we assume the observation $(\mathbf{y}, G)$ is a random draw from the stationary distribution $\mu^\vartheta(\mathbf{y}, G)$ given in Theorem 1. Then, the structural parameters of the model can be identified based on the maximum likelihood principle (Lewbel 2019). However, the evaluation of $\mu^\vartheta(\mathbf{y}, G)$ is computationally costly, if not impossible, for a large network due to the intractable denominator $\sum_{G' \in \mathcal{G}^n} \int_{\mathcal{Y}^n} e^{\vartheta \Phi(\mathbf{y}', G')} d\mathbf{y}'$. To overcome this computational difficulty, we take a *composite likelihood* approach (Lindsay 1988; Varin et al. 2011) based on the conditional distributions $\mu^\vartheta(\mathbf{y}|G)$ and $\mu^\vartheta(G|\mathbf{y})$ given in the following proposition. These conditional distributions are simple to compute. They also provide additional insight into the properties of the stationary distribution.

PROPOSITION 2. *The stationary distribution $\mu^\vartheta(\mathbf{y}, G)$ given in Theorem 1 admits the following conditional distributions.*

(i) Assume $\mathbf{M}(G) \equiv \mathbf{I}_n - \rho \mathbf{A} + \lambda \mathbf{B}$ is nonsingular, where $\mathbf{A} \equiv (a_{ij})_{1 \le i,j \le n}$ is the adjacency matrix and $\mathbf{B}$ is a zero-diagonal square matrix with all off-diagonal elements equal to one. Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$. Then, the stationary distribution of R&D efforts $\mathbf{y}$, conditional on the R&D network G, is given by

$$\mu^\vartheta(\mathbf{y}|G) = \left(\frac{2\pi}{\vartheta}\right)^{-\frac{n}{2}} |\mathbf{M}(G)|^{\frac{1}{2}} \exp\left\{-\frac{\vartheta}{2}\left(\mathbf{y} - \mathbf{M}(G)^{-1}\boldsymbol{\eta}\right)^\top \mathbf{M}(G)(\mathbf{y} - \mathbf{M}(G)^{-1}\boldsymbol{\eta})\right\}. \tag{6}$$

(ii) The stationary distribution of the R&D network G, conditional on R&D efforts $\mathbf{y}$, is given by

$$\mu^\vartheta(G|\mathbf{y}) = \prod_{i=1}^{n} \prod_{j=i+1}^{n} \frac{e^{\vartheta a_{ij}(\rho y_i y_j - \zeta)}}{1 + e^{\vartheta(\rho y_i y_j - \zeta)}}. \tag{7}$$

Proposition 2 (i) shows that, conditional on the R&D network G, the stationary distribution of R&D efforts $\mathbf{y}$ is jointly normal with mean $\mathbf{M}(G)^{-1}\boldsymbol{\eta}$ and variance $\frac{1}{\vartheta}\mathbf{M}(G)^{-1}$. The matrix $\mathbf{M}(G)$

captures the interdependence of R&D efforts due to the spillover effect (ρ) and the competition effect (λ). In the absence of these effects (i.e., $\rho = \lambda = 0$), the matrix $\mathbf{M}(G)$ reduces to the identity matrix $\mathbf{I}_n$ and the R&D effort of a firm y_i becomes independently normally distributed with mean η_i and variance $\frac{1}{\vartheta}$.

Proposition 2 (ii) indicates that, conditional on the R&D efforts $\mathbf{y}$, the network links a_{ij} are pairwise independent with the linking probability

$$\mathbb{P}(a_{ij} = 1|\mathbf{y}) = \frac{e^{\vartheta(\rho y_i y_j - \zeta)}}{1 + e^{\vartheta(\rho y_i y_j - \zeta)}}. \tag{8}$$

Equation (8) suggests that firms with higher R&D efforts are more likely to form an R&D collaboration. Although the network links are conditionally independent given the R&D efforts, they are not *unconditionally* independent due to the interdependence of R&D efforts in the presence of the spillover effect ($\rho \neq 0$). This is an important feature of our model that distinguishes it from an inhomogeneous random graph model (defined in Supplementary Appendix B).

2.4. Equilibrium Characterization of Stochastically Stable States

In this subsection, we derive the R&D effort profile and network $(\mathbf{y}, G)$ in the limit of vanishing noise (i.e., $\vartheta \to \infty$). In the following, we denote $\mu^*(\mathbf{y}, G) \equiv \lim_{\vartheta \to \infty} \mu^{\vartheta}(\mathbf{y}, G)$. We call the states in the support of $\mu^*(\mathbf{y}, G)$ the *stochastically stable* states (Kandori et al. 1993; Jackson and Watts 2002). The set of stochastically stable states is denoted by $\Omega^* \subseteq \Omega$. From the Gibbs distribution in Equation (5) it follows that $(\mathbf{y}, G) \in \Omega^*$ if and only if $\Phi(\mathbf{y}, G) \geq \Phi(\mathbf{y}', G')$ for all $\mathbf{y}' \in \mathcal{Y}^n$ and $G' \in \mathcal{G}^n$. The following proposition provides an explicit characterization of the equilibrium R&D effort profile and network in the stochastically stable state.[6]

PROPOSITION 3. *In the limit of $\vartheta \to \infty$, the stochastically stable network $G \in \mathcal{G}^n$ in the support of $\mu^*(\mathbf{y}, G)$ is a nested split graph*[7] *in which the link between firms i and j is determined by the*

---

6. Proposition 3 analyzes the general case of heterogeneous firms (that is, $\eta_i \neq \eta_j$ for at least one pair $i, j \in \mathcal{N}$) while additional results for the special case of ex-ante homogeneous firms ($\eta_i = \eta$ for all $i \in \mathcal{N}$) can be found in Supplementary Appendix D.

7. A *nested split graph* is characterized by the fact that the neighborhood of every node is contained in the neighborhoods of the nodes with higher degrees (Mahadev and Peled 1995). That is, the neighborhoods of the nodes are nested. Furthermore, the adjacency matrix of a nested split graph is stepwise, meaning that a step function separates the zero entries from the one entries in the matrix. See Supplementary Appendix B for further details.

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

FIGURE 4. The (stepwise) adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ of a nested split graph, with a_{ij} given in Proposition 3. The matrices from left to right correspond to increasing linking costs $\zeta \in \{0.0075, 0.01, 0.02\}$. The parameters used are $n = 10$, $\lambda = 0.06$, $\rho = 0.02$ and $\boldsymbol{\eta} = (1.00, 0.71, 0.58, 0.50, 0.45, 0.41, 0.38, 0.35, 0.33, 0.32)^\top$, where the firms with the highest (lowest) productivity η_i have the lowest (highest) index.

indicator function

$$a_{ij} = \mathbb{1}_{\{y_i y_j > \zeta/\rho\}}, \quad \mu^*\text{-a.s.}, \tag{9}$$

and the stochastically stable R&D effort profile $\mathbf{y} \in \mathcal{Y}^n$ *in the support of* $\mu^*(\mathbf{y}, G)$ *is the fixed point to the following system of equations*

$$y_i = \eta_i + \sum_{j \neq i}^{n} y_j \left(\rho a_{ij} - \lambda\right), \quad \mu^*\text{-a.s.}. \tag{10}$$

Moreover, if firm i *is more productive than firm* j $(\eta_i > \eta_j)$*, then* i *has higher R&D effort* $(y_i > y_j)$ *and more collaborations* $(d_i > d_j)$ *than* j*,* μ^**-a.s..*

Proposition 3 has some important implications. First, in the limit of $\vartheta \to \infty$, the network in the stochastically stable state is a nested split graph with a stepwise adjacency matrix. In Figure 4, we numerically generate some adjacency matrices $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$, where the entries of the adjacency matrix a_{ij} together with the equilibrium effort level y_i are jointly determined by Equations (9) and (10) in Proposition 3. The figure shows that the numerically generated adjacency matrix $\mathbf{A}$ is indeed stepwise, characterizing a nested split graph,[7] and becomes increasingly sparse with increasing linking cost ζ. This is consistent with the nestedness of empirical R&D networks in our data as shown in Figure 3. Furthermore, nested split graphs are paramount examples of core-periphery networks. The core-periphery structure of R&D alliance networks has been documented empirically in Kitsak et al. (2010) and Rosenkopf and Schilling (2007). Our model thus provides a theoretical explanation for why real-world R&D networks exhibit such a core-periphery structure.

Second, Proposition 3 shows that firms with higher productivity tend to have higher R&D effort levels and form more links in the stochastically stable state. As a firm's R&D effort level is associated

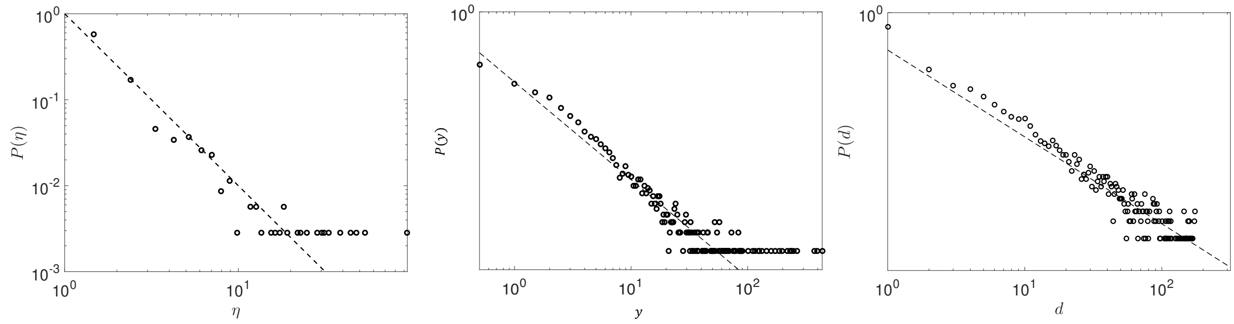


FIGURE 5. The distribution $P(\eta)$ of η following a power law with exponent 2 (left panel), the resulting R&D effort distribution $P(y)$ (middle panel) and the degree distribution $P(d)$ (right panel) from a numerical simulation of the stochastic process of Definition 1. Dashed lines indicate a power law fit. The parameters used are $n = 350$, $\lambda = 0.75$, $\rho = 2$ and $\zeta = 75$.

with its market value (see Supplementary Appendix C) and, in a nested split graph, firms with more links are more central in the network, Proposition 3 implies that firms with high R&D effort and market value are at the core of the R&D network. This is in line with what we observe in Figure 2 and findings in previous empirical studies (e.g., Kitsak et al. 2010).

Third, Proposition 3 provides a plausible explanation on the heavy-tailed distributions of R&D efforts and network degrees that are prevalent in empirical data. Taking network links a_{ij} as given, it follows from Equation (10) that R&D efforts are weighted sums of productivities, $\mathbf{y} = \mathbf{M}(G)^{-1}\boldsymbol{\eta}$, where $\mathbf{M}(G)$ is defined in Proposition 2. Hence, when productivities $(\eta_i)_{i=1}^n$ are ex-ante drawn from a power law distribution, the distribution of R&D efforts also follows a power law distribution in the tail (Wang and Tang 2006). On the other hand, taking the R&D efforts $\mathbf{y}$ as given, network links are determined by Equation (9). Hence, when R&D efforts are power law distributed, the degree distribution will also be a power law (Boguná and Pastor-Satorras 2003).[8] An example based on a numerical simulation of the stochastic process of Definition 1 with a power law distribution of η can be seen in Figure 5. Our model can thus provide an explanation for the heavy-tailed distributions of R&D efforts and network degrees as illustrated in Figure 1.

---

8. Equation (9) can be rewritten as $a_{ij} = \mathbb{1}_{\{y_i y_j > \zeta/\rho\}} = \mathbb{1}_{\{\log y_i + \log y_j > \log(\zeta/\rho)\}}$. When R&D effort levels are power law distributed, with density $f(y) = \frac{\gamma}{c}\left(\frac{c}{y}\right)^{\gamma+1}$ for $y > c$, where $c > 0$ is a lower-cut-off and $\gamma > 0$ is a positive parameter, the log-effort levels, $\ln y$, follow an exponential distribution with density $f(y) = \gamma c^{\gamma} e^{-\gamma y}$. Boguná and Pastor-Satorras (2003) show that the random graph with links a_{ij} determined by Equation (9) is characterized by a power law degree distribution, a negative clustering degree correlation, and a decaying average nearest neighbor degree distribution indicating a disassortative network. In Appendix E.1 we discuss how such network characteristics can also be obtained when firms are heterogeneous in terms of their marginal collaboration costs.

2.5. Efficiency

Social welfare, $W(\mathbf{y}, G)$, is given by the sum of consumer surplus, $U(\mathbf{y})$, and producer surplus, $\Pi(\mathbf{y}, G)$. Consumer surplus is given by $U(\mathbf{y}) = \frac{\psi}{2}\sum_{i=1}^n y_i^2 + \frac{\lambda}{2}\sum_{i=1}^n \sum_{j\neq i}^n y_i y_j$ with $\psi > \lambda > 0$ (Singh and Vives 1984; Vives 1999).[9] Producer surplus is given by firms' aggregate profits $\Pi(\mathbf{y}, G) = \sum_{i=1}^n \pi_i(\mathbf{y}, G)$. The *efficient* state is then defined by the network $G^* \in \mathcal{G}^n$ and R&D effort profile $\mathbf{y}^* \in \mathcal{Y}^n$ that maximize welfare $W(\mathbf{y}, G)$, that is, $W(\mathbf{y}^*, G^*) \geq W(\mathbf{y}, G)$ for all $G \in \mathcal{G}^n$ and $\mathbf{y} \in \mathcal{Y}^n$. The following proposition shows that similar to the equilibrium network (Proposition 3), the efficient network is a nested split graph. However, the equilibrium network tends to be under-connected, and the equilibrium R&D effort is low compared to what would be socially optimal.

PROPOSITION 4. *The efficient network $G^* \in \mathcal{G}^n$ is a nested split graph where the R&D effort profile $\mathbf{y}^* \in \mathcal{Y}^n$ is the solution to the following system of equations:*

$$y_i = \frac{\eta_i}{1-\psi} + \frac{1}{1-\psi}\sum_{j\neq i}^n y_j \left(\rho \mathbb{1}_{\{\rho y_i y_j > \zeta\}} - \lambda\right). \tag{11}$$

Furthermore, the stochastically stable equilibrium R&D efforts and the collaboration intensity are too low compared to the social optimum (μ^*-a.s.).

Comparing the recursive Equations (10) and (11) in Propositions 3 and 4, respectively, we observe that the equilibrium R&D effort is lower than the efficient R&D effort level (since $\frac{1}{1-\psi} > 1$), and this difference increases with ψ. Since both the equilibrium network and the efficient network are nested split graphs in which the link between firms i and j is determined by Equation (9), the equilibrium R&D effort being less than the social optimum implies that the equilibrium network tends to be under-connected. In Section 4 we will analyze the welfare-improving impact of a subsidy on firms' R&D collaboration costs, which gives firms additional incentives to form collaborations and thus increases the network connectivity.

---

9. See Supplementary Appendix C for further details and explanation.

2.6. Extensions

The model presented so far can be extended in a number of different directions that account for firm heterogeneity, which are summarized below and described in greater detail in Supplementary Appendix E.

a) Heterogeneous collaboration costs: We can extend the model by allowing for heterogenous collaboration costs (see Supplementary Appendix E.1 for further details). One can show that a similar equilibrium characterization using a Gibbs measure as in Theorem 1 is possible. Moreover, in the special case that the productivity is power law distributed, one can show that the degree distribution also follows a power law distribution (see Proposition E.1), consistent with previous empirical studies of R&D networks, together with other empirically relevant correlations (see Propositions E.2 and E.3).[10]

b) Heterogeneous technology spillovers: We can further extend the model by assuming that there are heterogeneous spillovers between collaborating firms depending on their technology portfolios (see Supplementary Appendix E.2 for further details). For example, under the assumption that firms can only benefit from collaborations if they have at least one technology in common, one can show that our model is a generalization of a "random intersection graph" (see supplementary Appendix B) for which positive degree correlations can be obtained (i.e., "assortativity", see Proposition E.4).

The above extensions show that our model is capable of generating networks with various properties that can be observed in real-world networks, such as power law degree distributions and degree correlations, once we introduce firm heterogeneity. This contrasts with (simple variants of) exponential random graphs often used in the statistics/econometrics literature, which often have difficulties in generating networks with empirically relevant characteristics (Chandrasekhar and Jackson 2012).

10. We also note that other statistics such as the clustering degree distribution can be computed. See Supplementary Appendix E.2 for further details. In particular, under the assumption of a power law productivity distribution, we can generate two-vertex and three-vertex degree correlations, such as a decreasing average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, indicating a disassortative network, as well as a decreasing clustering degree distribution, $C(d)$.

3. Empirical Analysis

3.1. Data

To get a comprehensive picture of R&D alliances we use data on interfirm R&D collaborations stemming from two sources that have been widely used in the literature (Schilling 2009). The first is the Cooperative Agreements and Technology Indicators (CATI) database (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 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 R&D collaborations. Supplementary Appendix F provides more information about the databases used for this study.

We then merged the CATI database with the Thomson SDC alliance database. For the matching of firms across datasets, we adopted and extended the name-matching algorithm developed as part of the NBER patent data project (Trajtenberg et al. 2009).[11] The systematic collection of inter-firm alliances in CATI started in 1987 and ended in 2006. As the CATI database only includes collaborations up to 2006, we take this year as the base year for our empirical analysis. We then construct the R&D alliance network by assuming that an alliance lasts for 5 years (similar to, e.g., Rosenkopf and Padula 2008). The corresponding entry in the adjacency matrix between two firms is coded as one if an alliance between them exists during this period, and zero otherwise. An illustration of the observed R&D network can be seen in Figure 2. The figure indicates two clusters representing the manufacturing and the pharmaceutical sectors, respectively, with most collaborations within a sector and a few across sectors. We will use the fact that these two sectors (markets) are not perfectly overlapping with the R&D collaboration clusters (within and across sectors) to separately identify the technology spillover and the product market competition effect.

11. 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.

The combined CATI-SDC database provides only the names of the firms in an alliance. To obtain information on their balance sheets and income statements we matched the firms' names in the CATI-SDC database with the firms' names in Standard & Poor's Compustat U.S. and Global Fundamentals databases, as well as Bureau van Dijk's Orbis database (Bloom et al. 2013). For the purpose of matching firms across databases, we employ the aforementioned name-matching algorithm. The balance sheet information is available for roughly 25% of the firms in the alliance data.[12] By matching the firms' names between the alliance database and the Compustat and Orbis databases, we obtain a firm's R&D expenditure, primary industry code, and location for a sample of 1,738 firms with 632 R&D collaborations.[13]

We use a firm's log R&D expenditures to measure its R&D effort. A firm's productivity is measured by its log R&D capital stock (lagged by one year) representing the learning-by-doing effect from conducting R&D.[14] The R&D capital stock is computed using a perpetual inventory method based on the firms' R&D expenditures with a 15% depreciation rate (Hall et al. 2000; Bloom et al. 2013; König et al. 2019).[15] We further identify the patent portfolios of the firms in our data using

---

12. For many small private firms, no information on R&D expenditures is available, either because these firms do not conduct R&D or due to missing data. Hence, these firms could not be matched by our algorithm. Nevertheless, R&D is mostly concentrated in larger firms, which cover most of the R&D activities in the economy and thus tend to generate larger R&D spillovers (cf. Bloom et al. 2013), and these firms are typically included in our sample.

13. After matching the CATI-SDC R&D alliance database with Compustat-Orbis and keeping only firms without missing information on R&D expenditures (the dependent variable in our empirical analysis), we obtain a sample composed of 1,931 firms (which corresponds to 25% of all firms in the alliance data). We further drop firms with missing information about the industry in which they operate and industries with only a single firm (in order to identify the competition effect). This leaves us with a sample of 1,738 firms. The two samples (with 1,738 firms and 1,931 firms respectively) show only minor differences in descriptive statistics: the average R&D expenditure is slightly higher (9.1467 vs. 9.0671), and the average number of R&D collaborations is slightly lower (0.7273 vs. 0.7281) in the sample of 1,738 compared to the sample of 1,931 firms.

14. We have also used value-added per employee as an additional measure of a firm's productivity. However, as many firms have missing information on value-added or employment, using value-added per employee as a productivity measure would reduce the sample size by 25% while leaving the estimation results qualitatively unchanged. In order to avoid this reduction in the sample size, we used the time-lagged log R&D capital stock to measure a firm's productivity as our main specification. The estimation results for the alternative specification with value-added per employee can be found in Supplementary Appendix Table I.1.

15. We also use different depreciation rates (10% and 20%) to calculate the R&D capital stock and find the estimation results are qualitatively similar. The detailed estimation results are in Supplementary Appendix Table I.2.

Table 1. Descriptive statistics.

Variable	All Sectors mean	s.d.	min	max	SIC 283 mean	s.d.	min	max
R&D effort	9.147	2.226	0.000	15.247	8.950	2.033	1.632	15.247
Collaborators' R&D efforts	8.453	23.729	0.000	242.357	11.085	25.696	0.000	192.290
Competitors' R&D efforts	1338.013	1386.564	15.227	3669.380	1440.998	557.776	159.886	1903.573
Productivity	10.598	2.267	0.643	17.061	10.172	2.070	3.838	16.816
R&D collaborations	0.727	2.057	0.000	24.000	0.966	2.340	0.000	19.000
Number of firms	1738				410			

Notes: R&D effort is measured by log R&D expenditure (measured in thousand U.S. dollars). The reference year is 2006. Competitors of a firm are the firms operating in the same three-digit SIC sector. A firm's productivity is measured by its log-R&D capital stock (lagged by one year). To compute the R&D capital stocks we use a perpetual inventory method based on the firms' R&D expenditures with a 15% depreciation rate (cf. Hall et al. 2000; Bloom et al. 2013). SIC 283 refers to the drugs development sector.

the EPO Worldwide Patent Statistical (PATSTAT) database (Jaffe and Trajtenberg (2002); see also Supplementary Appendix F.4). 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 (Copeland and Fixler 2012). The technology classes are identified using the main international patent classification (IPC) numbers at the 4-digit level.

Table 1 shows descriptive statistics of the full sample covering all sectors and the subsample of the drugs development sector with standard industry classification (SIC) code 283, which is not only the largest three-digit SIC sector in our sample (with 410 firms) but also has the largest number of R&D collaborations (with 198 collaborations). We focus on this sector (SIC 283) to conduct an analysis of R&D collaboration subsidies in Section 4. Table 1 shows that the total R&D effort of the competitors of a firm is much higher than that of collaborating firms. This is due to the fact that firms typically have more competitors in the product market (defined as all other firms in the same three-digit SIC sector) than R&D collaborators. In the SIC 283 subsample, the average R&D effort of a firm is slightly lower than that in the full sample. However, due to a relatively larger number of firms and R&D collaborations in SIC 283, on average, the total R&D effort of both, the collaborators and competitors in SIC 283, are higher than in the full sample including all sectors. Moreover, in Table 2 we show the correlations between a firm's own R&D effort, the collaborators' total efforts, and the competitors' total efforts. We observe that, for both the full sample and the SIC 283 subsample, the correlation between the firm's own effort and the collaborators' efforts is significantly positive, while the correlation between the firm's own effort and the competitors' efforts is significantly negative. This is consistent with the predictions of the theoretical model.

Table 2. Correlations between a firm's own R&D effort and peers' (collaborators' or competitors') R&D efforts.

	All Sectors			SIC 283		
	Own effort	Collab. efforts	Compet. efforts	Own effort	Collab. efforts	Compet. efforts
Own R&D effort	1.000			1.000		
Collaborators' efforts	0.399	1.000		0.515	1.000	
Competitors' efforts	-0.058	0.131	1.000	-0.156	0.090	1.000

Notes: The correlation coefficients reported in this table are all statistically significant at the 10% level.

3.2. Empirical Model

To account for the firm level heterogeneity in the empirical analysis, we extend the profit function of Equation (1) in Section 2 to accommodate heterogeneous spillover effects and collaboration costs (see also Section 2.6), so that the profit of firm i can be written as follows:

$$\pi_i(\mathbf{y}, G) = \eta_i y_i - \frac{1}{2} y_i^2 + \rho \sum_{j=1}^{n} f_{ij} a_{ij} y_j y_i - \lambda \sum_{j \neq i}^{n} b_{ij} y_j y_i - \sum_{j=1}^{n} a_{ij} \zeta_{ij}, \tag{12}$$

where η_i represents firm level heterogeneity in productivity. We define

$$\eta_i = \mathbf{X}_i \boldsymbol{\delta} + \kappa z_i, \tag{13}$$

where $\mathbf{X}_i$ is a (row) vector of observable measures for productivity including one-year-lagged log R&D capital stock, a sector dummy (at the four-digit SIC level), and z_i is a latent variable that captures unobserved heterogeneity in productivity. When the coefficient κ is positive, a firm with a higher z_i is more productive ceteris paribus.

The coefficient ρ multiplied on the term $\sum_{j=1}^{n} f_{ij} a_{ij} y_j y_i$ in Equation (12) measures the R&D spillover effect. The indicator variable a_{ij} is equal to one if firms i and j form an R&D collaboration and zero otherwise. To capture heterogeneous spillover effects across firms, we further introduce the weights $(f_{ij})_{1 \leq i,j \leq n}$ (with $f_{ij} = f_{ji}$) based on the technological similarity of firms i and j (in terms of their patent portfolios),[16] so that firms with high (low) technological similarities can benefit more (less) from their collaborations.

The structure of the product market is captured by the indicator variable $b_{ij} \in \{0, 1\}$ such that $b_{ij} = 1$ if firms i and j are in the same three-digit SIC sector and $b_{ij} = 0$ otherwise. Each firm faces a substitutability effect (or competition effect) from all other firms within the same sector (market) measured by the coefficient λ.

16. See Supplementary Appendix E.2 for a simple example and the implications for the network structure.

The total cost of R&D collaborations for firm i is given by $\sum_{j=1}^{n} a_{ij}\zeta_{ij}$, with the pairwise symmetric marginal cost function

$$\zeta_{ij} = \mathbf{W}_{ij}\boldsymbol{\gamma} - z_i - z_j. \tag{14}$$

The r-dimensional (row) vector of dyad-specific variables, $\mathbf{W}_{ij}$, captures the similarity between firms i and j regarding the sector, location, productivity, etc., that might affect the collaboration cost (see, e.g., Lychagin et al. 2016). We also include individual latent variables z_i and z_j in Equation (14) to capture unobserved degree heterogeneity (Graham 2017).[17] More specifically, as the latent variable z_i is positively associated with a firm's productivity when $\kappa > 0$ in Equation (13), subtracting z_i and z_j from $\mathbf{W}_{ij}\boldsymbol{\gamma}$ in Equation (14) implies that more productive firms have lower costs to form R&D collaborations and are more likely to collaborate with each other.

The generalized potential function corresponding to Equation (12) is given by

$$\Phi(\mathbf{y}, G) = \sum_{i=1}^{n} \eta_i y_i - \frac{1}{2}\sum_{i=1}^{n} y_i^2 + \frac{\rho}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} f_{ij}a_{ij}y_i y_j - \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} b_{ij}y_i y_j - \frac{1}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} a_{ij}\zeta_{ij}.$$

In vector-matrix form, this can be written as

$$\Phi(\mathbf{y}, G) = \boldsymbol{\eta}^{\top}\mathbf{y} - \frac{1}{2}\mathbf{y}^{\top}\mathbf{M}(G)\mathbf{y} - \frac{1}{2}\mathrm{tr}(\mathbf{A}\boldsymbol{\zeta}), \tag{15}$$

with $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^{\top}$, $\boldsymbol{\zeta} = (\zeta_{ij})_{1\leq i,j\leq n}$, $\mathbf{M}(G) = \mathbf{I}_n - \rho(\mathbf{A}\circ\mathbf{F}) + \lambda\mathbf{B}$, $\mathbf{A} = (a_{ij})_{1\leq i,j\leq n}$, $\mathbf{B} = (b_{ij})_{1\leq i,j\leq n}$, $\mathbf{F} = (f_{ij})_{1\leq i,j\leq n}$, and $\circ$ denotes the Hadamard element-wise matrix product. The stationary distribution of the Markov process described in Definition 1 is then given by the Gibbs measure $\mu^{\vartheta}(\mathbf{y}, G)$ defined in Equation (5) of Theorem 1 with the potential function $\Phi(\mathbf{y}, G)$ given by Equation (15). In the following, we introduce an estimation approach that is based on this stationary distribution $\mu^{\vartheta}(\mathbf{y}, G)$.

3.3. Composite Likelihood Estimation

Assuming the observation $(\mathbf{y}, G)$ is a random draw from the stationary distribution $\mu^{\vartheta}(\mathbf{y}, G)$ given in Theorem 1, the structural parameters of the model can be identified and estimated based on the maximum likelihood principle (Lewbel 2019). However, the evaluation of $\mu^{\vartheta}(\mathbf{y}, G)$

17. Supplementary Appendix E.1 discusses a simple example and what this implies for the network structure.

is computationally cumbersome due to the intractable denominator $\sum_{G' \in \mathcal{G}^n} \int_{\mathcal{Y}^n} e^{\vartheta \Phi(\mathbf{y}', G')} d\mathbf{y}'$. To overcome this computational difficulty, Hsieh et al. (2022) adopt the Double Metropolis-Hastings (DMH) algorithm that avoids the evaluation of the intractable denominator of $\mu^{\vartheta}(\mathbf{y}, G)$. While the DMH algorithm can accommodate a more general linking cost function than the one given by Equation (14) (see, e.g., Mele 2017; Hsieh et al. 2022), it is computationally costly for large networks since the DMH algorithm needs to repeatedly simulate auxiliary network data in the Markov Chain Monte Carlo (MCMC) procedure. In this paper, we show that, for the linking cost function given by Equation (14), the structural parameters of the model can be estimated by the *composite likelihood* approach (Lindsay 1988; Varin et al. 2011). The composite likelihood method is computationally simple even for large networks.

In the absence of the latent variable z_i, Equations (13) and (14) become $\eta_i = \mathbf{X}_i \boldsymbol{\delta}$ and $\zeta_{ij} = \mathbf{W}_{ij} \boldsymbol{\gamma}$. In this case, the composite likelihood method maximizes the objective function

$$\ln \mu^{\vartheta}(\mathbf{y}|G) \mu^{\vartheta}(G|\mathbf{y}) = \ln \mu^{\vartheta}(\mathbf{y}|G) + \ln \mu^{\vartheta}(G|\mathbf{y}),$$

where $\mu^{\vartheta}(\mathbf{y}|G)$ and $\mu^{\vartheta}(G|\mathbf{y})$ are defined in Proposition 2. More specifically, $\mu^{\vartheta}(\mathbf{y}|G)$ is the probability density function of a multivariate normal distribution with mean $\mathbf{M}(G)^{-1}\boldsymbol{\eta} = \mathbf{M}(G)^{-1}\mathbf{X}\boldsymbol{\delta}$, where $\mathbf{X} = (\mathbf{X}_1^{\top}, \ldots, \mathbf{X}_n^{\top})^{\top}$, and variance $\frac{1}{\vartheta}\mathbf{M}(G)^{-1}$; and $\mu^{\vartheta}(G|\mathbf{y})$ is the joint probability function of conditionally pairwise independent network links[18] with

$$\mathbb{P}(a_{ij} = 1|\mathbf{y}) = \frac{e^{\vartheta(\rho f_{ij} y_i y_j - \zeta_{ij})}}{1 + e^{\vartheta(\rho f_{ij} y_i y_j - \zeta_{ij})}} = \frac{e^{\vartheta(\rho f_{ij} y_i y_j - \mathbf{W}_{ij}\boldsymbol{\gamma})}}{1 + e^{\vartheta(\rho f_{ij} y_i y_j - \mathbf{W}_{ij}\boldsymbol{\gamma})}}. \tag{16}$$

As in a conditional autoregressive (CAR) model in the spatial statistics literature (see, e.g., Besag 1974, 1975), the structural parameters ρ, λ, $\boldsymbol{\delta}$ and ϑ can be identified from the conditional distribution $\mu^{\vartheta}(\mathbf{y}|G)$.[19] In particular, ρ and λ can be separately identified when the R&D collaboration network and the product competition network do not perfectly overlap with each other (i.e., $\mathbf{A} \circ \mathbf{F} \neq \mathbf{B}$). Given that ϑ and ρ are identified, $\boldsymbol{\gamma}$ can be identified from the conditional

18. It is worth pointing out that the network links are not *unconditionally* independent due to the interdependence of R&D efforts in the presence of the spillover effect ($\rho \neq 0$).

19. In Supplementary Appendix H.1, we show that, when the latent variable z_i is absent in the data generating process, the spillover effect ρ and the substitutability effect λ can be successfully recovered from the conditional distribution $\mu^{\vartheta}(\mathbf{y}|G)$ alone.

distribution $\mu^{\vartheta}(G|\mathbf{y})$. Using the terminology in Lewbel (2019), the above identification strategy is distribution based. Since the score function of the composite likelihood is a linear combination of valid score functions associated with each log-likelihood term (i.e., $\ln \mu^{\vartheta}(\mathbf{y}|G)$ and $\ln \mu^{\vartheta}(G|\mathbf{y})$), the composite likelihood estimator is consistent and asymptotically normal (see, e.g., Pauli et al. 2011).

A well-known identification problem for network models is the so-called "reflection problem" (Manski 1993), which refers to the difficulty of disentangling the spillover effect from other confounding effects. In the context of R&D networks, if one fails to control for an unobservable firm-specific attribute that is associated with a firm's tendency to increase R&D spending and form more R&D collaborations, then the R&D spillover effect is likely to be overestimated. Hence, including the latent variable z_i in Equations (13) and (14) can alleviate the potential reflection problem. In the presence of the latent variable z_i, the R&D effort profile $\mathbf{y}$ follows a joint normal distribution conditional on the network G and latent variables $\mathbf{z}$ with mean $\mathbf{M}(G)^{-1}\boldsymbol{\eta} = \mathbf{M}(G)^{-1}(\mathbf{X}\boldsymbol{\delta} + \kappa\mathbf{z})$, where $\mathbf{z} = (z_1, \ldots, z_n)^{\top}$, and variance $\frac{1}{\vartheta}\mathbf{M}(G)^{-1}$. Since we use a single snapshot of $\mathbf{y}$, instead of repeated observations of $\mathbf{y}$ over time, we cannot identify the latent variable z_i from $\mu^{\vartheta}(\mathbf{y}|G, \mathbf{z})$ alone. However, as the latent variable z_i shows up in all network links of firm i, it can be identified from

$$\mathbb{P}(a_{ij} = 1|\mathbf{y}, \mathbf{z}) = \frac{e^{\vartheta(\rho f_{ij} y_i y_j - \zeta_{ij})}}{1 + e^{\vartheta(\rho f_{ij} y_i y_j - \zeta_{ij})}} = \frac{e^{\vartheta(\rho f_{ij} y_i y_j - \mathbf{W}_{ij}\boldsymbol{\gamma} + z_i + z_j)}}{1 + e^{\vartheta(\rho f_{ij} y_i y_j - \mathbf{W}_{ij}\boldsymbol{\gamma} + z_i + z_j)}}$$

following the same argument as in Graham (2017). Intuitively, assuming that ϑ can be identified from the conditional distribution $\mu^{\vartheta}(\mathbf{y}|G, \mathbf{z})$, we can identify z_i from the different linking decisions, $a_{i1}, a_{i2}, a_{i3}, \ldots$, of firm i in a similar manner as the individual effect in panel data. On the other hand, the identification of ϑ from the conditional distribution $\mu^{\vartheta}(\mathbf{y}|G, \mathbf{z})$ relies on the identification of z_i from $\mu^{\vartheta}(G|\mathbf{y}, \mathbf{z})$. Therefore, it is crucial to estimate $\mu^{\vartheta}(\mathbf{y}|G, \mathbf{z})$ and $\mu^{\vartheta}(G|\mathbf{y}, \mathbf{z})$ jointly in the composite likelihood framework. We assume the latent variable z_i is normally distributed with zero mean and variance σ_z^2, and take the Bayesian data augmentation approach (Zeger and Karim 1991) to implement the composite likelihood estimation.[20] The technical details regarding the Bayesian estimation procedure are provided in Supplementary Appendix G. We also perform Monte Carlo simulations to gauge the finite-sample performance of the proposed Bayesian estimation procedure

20. By contrast, the frequentist estimation approach needs to evaluate the high-dimensional integrals in $\mu^{\vartheta}(G|\mathbf{y})\mu^{\vartheta}(\mathbf{y}|G) = \int \mu^{\vartheta}(G|\mathbf{y}, \mathbf{z})\mu^{\vartheta}(\mathbf{y}|G, \mathbf{z}) f(\mathbf{z}) d\mathbf{z}$, where $\mathbf{z} = (z_1, \ldots, z_n)^{\top}$ and $f(\mathbf{z})$ denotes the density function of $\mathbf{z}$, and hence is computationally cumbersome. A similar Bayesian approach using the composite likelihood has been considered in Pauli et al. (2011), Friel (2012) and Ribatet et al. (2012).

and the results in Supplementary Appendix H.2 show that the estimator can successfully recover the true parameter values of the model.

3.4. Estimation Results

Table 3 presents the estimation results for the homogeneous spillover case, i.e., when the weight $f_{ij} = 1$ is applied to all pairs $i, j \in \mathcal{N}$.[21] We consider two alternative models: Model (A) ignores the latent variables z_i's in Equations (13) and (14) while Model (B) includes them. The estimated technology spillover effect ρ and substitutability effect λ are statistically significant and have the expected signs in both models, which match the theoretical predictions from Section 2. This shows that firms face a positive complementary effect from R&D collaborations and a negative product substitutability effect from competing firms in the same market. The estimated substitutability effect coefficient λ is small because it is associated with the total R&D effort of all the other firms in the same sector (i.e., the competitors). Since the total R&D effort of the competitors is high on average (see Table 1), the substitutability effect is not negligible as shown in the following marginal effect analysis. Moreover, we find that failing to control for unobserved individual heterogeneity with the latent variables in Model (A) leads to a significant upward bias on the estimate of the spillover effect ρ, which is consistent with the Monte Carlo simulation results reported in Supplementary Appendix H.2. Finally, we find that more productive firms are more active in R&D investment (see, e.g., Cohen et al. 1987) and firms in the same sector or in the same country are more likely to form R&D collaborations.

The magnitudes of the estimated ρ and λ are not directly comparable due to the different densities of the collaboration and competition networks. To gain a deeper understanding of the relative magnitudes of the estimated ρ and λ, we carry out the following marginal effect analysis. Recall the conditional expectation of the R&D profile $\mathbf{y}$ given the network G is $\mathbb{E}(\mathbf{y}|G) = \mathbf{M}(G)^{-1}(\mathbf{X}\boldsymbol{\delta} + \kappa\mathbf{z})$, where $\mathbf{M}(G) = \mathbf{I}_n - \rho(\mathbf{A} \circ \mathbf{F}) + \lambda\mathbf{B}$. The (i, k)th element of $\mathbf{X}$, denoted by x_{ik}, represents the ith observation on the kth explanatory variable. Then, the marginal effect of x_{ik} is given by $\text{ME}_{ik} = \sum_{j=1}^{n} \partial\mathbb{E}(y_j|G)/\partial x_{ik}$. In the absence of the spillover effect and substitutability effect (i.e., $\rho = \lambda = 0$), $\mathbf{M}(G) = \mathbf{I}_n$ and hence the marginal effect of x_{ik} is simply $\text{ME}_{ik} = \sum_{j=1}^{n} \partial\mathbb{E}(y_j|G)/\partial x_{ik} = \delta_k$, where

21. In Supplementary Appendix Tables I.3 and I.4, we provide additional estimation results by taking into account heterogeneous technology spillovers among collaborating firms.

Table 3. Estimation results for the case of homogeneous technology spillovers.

		Model (A) W/o Unobs. Heterogeneity		Model (B) With Unobs. Heterogeneity	
Profits					
R&D Spillover	(ρ)	0.017***	(0.001)	0.010***	(0.001)
Substitutability	(λ)	3.77e-5***	(1.35e-5)	3.45e-5***	(1.35e-5)
Productivity	(δ)	0.848***	(0.002)	0.853***	(0.002)
Unobs. Heterogeneity	(κ)	–		0.010***	(0.004)
Sector Dummies		Yes		Yes	
Linking Cost					
Constant	(γ_0)	6.843***	(0.180)	8.454***	(0.274)
Same Sector	(γ_1)	-1.194***	(0.055)	-1.479***	(0.083)
Same Country	(γ_2)	-0.379***	(0.048)	-0.648***	(0.077)
Diff-in-Productivity	(γ_3)	-0.090***	(0.011)	0.002	(0.014)
Noise/Uncertainty					
Noise in Decisions	(ϑ)	1.736***	(0.048)	1.421***	(0.043)
Unobs. Heterogeneity	(σ_z^2)	–		1.020***	(0.129)
Sample Size	(n)	1,738			

Notes: The dependent variables are log-R&D expenditures and (presence or absence of) R&D collaborations between firms. A firm's productivity is measured by its one-year-lagged log-R&D capital stock. To compute the R&D capital stock we use a perpetual inventory method based on the firms' R&D expenditures with a 15% depreciation rate (cf. Hall et al. 2000; Bloom et al. 2013). The parameters are given in the empirical profit function of Equation (12). We make 50,000 MCMC draws where the first 10,000 draws are dropped during a burn-in phase and every 10th of the remaining draws are kept to calculate the posterior mean (as point estimates) and posterior standard deviation (shown in parenthesis). All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery et al. (1992). The asterisks *** (**,*) indicate that a parameter's 99% (95%, 90%) highest posterior density range does not cover zero (Kruschke 2015).

δ_k denotes the kth element of $\boldsymbol{\delta}$. Figure 6 plots the distribution of the marginal effect of productivity based on the estimates of Model (B) in Table 3. In the left panel, we set the spillover effect to zero ($\rho = 0$) and the substitutability effect parameter λ to its estimate. In this case, the average marginal effect $\text{AME}_k = \frac{1}{n}\sum_{i=1}^{n} \text{ME}_{ik}$ (indicated by the blue line) is lower than the estimated coefficient of productivity (δ). In the middle panel, we set the substitutability effect to zero ($\lambda = 0$) and the spillover effect parameter ρ to its estimate. In this case, the average marginal effect is higher than the estimated δ. In the right panel, we set both ρ and λ to their estimates. In this case, the average marginal effect is also higher than the estimated δ. In summary, the spillover effect and the substitutability effect amplify and reduce the marginal effect of productivity respectively, and the spillover effect dominates the substitutability effect. This echoes the previous findings by Bloom et al. (2013).

3.5. Goodness-of-fit

In this subsection, we investigate how well our model fits the observed network data. We simulate one hundred networks using our model with the estimates reported for Model (B) in Table 3. The goodness-of-fit of our model is examined by comparing the simulated $\mathbf{y}$ and G with the observed

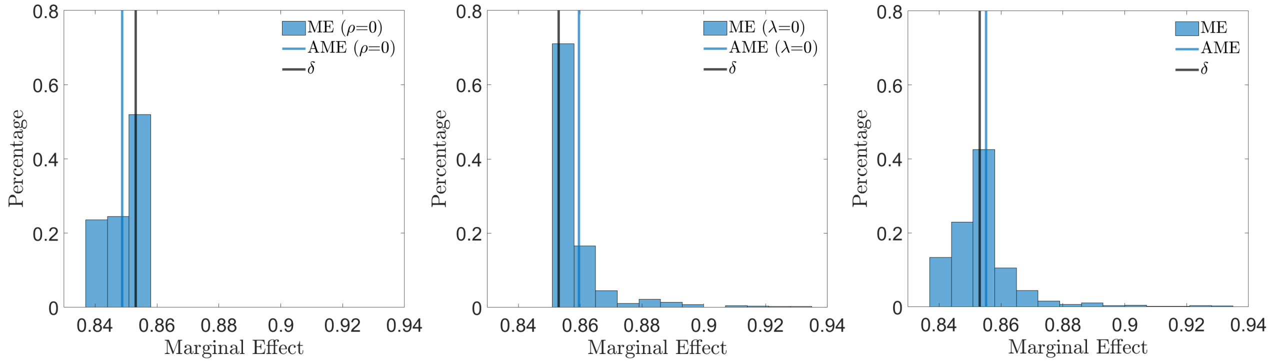


FIGURE 6. Marginal effect (ME) of productivity.

ones from the full sample of 1,738 firms in terms of the distributions of the R&D effort and the four network statistics considered in Hunter et al. (2008): the degree, the geodesic distance, the number of edge-wise shared partners, and the degree-specific clustering coefficient. The degree is the number of links of a node. The geodesic distance between any two nodes refers to the length of the shortest path joining these two nodes. The edge-wise shared nodes contain information related to the count of triangles in a network. The clustering coefficient is the fraction of the number of links between the neighbors of a node divided by the maximum number of links that could possibly exist between the neighbors. In Figure 7, we plot the empirical distributions of the above statistics from the observed data (by solid lines) and the corresponding means and 95% confidence intervals (by dashed lines) from the simulated data. From the figure, we can see that the simulated $\mathbf{y}$ and G from our model match the observed ones for all statistics considered, which suggests our model does a good job of capturing the underlying data generating process.[22]

4. R&D Collaboration Subsidy Analysis

Many governments provide R&D subsidies to foster the R&D activities of firms (see e.g., Cohen 1994). One example is the Advanced Technology Program (ATP) which was administered by the National Institute of Standards and Technology (NIST) in the U.S. (Feldman and Kelley 2003). In Europe, Eureka is a Europe-wide network for promoting market-driven collaborative research and

22. To further illustrate the goodness-of-fit of our model, we compare the observed links with the simulated links for the largest 150 firms in our sample in Supplementary Appendix Figure I.1.

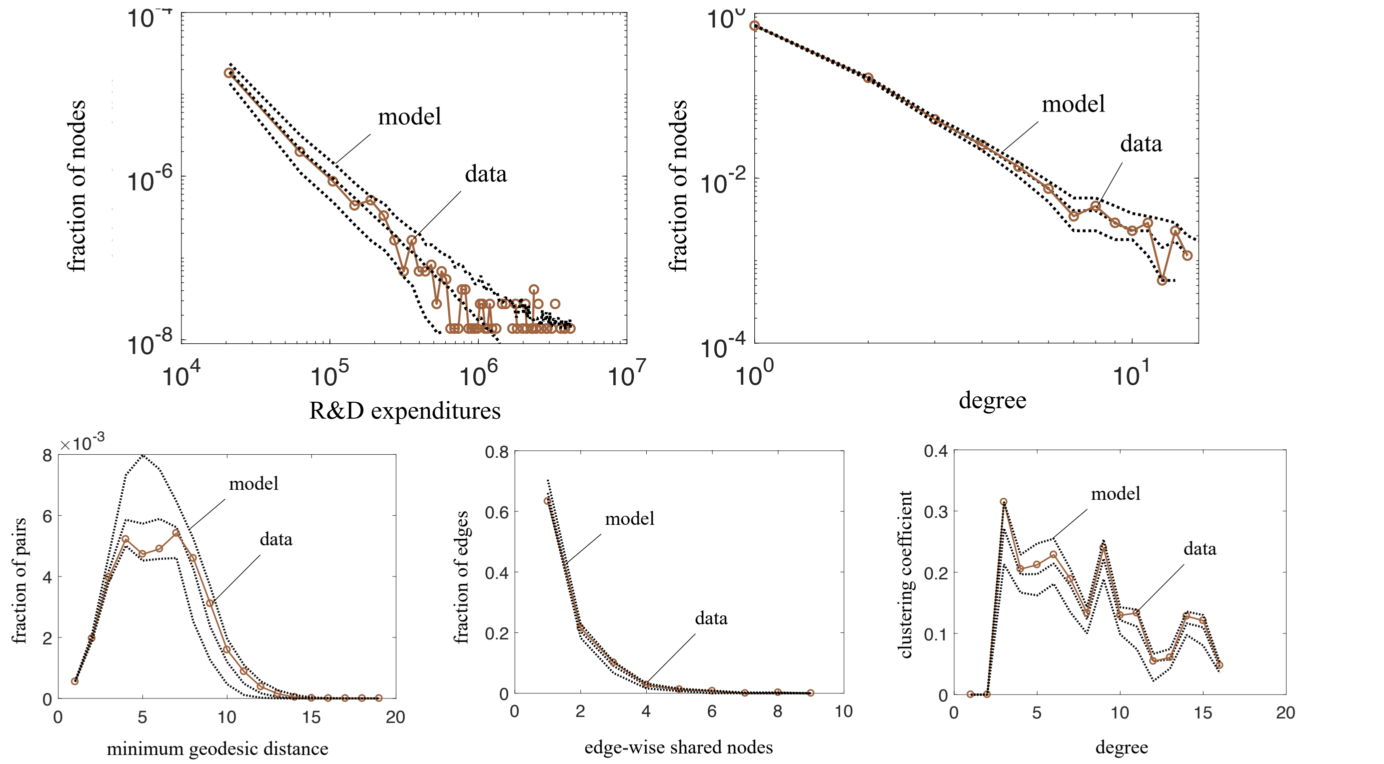


FIGURE 7. Goodness-of-fit statistics.

technology development. Its subsidies amounted to more than €682 million in 2021, with a total of €7.1 billion since 2014, and most subsidies awarded to firms in the biotech sector (Eureka 2021).[23]

In this section, we analyze the welfare impact of subsidizing R&D collaboration costs in the framework of an endogenous R&D network. More specifically, suppose that the government subsidizes the R&D collaboration cost between firms with a subsidy rate of $s \in [0, 1]$. With this subsidy rate, the R&D collaboration cost of firms i and j reduces from ζ_{ij} to $(1 - s)\zeta_{ij}$, while the government bears the cost of the subsidy, $s\zeta_{ij}$. To evaluate the welfare impact of the subsidy, we perform 300 simulations based on the network formation process in Definition 1 with the estimates of Model (B) reported in Table 3, and calculate welfare (defined in Section 2.5) across the 300 simulated networks in the stationary state with and without the subsidy. For this exercise, the parameter ψ in consumer surplus is normalized to one. The welfare gain is defined as the difference between the average welfare with and without the subsidy, and the *return on subsidy* (ROS) is defined as the welfare gain divided by the total cost of the subsidy. Because the computational cost of simulating

23. See Takalo et al. (2013) for an analysis of R&D subsidies provided by the Finnish Funding Agency for Technology and Innovation. The effect of federal government subsidies to stimulate R&D collaboration activities in Germany is analyzed in Broekel and Graf (2012). For a general discussion about the effectiveness of public R&D funding see Bloom et al. (2019).

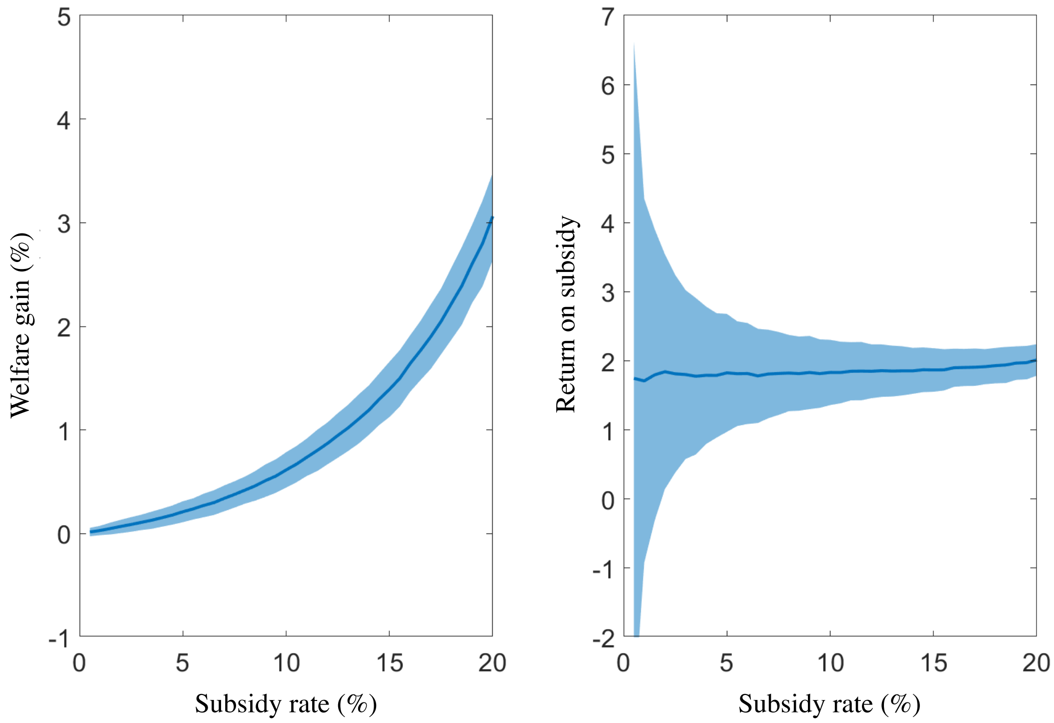


FIGURE 8. The left panel shows the welfare gain and the right panel the return on subsidy (ROS) for varying uniform subsidy rates in the drugs development sector (SIC code 283). Solid lines indicate the average across 300 simulations, while the shaded areas represent the 95% confidence interval.

the network formation process increases exponentially with the network size, we only focus on the drugs development sector (SIC 283) for the subsidy analysis, which also has the largest number of R&D collaborations across sectors.

In the following subsections, we consider four different cases for the analysis of the effectiveness of the R&D collaboration subsidy: a uniform subsidy (Section 4.1), a firm-size-specific subsidy (Section 4.2), a sector-specific subsidy (Section 4.3), and a subsidy targeting specific collaborations (Section 4.4).

4.1. Uniform Subsidy

In this subsection, we consider a subsidy that reduces the collaboration costs uniformly across all firm pairs. Figure 8 plots the welfare gain and the ROS under this subsidy scheme. We find that, for a subsidy rate of up to 20%, a welfare gain of up to 3% can be achieved with an ROS of up to 2. The ROS shows that the welfare gain from the subsidy can be twice as large as the cost of the subsidy. As discussed in more detail in Section 2.5, the welfare gain due to the subsidy stems from the R&D externalities that are not fully internalized by the firm's profit-maximizing decisions.

4.2. Firm-Size-Specific Subsidy

Many countries have introduced R&D support programs that discriminate between small and large firms. For example, Dechezleprêtre et al. (2016) analyze an R&D policy in the United Kingdom targeting specifically small firms. Relatedly, Mindruta et al. (2016) provide an empirical analysis of the importance of firm size on the formation of R&D collaborations in the biotech-pharmaceutical industry. Motivated by the empirical relevance, in this subsection, we consider a firm-size-specific subsidy.

Figure 9 plots the ROS for three different subsidy schemes that reduce the collaboration cost between (i) small firms (firms with size below the 10% firm size percentile), (ii) small and large firms, and (iii) large firms (firms with size above the 90% firm size percentile), by 10%, respectively. We find that subsidizing a collaboration between two small firms yields a ROS lower than one, indicating that the cost of the subsidy exceeds the welfare gain; while subsidizing a collaboration involving a large firm yields a ROS higher than one, suggesting that the welfare gain of the subsidy exceeds its cost. In particular, subsidizing a collaboration between two large firms generates a welfare gain that is more than double the cost of the subsidy.

It is not surprising that subsidizing collaborations involving large firms leads to higher welfare gains as larger firms also tend to be more productive. Figure 10 shows that the productivity of large firms is significantly higher than that of small firms in the drugs development sector (SIC 283). An important implication of Proposition 3 is that firms with higher productivity are better connected and more central in the R&D network. Hence, subsidizing collaborations involving large firms will not only foster the R&D activities of these more productive firms but also generate larger technology spillovers to other firms in the network. This complements the findings in Mindruta et al. (2016) who emphasize the importance of firm size in facilitating knowledge transfers and document a tendency for large firms to form R&D collaborations in the biopharmaceutical industry.

4.3. Sector-Specific Subsidy

In this subsection, we investigate how the effectiveness of the subsidy varies across different sub-sectors. The four largest sub-sectors in the drugs-development sector (SIC 283) in our sample are pharmaceutical preparations (SIC 2834) with 204 firms, biological products except diagnostic substances (SIC 2836) with 147 firms, in vitro and in vivo diagnostic substances (SIC 2835) with 41 firms, and medicinal chemicals and botanical products (SIC 2833) with 18 firms. Figure 11 plots the

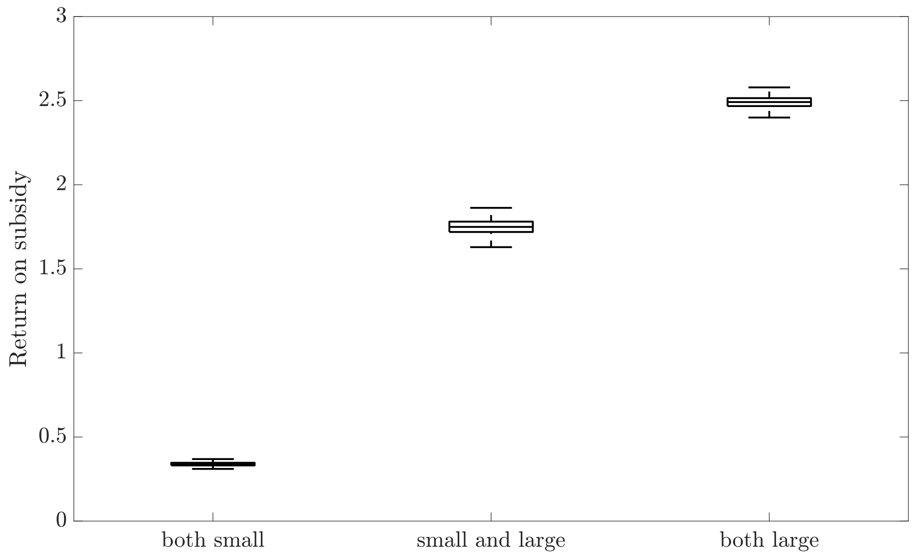


FIGURE 9. Return on subsidy (ROS) for collaborations between firms of different sizes.

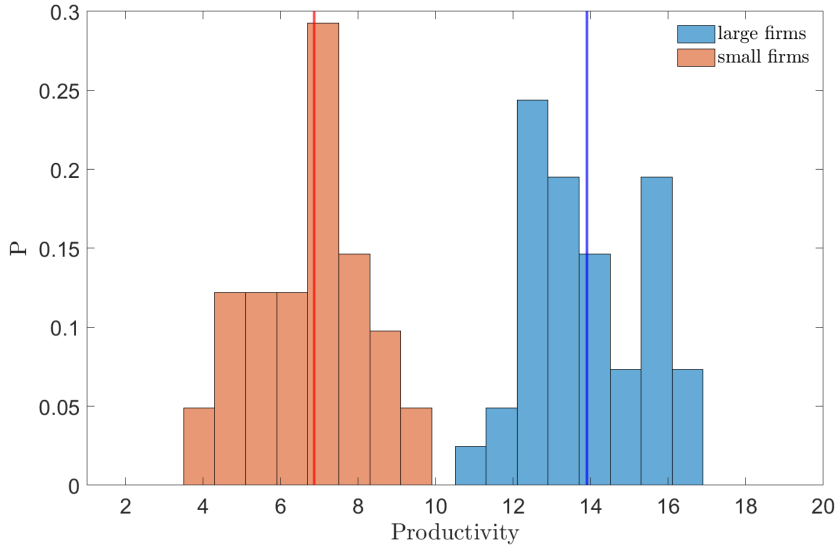


FIGURE 10. Distribution of productivity for large and small firms in the drugs development sector (SIC 283).

ROS for three different subsidy schemes that reduce the collaboration cost between firms (i) within SIC 2834, (ii) within other sub-sectors (SIC 2836, SIC 2835, and SIC 2833), and (iii) across SIC 2834 and other sub-sectors, by 10%, respectively. We find that the ROS is slightly higher subsidizing collaborations within SIC 2834 (with a median ROS of 1.86) than subsidizing collaborations within other sub-sectors (with a median ROS of 1.48) or collaborations across SIC 2834 and other sub-sectors (with a median ROS of 1.40). Based on the discussion in Section 4.2, the higher ROS from subsidizing collaborations within SIC 2834 is in line with the higher productivity of firms in that sub-sector as shown in Figure 12.

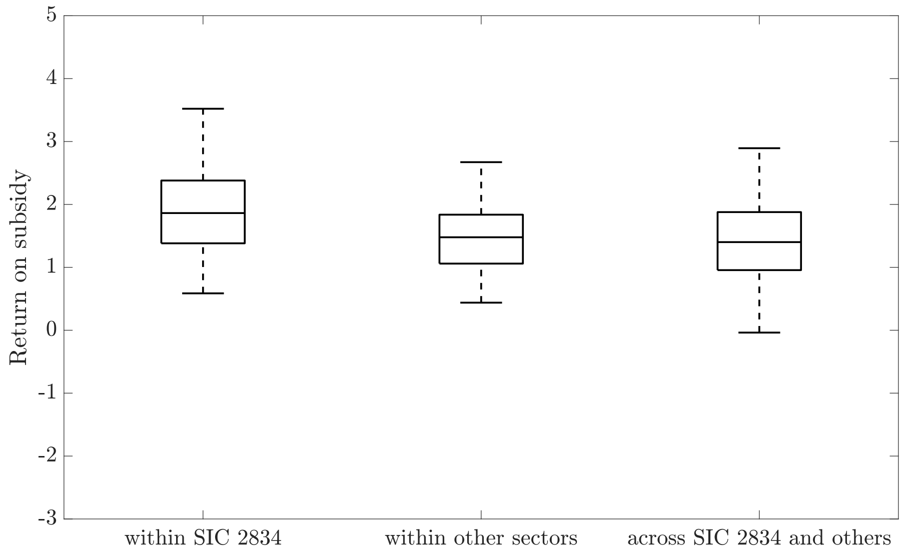


FIGURE 11. Return on a uniform R&D collaboration subsidy within and across sub-sectors.

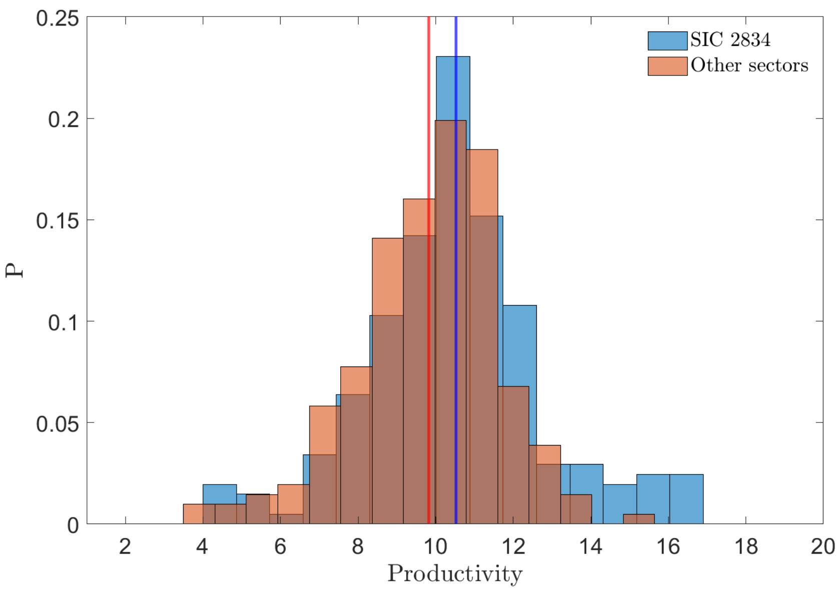


FIGURE 12. The distribution of productivity in the pharmaceutical preparations sector (SIC 2834) and the other drugs development sectors (SIC 2836, 2835, and 2833).

4.4. Subsidies Targeting Specific Collaborations

In this subsection, we consider a subsidy scheme where the government compensates the total collaboration cost of a specific pair of firms through a subsidy. We assume the collaboration between firms i and j is contingent on the availability of this subsidy (i.e., the link $a_{ij} = 1$ if and only if firms i and j receive this subsidy). In Table 4, we provide a ranking of the potential collaborations to be subsidized according to the ROS.

To calculate the ROS, we consider two cases in Table 4: the short-run ROS and the long-run ROS. To obtain the short-run ROS, we assume the other firms in the network can adjust their R&D efforts but cannot change their R&D collaborations in response to the subsidy received by firms i and j. To obtain the long-run ROS, we assume the other firms in the network can adjust their R&D efforts and collaborations according to the stochastic process in Definition 1 in response to the subsidy received by firms i and j.[24] Table 4 shows that the long-run ROS is constantly higher than the short-run ROS and the short-run and the long-run ROS rankings differ considerably. This highlights the importance of taking network endogeneity into account when evaluating the effectiveness of R&D subsidy policies.

From Table 4 we find that a subsidy for the collaboration between *Novartis*, a Swiss multinational pharmaceutical company, and *Pfizer*, an American multinational pharmaceutical and biotechnology corporation, would yield the highest long-run ROS of 5.17. This is more than two times higher than the ROS from the uniform subsidy in Section 4.1, which is around 2. This suggests that a targeted subsidy can be potentially much more effective than a uniform subsidy. The ranking in Table 4 further shows that higher ROS can be obtained from subsidizing R&D collaborations between more productive firms (the correlation between the long-run ROS and productivity is 0.78). This is consistent with the findings and discussion in Section 4.2. Our findings show that it is important to subsidize the "right firms" (i.e., more productive firms) for an R&D collaboration subsidy program to be effective (cf. e.g. König et al. 2022).

5. Conclusion

In this paper, we introduce a tractable framework to model the coevolution of networks and behavior, and we apply it to study the formation of R&D networks in which firms form R&D collaborations to benefit from technology spillovers while competing in the product market. We provide a complete equilibrium characterization and show that our model can reproduce the important patterns of observed networks in the real world. Moreover, the model can be conveniently estimated even in large networks based on a composite-likelihood function. Finally, our model is amenable to policy analysis as we demonstrated with the example of subsidizing R&D collaborations. While our policy analysis focuses on the stationary equilibrium, we leave a dynamic policy analysis that takes into account transitional dynamics toward the stationary state to future work. Other important avenues for future

24. Note that the ROS reported in Sections 4.1-4.3 refers to the long-run ROS.

TABLE 4. Key link analysis for firms in the drugs development sector (SIC 283).

Firm i	Firm j	Relative[a] Prod. i	Relative Prod. j	Degree (d_i)	Degree (d_j)	Short Run[b] ROS[c]	Short Run Rank	Long Run[b] ROS	Long Run Rank
Novartis	Pfizer	1.592	1.653	19	16	4.261	3	5.171	1
Merck	Pfizer	1.582	1.653	15	16	4.433	1	5.147	2
Amgen	Pfizer	1.526	1.653	14	16	4.057	4	4.972	3
GlaxoSmithKline	Novartis	1.509	1.592	14	19	4.371	2	4.941	4
Amgen	Merck	1.526	1.582	14	15	3.471	9	4.458	5
GlaxoSmithKline	Pfizer	1.509	1.653	14	16	3.597	7	4.449	6
Merck	Novartis	1.582	1.592	15	19	3.675	6	4.371	7
Merck	Johnson & Johnson	1.582	1.617	15	11	3.487	8	4.331	8
Novartis	Johnson & Johnson	1.592	1.617	19	11	3.376	10	4.325	9
Bristol-Myers Squibb	Pfizer	1.564	1.653	7	16	3.033	13	4.286	10
Amgen	Johnson & Johnson	1.526	1.617	14	11	3.181	12	4.176	11
Amgen	Novartis	1.526	1.592	14	19	3.361	11	4.145	12
GlaxoSmithKline	Johnson & Johnson	1.509	1.617	14	11	2.736	16	4.120	13
GlaxoSmithKline	Amgen	1.509	1.526	14	14	2.827	15	4.005	14
GlaxoSmithKline	Merck	1.509	1.582	14	15	2.983	14	3.967	15

[a] Relative productivity shows the firm's productivity relative to the average productivity of all firms in the sample.
[b] In the short run the network is assumed to be fixed, while in the long run the network rewires endogenously responding to the presence (and absence) of the link between firms i and j.
[c] The return on subsidy (ROS) is defined as the ratio of the expected welfare gain to the subsidy cost.

work would be to incorporate convex link formation costs, multi-product firms, and diminishing returns to scale in the marginal cost specification.

Due to the generality of the payoff function that we consider in this paper, we believe that our model, both from theoretical and empirical perspectives, can be applied to a variety of related contexts, where externalities can be modeled in the form of an endogenous network. Examples include peer effects in education, crime, risk sharing, scientific co-authorship, etc. (cf. Jackson and Zenou 2015). Our methodology can also be extended to study network games with local substitutes (Bramoullé and Kranton 2007), when we assume a negative sign for the local externality parameter in our payoff function.

References

Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.

Atmar, W. and Patterson, B. D. (1993). The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, 96(3):373–382.

Badev, A. (2021). Nash equilibria on (un) stable networks. *Econometrica*, 89(3):1179–1206.

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

Belderbos, R., Carree, M., and Lokshin, B. (2004). Cooperative R&D and firm performance. *Research Policy*, 33(10):1477–1492.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Besag, J. (1975). Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 24(3):179–195.

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

Bloom, N., Van Reenen, J. and Williams, H. (2019). A toolkit of policies to promote innovation. *Journal of Economic Perspectives*, 33(3):163–84.

Boguná, M. and Pastor-Satorras, R. (2003). Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112.

Bramoullé, Y. and Kranton, R. (2007). Public goods in networks. *Journal of Economic Theory*, 135(1):478–494.

Broekel, T. and Graf, H. (2012). Public research intensity and the structure of german R&D networks: A comparison of 10 technologies. *Economics of Innovation and New Technology*, 21(4):345–372.

Bronshtein, I. N., Semendyayev, K. A., Musiol, G., and Muehlig, H. (2013). *Handbook of Mathematics*. Springer.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Chandrasekhar, A. and Jackson, M. (2012). Tractable and consistent random graph models. *Available at SSRN 2150428*.

Cohen, L. (1994). When can government subsidize research joint ventures? Politics, economics, and limits to technology policy. *American Economic Review*, 84(2):159–163.

Cohen, W., Levin, R., and Mowery, D. (1987). Firm size and R&D intensity: A re-examination. *Journal of Industrial Economics*, 35(4):543–565.

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

Cournot, A. (1838). *Researches into the Mathematical Principles of the Theory of Wealth.* Trans. N.T. Bacon, New York: Macmillan.

Czarnitzki, D., Hussinger, K., and Schneider, C. (2015). R&D collaboration with uncertain intellectual property rights. *Review of Industrial Organization*, 46(2):183–204.

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

Daughety, A. F. (2008). *Cournot competition.* From The New Palgrave Dictionary of Economics, Second Edition, Edited by Steven N. Durlauf and Lawrence E. Blum.

Dawid, H. and Hellmann, T. (2014). The evolution of R&D networks. *Journal of Economic Behavior & Organization*, 105:158–172.

Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K.-T. and Van Reenen, J. (2016). Do tax incentives for research increase firm innovation? An RD design for R&D. National Bureau of Economic Research, Working Paper No. w22405.

Eureka Association (2021). Annual Report, Eureka, Innovation Beyond Borders. URL: www.eurekanetwork.org.

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

Friel, N. (2012). Bayesian inference for Gibbs random fields using composite likelihoods. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 1–8. IEEE.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, 169–193.

Gibson, M. A. and Bruck, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889.

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

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Grimmett, G. (2010). *Probability on graphs*. Cambridge University Press.

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., and Trajtenberg, M. (2000). Market value and patent citations: A first look. *National Bureau of Economic Research, Working Paper No. w7741*.

Hsieh, C.-S., König, M., and Liu, X. (2022). A structural model for the coevolution of networks and behavior. *Review of Economics and Statistics*, 104(2):355–367.

Hsieh, C.-s., Lee, L.-f., and Boucher, V. (2020). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Quantitative Economics*, 11(4): 1349–1390.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481).

Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, 106(2):265–295.

Jackson, M. O. and Zenou, Y. (2015). Games on networks. *Handbook of Game Theory with Economic Applications*, 4:95–163.

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

Kandori, M., Mailath, G., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56.

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

König, M. D., Liu, X., and Zenou, Y. (2019). R&D networks: Theory, empirics and policy implications. *Review of Economics and Statistics*, 101(3):476–491.

König, M. D., Storesletten, K., Song, Z., and Zilibotti, F. (2022). From imitation to innovation: Where is all that Chinese R&D going? *Econometrica*, 90(4):1615–1654.

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis*, Vol. 476.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–239.

Lychagin, S., Pinkse, J., Slade, M. E., and Van Reenen, J. (2016). Spillovers in space: Does geography matter? *Journal of Industrial Economics*, 64(2):295–335.

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

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.

Mindruta, D., Moeen, M. and Agarwal, R. (2016). A two-sided matching approach for partner selection and assessing complementarities in partners' attributes in inter-firm alliances. *Strategic Management Journal*, 37(1):206–231.

Monderer, D. and Shapley, L. (1996). Potential Games. *Games and Economic Behavior*, 14(1):124–143.

Norris, J. R. (1998). *Markov Chains*. Cambridge University Press.

OECD (2001). *Innovative Networks. Co-operation in National Innovation Systems*. OECD Publishing.

OECD (2008). *Open Innovation in Global Networks*. OECD Publishing.

Pauli, F., Racugno, W., and Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, 149–164.

Raftery, A. E., Lewis, S., Banfield, J. D., and Rafterv, A. E. (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, 4:115–130.

Ribatet, M., Cooley, D., and Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 813–845.

Rosenkopf, L. and Padula, G. (2008). Investigating the microstructure of network evolution: Alliance formation in the mobile communications industry. *Organization Science*, 19(5):669–687.

Rosenkopf, L. and Schilling, M. (2007). Comparing alliance network structure across industries: Observations and explanations. *Strategic Entrepreneurship Journal*, 1:191–209.

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

Singh, N. and Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *RAND Journal of Economics*, 15(4):546–554.

Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395.

Takalo, T., Tanayama, T., and Toivanen, O. (2013). Estimating the benefits of targeted R&D subsidies. *Review of Economics and Statistics*, 95(1):255–272.

Trajtenberg, M., Shiff, G., and Melamed, R. (2009). The "names game": Harnessing inventors, patent data for economic research. *Annals of Economics and Statistics*, (93/94):79–108.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.

Vives, X. (1999). *Oligopoly Pricing: Old ideas and new tools*. MIT Press.

Wang, D. and Tang, Q. (2006). Tail probabilities of randomly weighted sums of random variables with dominated variation. *Stochastic Models*, 22(2):253–272.

Watts, A. (2001). A dynamic model of network formation. *Games and Economic Behavior*, 34(2):331–341.

Westbrock, B. (2010). Natural concentration in industrial research collaboration. *RAND Journal of Economics*, 41(2):351–371.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86.

Appendix

Appendix A: Proofs

We first prove that the potential function has the property that the marginal profit of a firm from adding or removing a link is equivalent to the difference in the potential function from adding or removing a link. Similarly, the marginal profit of a firm from changing its R&D effort level is equivalent to the change of the potential function.

Proof of Proposition 1. The potential $\Phi(\mathbf{y}, G)$ has the property that $\Phi(\mathbf{y}, G \oplus (i,j)) - \Phi(\mathbf{y}, G) = \rho y_i y_j - \zeta = \pi_i(\mathbf{y}, G \oplus (i,j)) - \pi_i(\mathbf{y}, G)$, and similarly, $\Phi(\mathbf{y}, G \ominus (i,j)) - \Phi(\mathbf{y}, G) = \zeta - \rho y_i y_j = \pi_i(\mathbf{y}, G \ominus (i,j)) - \pi_i(\mathbf{y}, G)$ for any $\mathbf{y} \in \mathcal{Y}^n$ and $G \in \mathcal{G}^n$. From the properties of $\pi_i(\mathbf{y}, G)$ it also follows that $\Phi(y_i', \mathbf{y}_{-i}, G) - \Phi(y_i, \mathbf{y}_{-i}, G) = \pi_i(y_i', \mathbf{y}_{-i}, G) - \pi_i(y_i, \mathbf{y}_{-i}, G)$. $\square$

We next introduce some definitions and notations that allow us to formally characterize the stochastic process in Definition 1. Let $\mathcal{F}$ denote the smallest σ-algebra generated by $\sigma(\boldsymbol{\omega}_t : t \in \mathbb{R}_+)$. The filtration is the non-decreasing family of sub-σ-algebras $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ on the measure space $(\Omega, \mathcal{F})$, $\Omega = \mathcal{Y}^n \times \mathcal{G}^n$, with the property that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_t \subseteq \cdots \subseteq \mathcal{F}$. The probability space is given by the triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P} \colon \mathcal{F} \to [0,1]$ is the probability measure satisfying $\int_\Omega \mathbb{P}(d\boldsymbol{\omega}) = 1$. As we will see below in Theorem 1 the sequence of states $(\boldsymbol{\omega}_t)_{t \in \mathbb{R}_+}$, $\boldsymbol{\omega}_t \in \Omega$, induces an irreducible and positive recurrent (i.e., ergodic) time homogeneous Markov chain.

The one-step transition probability matrix $\mathbf{P}^\vartheta(t) : \Omega^2 \to [0,1]^{|\Omega|^2}$ has elements which determine the probability of a transition from a state $\boldsymbol{\omega} \in \Omega$ to a state $\boldsymbol{\omega}' \in \Omega$ in a small time interval $[t, t+\Delta t)$ of length Δt given by $\mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = \boldsymbol{\omega}' | \mathcal{F}_t = \sigma(\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_t = \boldsymbol{\omega})) = \mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = \boldsymbol{\omega}' | \boldsymbol{\omega}_t = \boldsymbol{\omega}) = y^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}')\Delta t + o(\Delta t)$ if $\omega' \neq \omega$ and $\mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = \boldsymbol{\omega} | \mathcal{F}_t = \sigma(\boldsymbol{\omega}_0, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_t = \boldsymbol{\omega})) = \mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = \boldsymbol{\omega} | \boldsymbol{\omega}_t = \boldsymbol{\omega}) = 1 + x^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega})\Delta t + o(\Delta t)$, where $y^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is the transition rate from state $\boldsymbol{\omega}$ to state $\boldsymbol{\omega}'$ and $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$ (see Theorem 2.8.2 in Norris (1998)). From the stochastic process of Definition 1 we see that the transition rate matrix (or infinitesimal generator) $\mathbf{Q}^\vartheta = (y^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}'))_{\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega}$ of the Markov chain has the elements

$$x^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') = \begin{cases} \chi \frac{e^{\vartheta \pi_i(y', \mathbf{y}_{-i}, G)}}{\int_{\mathcal{Y}} e^{\vartheta \pi_i(y'', \mathbf{y}_{-i}, G)} dy''} & \text{if } \boldsymbol{\omega}' = (y', \mathbf{y}_{-i}, G) \text{ and } \boldsymbol{\omega} = (\mathbf{y}, G), \\ \xi \frac{e^{\vartheta \Phi(\mathbf{y}, G \oplus (i,j))}}{e^{\vartheta \Phi(\mathbf{y}, G \oplus (i,j))} + e^{\vartheta \Phi(\mathbf{y}, G)}} & \text{if } \boldsymbol{\omega}' = (\mathbf{y}, G \oplus (i,j)) \text{ and } \boldsymbol{\omega} = (\mathbf{y}, G), \\ \xi \frac{e^{\vartheta \Phi(\mathbf{y}, G \ominus (i,j))}}{e^{\vartheta \Phi(\mathbf{y}, G \ominus (i,j))} + e^{\vartheta \Phi(\mathbf{y}, G)}} & \text{if } \boldsymbol{\omega}' = (\mathbf{y}, G \ominus (i,j)) \text{ and } \boldsymbol{\omega} = (\mathbf{y}, G), \\ -\sum_{\boldsymbol{\omega}' \neq \boldsymbol{\omega}} x^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') & \text{if } \boldsymbol{\omega}' = \boldsymbol{\omega}, \\ 0 & \text{otherwise,} \end{cases} \tag{A.1}$$

with $\sum_{\boldsymbol{\omega}' \in \Omega} x^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') = 0$. The transition rate matrix satisfies the Chapman-Kolmogorov forward equation $\frac{d}{dt}\mathbf{P}^\vartheta(t) = \mathbf{P}^\vartheta(t)\mathbf{Q}^\vartheta$ so that $\mathbf{P}^\vartheta(t) = \mathbf{I}_{|\Omega|} + \mathbf{Q}^\vartheta \Delta t + o(\Delta t)$ (Norris 1998). Conversely, we have that $\mathbf{Q}^\vartheta = \lim_{\Delta t \to 0} \frac{\mathbf{P}^\vartheta(t+\Delta t) - \mathbf{I}_{|\Omega|}}{\Delta t}$.As the Markov chain is time homogeneous, the transition rates are independent of time. The stationary distribution $\mu^\vartheta : \Omega \to [0,1]$ is then the solution to $\mu^\vartheta \mathbf{P}^\vartheta = \mu^\vartheta$, or equivalently $\mu^\vartheta \mathbf{Q}^\vartheta = \mathbf{0}$ (Norris 1998).

We are now in the position to give a proof of Theorem 1 where it is shown that the stationary distribution can be characterized by a Gibbs measure (Grimmett 2010).

Proof of Theorem 1. First, note from Equation (A.1) that $y^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') > 0$ for any $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$ and finite ϑ, so that there is a positive probability of a transition from any state $\boldsymbol{\omega}$ to any other state $\boldsymbol{\omega}'$, and there can be no absorbing state. The generator matrix $\mathbf{Q}^\vartheta = (y^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}'))_{\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega}$ is therefore irreducible. Moreover, for an irreducible Markov chain on a finite state space Ω?all states are positive recurrent. The Markov chain then is ergodic and has a unique stationary distribution (Norris 1998).

The stationary distribution solves $\mu^\vartheta \mathbf{Q}^\vartheta = \mathbf{0}$ with the transition rates matrix $\mathbf{Q}^\vartheta$ of Equation (A.1). This equation is satisfied when the probability distribution μ^ϑ satisfies the following detailed balance condition (Norris 1998)

$$\forall \boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega : \quad \mu^\vartheta(\boldsymbol{\omega}) y^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') = \mu^\vartheta(\boldsymbol{\omega}') y^\vartheta(\boldsymbol{\omega}', \boldsymbol{\omega}). \tag{A.2}$$

Observe that the detailed balance condition is trivially satisfied if $\boldsymbol{\omega}'$ and $\boldsymbol{\omega}$ differ in more than one link or more than one R&D effort level. Hence, we consider only the case of link creation $G' = G \oplus (i,j)$ (and removal $G' = G \ominus (i,j)$) or an adjustment in quantity $y_i' \neq y_i$ for some $i \in \mathcal{N}$. For the case of link creation with a transition from $\boldsymbol{\omega} = (\mathbf{y}, G)$ to $\boldsymbol{\omega}' = (\mathbf{y}, G \oplus (i,j))$ we can write the detailed balance condition as follows

$$\frac{1}{\mathscr{Z}_\theta} e^{\vartheta \Phi(\mathbf{y}, G)} \frac{e^{\vartheta \Phi(\mathbf{y}, G \oplus (i,j))}}{e^{\vartheta \Phi(\mathbf{y}, G \oplus (i,j))} + e^{\vartheta \Phi(\mathbf{y}, G)}} \xi = \frac{1}{\mathscr{Z}_\theta} e^{\vartheta \Phi(\mathbf{y}, G \oplus (i,j))} \frac{e^{\vartheta \Phi(\mathbf{y}, G)}}{e^{\vartheta \Phi(\mathbf{y}, G)} + e^{\vartheta \Phi(\mathbf{y}, G \oplus (i,j))}} \xi.$$

This equality is trivially satisfied. A similar argument holds for the removal of a link with a transition from $\boldsymbol{\omega} = (\mathbf{y}, G)$ to $\boldsymbol{\omega}' = (\mathbf{y}, G \ominus (i,j))$ where the detailed balance condition reads

$$\frac{1}{\mathscr{Z}_\theta} e^{\vartheta \Phi(\mathbf{y}, G)} \frac{e^{\vartheta \Phi(\mathbf{y}, G \ominus (i,j))}}{e^{\vartheta \Phi(\mathbf{y}, G \ominus (i,j))} + e^{\vartheta \Phi(\mathbf{y}, G)}} \xi = \frac{1}{\mathscr{Z}_\theta} e^{\vartheta \Phi(\mathbf{y}, G \ominus (i,j))} \frac{e^{\vartheta \Phi(\mathbf{y}, G)}}{e^{\vartheta \Phi(\mathbf{y}, G)} + e^{\vartheta \Phi(\mathbf{y}, G \ominus (i,j))}} \xi.$$

For a change in the R&D effort level with a transition from $\boldsymbol{\omega} = (y_i, \mathbf{y}_{-i}, G)$ to $\boldsymbol{\omega}' = (y_i', \mathbf{y}_{-i}, G)$ we get for the following detailed balance condition

$$\frac{1}{\mathscr{Z}_\theta} e^{\vartheta\Phi(y_i,\mathbf{y}_{-i},G)} \frac{e^{\vartheta\pi_i(y_i',\mathbf{y}_{-i},G)}}{\int_{\mathcal{Y}} e^{\vartheta\pi_i(y',\mathbf{y}_{-i},G)}dy'}\chi = \frac{1}{\mathscr{Z}_\theta} e^{\vartheta\Phi(y_i',\mathbf{y}_{-i},G)} \frac{e^{\vartheta\pi_i(y_i,\mathbf{y}_{-i},G)}}{\int_{\mathcal{Y}} e^{\vartheta\pi_i(y',\mathbf{y}_{-i},G)}dy'}\chi.$$

This can be written as $e^{\vartheta(\Phi(y_i,\mathbf{y}_{-i},G)-\Phi(y_i',\mathbf{y}_{-i},G))} = e^{\vartheta(\pi_i(y_i,\mathbf{y}_{-i},G)-\pi_i(y_i',\mathbf{y}_{-i},G))}$, which is satisfied since we have for the potential $\Phi(y_i, \mathbf{y}_{-i}, G) - \Phi(y_i', \mathbf{y}_{-i}, G) = \pi_i(y_i, \mathbf{y}_{-i}, G) - \pi_i(y_i', \mathbf{y}_{-i}, G)$. Hence, the probability measure μ^ϑ satisfies a detailed balance condition of Equation (A.2) and therefore is the stationary distribution of the Markov chain with transition rate matrix $\mathbf{Q}^\vartheta$. □

We next state a useful lemma that will be needed in the proofs that follow.

Lemma A.1. *Consider a binary sequence $h_1, h_2, \ldots, h_n$ with elements $h_i \in \{0, 1\}$ and a real sequence $c_1, c_2, \ldots, c_n$ with elements $c_i \in \mathbb{R}$ for $i = 1, \ldots, n$ and $n \geq 1$. Let $\mathcal{H}^n$ be the set of all binary sequences $\mathbf{h} = (h_1, \ldots, h_n)$ with n elements. Then we have that*

$$\sum_{\mathbf{h}\in\mathcal{H}^n} e^{\sum_{i=1}^n h_i c_i} = \prod_{i=1}^n \sum_{h_i\in\{0,1\}} e^{h_i c_i}. \tag{A.3}$$

Proof of Lemma A.1. We give proof by induction. For the induction, the basis considers $n = 2$ (the case of $n = 1$ is trivially true). Then $\mathcal{H}^2 = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$, and we have that

$$\sum_{\mathbf{h}\in\mathcal{H}^2} e^{\sum_{i=1}^2 h_i c_i} = 1 + e^{c_1} + e^{c_2} + e^{c_1+c_2}.$$

On the other hand, we have that

$$\prod_{i=1}^2 \sum_{h_i\in\{0,1\}} e^{h_i c_i} = \prod_{i=1}^2 (1 + e^{c_i}) = 1 + e^{c_1} + e^{c_2} + e^{c_1+c_2}.$$

Next, for the induction step, assume that Equation (A.3) holds for some $n \geq 2$. Note that all binary sequences $\mathbf{h} \in \mathcal{H}^{n+1}$ can be constructed from a binary sequence $\mathbf{h} \in \mathcal{H}^n$ with one additional element, h_{n+1}, added to the sequence $h_1, \ldots, h_n$ where h_{n+1} takes on the two possible values 0 or 1. Hence, we can write

$$\begin{aligned}
\sum_{\mathbf{h}\in\mathcal{H}^{n+1}} e^{\sum_{i=1}^{n+1} h_i c_i} &= \sum_{\mathbf{h}\in\mathcal{H}^{n+1}} \prod_{i=1}^{n+1} e^{h_i c_i} \\
&= \sum_{\mathbf{h}\in\mathcal{H}^n} \prod_{i=1}^{n} e^{h_i c_i} + \sum_{\mathbf{h}\in\mathcal{H}^n} \prod_{i=1}^{n} e^{h_i c_i} e^{c_{n+1}} \\
&= \sum_{\mathbf{h}\in\mathcal{H}^n} \prod_{i=1}^{n} e^{h_i c_i} (1 + e^{c_{n+1}}) \\
&= \prod_{i=1}^{n} \sum_{h_i\in\{0,1\}} e^{h_i c_i} (1 + e^{c_{n+1}}) \\
&= \prod_{i=1}^{n+1} \sum_{h_i\in\{0,1\}} e^{h_i c_i},
\end{aligned}$$

where we have used the induction hypothesis that Equation (A.3) holds for n. This concludes the proof. □

For the proof of Proposition 2 we first introduce some notation and definitions. The *partition function* is defined as

$$\mathscr{Z}_\vartheta = \sum_{G\in\mathcal{G}^n} \int_{\mathcal{Y}^n} e^{\vartheta\Phi(\mathbf{y},G)} d\mathbf{y}, \tag{A.4}$$

so that we can write $\mu^\vartheta(\mathbf{y}, G) = \frac{1}{\mathscr{Z}_\vartheta} e^{\vartheta\Phi(\mathbf{y},G)}$ for any $\mathbf{y} \in \mathcal{Y}^n$ and $G \in \mathcal{G}^n$. We also introduce the *Hamiltonian*, defined by $\mathscr{H}_\vartheta(\mathbf{y}) \equiv \frac{1}{\vartheta} \ln\left(\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)}\right)$, which allows us to write the partition function more compactly as $\mathscr{Z}_\vartheta = \int_{\mathcal{Y}^n} e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y})} d\mathbf{y}$.

Proof of Proposition 2. We first give proof of part (ii) of the proposition. We have that $\mu^{\vartheta}(\mathbf{y},G) = \mu^{\vartheta}(G|\mathbf{y})\mu^{\vartheta}(\mathbf{y})$. Observe that the potential of Equation (2) can be written as

$$\Phi(\mathbf{y},G) = \underbrace{\sum_{i=1}^{n}\left(\eta_i - \frac{1}{2}y_i - \frac{\lambda}{2}\sum_{j\neq i}^{n} y_j\right) y_i}_{\varphi(\mathbf{y})} + \sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}\underbrace{(\rho y_i y_j - \zeta)}_{\sigma_{ij}} = \varphi(\mathbf{y}) + \sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}\sigma_{ij}. \tag{A.5}$$

We then have that $e^{\vartheta\Phi(\mathbf{y},G)} = e^{\vartheta\varphi(\mathbf{y})}e^{\vartheta\sum_{i<j}^{n} a_{ij}\sigma_{ij}}$, where only the last factor on the RHS is network dependent. Observing that the sequence $(a_{ij})_{1\leq i<j\leq n} = (a_{12}, a_{13}, \ldots, a_{n-1,n})$ is a binary sequence as in Lemma A.1, we can use the fact that for any constant, symmetric $\sigma_{ij} = \sigma_{ji}$, $1 \leq i,j \leq n$, we can write

$$\sum_{G\in\mathcal{G}^n} e^{\vartheta\sum_{i<j}^{n} a_{ij}\sigma_{ij}} = \prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1 + e^{\vartheta\sigma_{ij}}\right). \tag{A.6}$$

From Equation (A.6) we then obtain

$$\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)} = e^{\vartheta\varphi(\mathbf{y})}\prod_{i<j}^{n}\left(1 + e^{\vartheta\sigma_{ij}}\right) = \prod_{i=1}^{n} e^{\vartheta\left(\eta_i - \frac{1}{2}y_i - \frac{\lambda}{2}\sum_{j\neq i}^{n} y_j\right)y_i}\prod_{i<j}^{n}\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right). \tag{A.7}$$

We can use Equation (A.7) to compute the marginal distribution

$$\begin{aligned}
\mu^{\vartheta}(\mathbf{y}) &= \frac{1}{\mathscr{Z}_{\vartheta}}\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)} \\
&= \frac{1}{\mathscr{Z}_{\vartheta}}\prod_{i=1}^{n} e^{\vartheta\left(\eta_i - \frac{1}{2}y_i - \frac{\lambda}{2}\sum_{j\neq i} y_j\right)y_i}\prod_{i<j}^{n}\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right) \\
&= \frac{1}{\mathscr{Z}_{\vartheta}} e^{\vartheta\sum_{i=1}^{n}\left(\eta_i - \frac{1}{2}y_i - \frac{\lambda}{2}\sum_{j\neq i} y_j\right)y_i} e^{\sum_{i<j}^{n}\ln\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right)} \\
&= \frac{1}{\mathscr{Z}_{\vartheta}} e^{\vartheta\mathscr{H}_{\vartheta}(\mathbf{y})},
\end{aligned} \tag{A.8}$$

where we have introduced the Hamiltonian

$$\mathscr{H}_{\vartheta}(\mathbf{y}) \equiv \sum_{i=1}^{n}\left(\eta_i y_i - \frac{1}{2}y_i^2 + \sum_{j>i}^{n}\left(\frac{1}{\vartheta}\ln\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right) - \lambda y_i y_j\right)\right). \tag{A.9}$$

With the marginal distribution from Equation (A.8) and the potential in Equation (A.5) we then can write the conditional distribution as

$$\begin{aligned}
\mu^{\vartheta}(G|\mathbf{y}) = \frac{\mu^{\vartheta}(\mathbf{y},G)}{\mu^{\vartheta}(\mathbf{y})} = \frac{e^{\vartheta\Phi(\mathbf{y},G)}}{\sum_{G'\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G')}} &= \frac{e^{\varphi(\mathbf{y})}e^{\vartheta\sum_{i<j}^{n} a_{ij}(\rho y_i y_j - \zeta)}}{e^{\varphi(\mathbf{y})}\prod_{i<j}\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right)} \\
&= \frac{e^{\vartheta\sum_{i<j}^{n} a_{ij}(\rho y_i y_j - \zeta)}}{\prod_{i<j}\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right)} \\
&= \prod_{i<j}\frac{e^{\vartheta a_{ij}(\rho y_i y_j - \zeta)}}{1 + e^{\vartheta(\rho y_i y_j - \zeta)}} \\
&= \prod_{i<j}\left(\frac{e^{\vartheta(\rho y_i y_j - \zeta)}}{1 + e^{\vartheta(\rho y_i y_j - \zeta)}}\right)^{a_{ij}}\left(1 - \frac{e^{\vartheta(\rho y_i y_j - \zeta)}}{1 + e^{\vartheta(\rho y_i y_j - \zeta)}}\right)^{1-a_{ij}} \\
&= \prod_{i<j} p^{\vartheta}(y_i, y_j)^{a_{ij}}\left(1 - p^{\vartheta}(y_i, y_j)\right)^{1-a_{ij}}.
\end{aligned} \tag{A.10}$$

Hence, we obtain the likelihood of an inhomogeneous random graph with a link probability

$$p^{\vartheta}(y_i, y_j) = \frac{e^{\vartheta(\rho y_i y_j - \zeta)}}{1 + e^{\vartheta(\rho y_i y_j - \zeta)}}.$$

See also Supplementary Appendix B for further details about inhomogeneous random graphs.

We next give proof of part (i) of the proposition. The conditional probability $\mu^{\vartheta}(\mathbf{y}|G)$ of the R&D effort profile $\mathbf{y}$ given the network G is

$$\mu^{\vartheta}(\mathbf{y}|G) = \frac{\mu^{\vartheta}(\mathbf{y},G)}{\mu^{\vartheta}(G)} = \frac{e^{\vartheta\left(\boldsymbol{\eta}^{\top}\mathbf{y} - \frac{1}{2}\mathbf{y}\mathbf{M}(G)\mathbf{y}\right)}}{\int_{\mathcal{Y}^n} e^{\vartheta\left(\boldsymbol{\eta}^{\top}\mathbf{y}'^{\top} - \frac{1}{2}\mathbf{y}'^{\top}\mathbf{M}(G)\mathbf{y}'\right)}d\mathbf{y}'}, \tag{A.11}$$

where $\mathbf{M} = \mathbf{I}_n + \lambda\mathbf{B} - \rho\mathbf{A}$. Using the Gaussian integral formula (Bronshtein et al. 2013), we can write the denominator in Equation (A.11) as

$$\int_{\mathcal{Y}^n} \exp\left\{\vartheta\left(\boldsymbol{\eta}^\top\mathbf{y}' - \frac{1}{2}\mathbf{y}'^\top\mathbf{M}(G)\mathbf{y}'\right)\right\} d\mathbf{y}' = \left(\frac{2\pi}{\vartheta}\right)^{\frac{n}{2}} |\mathbf{M}(G)|^{-\frac{1}{2}} \exp\left\{\frac{\vartheta}{2}\boldsymbol{\eta}^\top\mathbf{M}(G)^{-1}\boldsymbol{\eta}\right\},$$

and hence

$$\mu^\vartheta(\mathbf{y}|G) = \left(\frac{2\pi}{\vartheta}\right)^{-\frac{n}{2}} |\mathbf{M}(G)|^{\frac{1}{2}} \exp\left\{-\frac{\vartheta}{2}\left(\mathbf{y} - \mathbf{M}(G)^{-1}\boldsymbol{\eta}\right)^\top \mathbf{M}(G)(\mathbf{y} - \mathbf{M}(G)^{-1}\boldsymbol{\eta})\right\},$$

which implies that the R&D effort $\mathbf{y}$, conditional on the R&D network G, follows a Gaussian normal density function with mean $\mathbf{M}(G)^{-1}\boldsymbol{\eta}$ and variance $\frac{1}{\vartheta}\mathbf{M}(G)^{-1}$. □

We next provide the proof of Proposition 3 which is concerned with the case of large ϑ and the characterization of the stochastically stable states. In Proposition D.1 in Supplementary Appendix D we provide an analogous characterization for the special case of homogenous firms.

Proof of Proposition 3. We first show that the networks G in support of the stationary distribution $\mu^\vartheta(\mathbf{y}, G)$ in the limit of vanishing noise $\vartheta \to \infty$ are nested split graphs. Before proceeding with the proof we give an alternative definition of nested split graphs following Mahadev and Peled (1995), which are also known as *threshold graphs*.

DEFINITION A.1 (Mahadev and Peled (1995)). *A graph G is a nested split graph if for every node $i \in \mathcal{N}$ there exists a real weight y_i and a real threshold τ such that nodes i and j are linked if and only if $y_i + y_j \geq \tau$.*

In the limit $\vartheta \to \infty$ the conditional probability of the network G can be written as $\mu^*(G|\mathbf{y}) = \lim_{\vartheta\to\infty} \mu^\vartheta(G|\mathbf{y}) = \prod_{i<j}^n \mathbb{1}_{\{\rho y_i y_j > \zeta\}}^{a_{ij}} \mathbb{1}_{\{\rho y_i y_j < \zeta\}}^{1-a_{ij}}$. Assume that G is a stochastically stable network, that is for $G \in \Omega^*$, we must have that $\mu^*(\mathbf{y}, G) = \lim_{\vartheta\to\infty} \mu^\vartheta(\mathbf{y}, G) > 0$. Since, $\mu^*(\mathbf{y}, G) = \mu^*(G|\mathbf{y})\mu^*(\mathbf{y})$ this implies that $\mu^*(G|\mathbf{y}) > 0$. It follows that $\rho y_i y_j > \zeta$ for all $a_{ij} = 1$ and $\rho y_i y_j < \zeta$ for all $a_{ij} = 0$. We then define the weights $y_i \equiv \log y_i$, $y_j \equiv \log y_j$ and a threshold $\tau \equiv \log\zeta - \log\rho$, and conclude that G is a nested split graph as defined in Definition A.1.

Moreover, the R&D effort distribution is given by $\mu^\vartheta(\mathbf{y}) = \frac{1}{\mathscr{Z}_\vartheta}\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)} = \frac{1}{\mathscr{Z}_n^\vartheta} e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y})}$, where the Hamiltonian is given by Equation (A.9). The R&D effort profile that maximizes the Hamiltonian can be found from the FOC, $\frac{\partial\mathscr{H}_\vartheta}{\partial y_i} = 0$, from which we get

$$y_i = \eta_i + \sum_{j\neq i}^n \left(\frac{\rho}{2}\left(1 + \tanh\left(\frac{\vartheta}{2}(\rho y_i y_j - \zeta)\right)\right) - \lambda\right) y_j.$$

Taking the limit $\vartheta \to \infty$ and noting that

$$\lim_{\vartheta\to\infty} \frac{1}{2}\left(1 + \tanh\left(\frac{\vartheta}{2}(\rho y_i y_j - \zeta)\right)\right) = \begin{cases} 1, & \text{if } \rho y_i y_j > \zeta, \\ 0, & \text{if } \rho y_i y_j < \zeta, \end{cases}$$

we thus obtain

$$y_i = \eta_i + \sum_{j\neq i}^n y_j \left(\rho\mathbb{1}_{\{y_i y_j > \frac{\zeta}{\rho}\}} - \lambda\right). \tag{A.12}$$

Note that for any profile of R&D effort levels $\mathbf{y}$ there exists a unique nested split graph with adjacency matrix $\mathbf{A} = (a_{ij})_{1\leq i,j,n}$ whose elements are given by $a_{ij} = \mathbb{1}_{\{y_i y_j > \zeta/\rho\}}$. Then we can write Equation (A.12) as follows

$$y_i = \eta_i + \rho\sum_{j\neq i}^n a_{ij} y_j - \lambda\sum_{j\neq i}^n y_j. \tag{A.13}$$

Moreover, assume that $\eta_i > \eta_j$, then we want to show that $y_j > y_i$ for the R&D effort profile $\mathbf{y}$ solving Equation (A.12). For this purpose we consider the iteration $(\mathbf{y}_t)_{t=0}^\infty$ with updates

$$y_{i,t+1} = f_i(\mathbf{y}_t) \equiv \max\left\{0, \eta_i + \rho\sum_{j\neq i}^n y_{j,t}\mathbb{1}_{\{y_{i,t} y_{j,t} > \frac{\zeta}{\rho}\}} - \lambda\sum_{j\neq i}^n y_j\right\}, \tag{A.14}$$

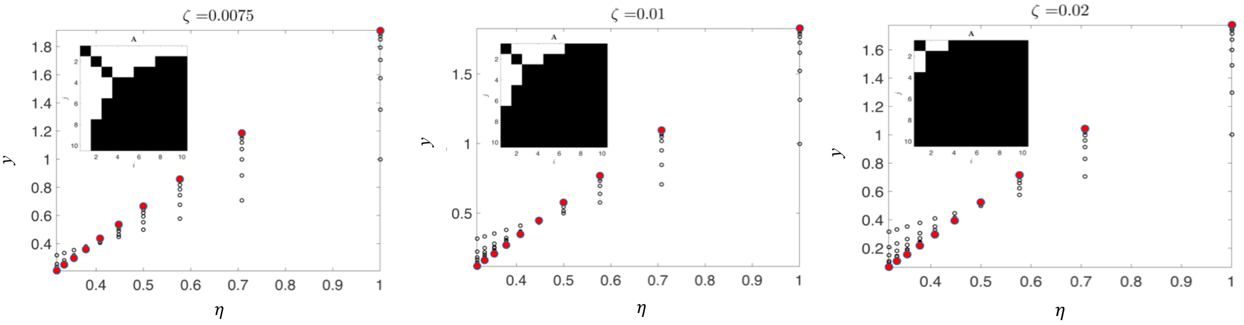


FIGURE A.1. The R&D effort iteration of Equation (A.14) over the firm fixed effects η_i. Filled circles indicate the fixed points. The insets show the adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j,n}$ (where ones are indicated with white squares and zeros are indicated with black squares) with elements are given by $a_{ij} = \mathbb{1}_{\{y_i y_j > \zeta/\rho\}}$ (see also Figure 4), where the vector $\mathbf{y}$ is the fixed point of Equation (A.14). The panels from left to right correspond to increasing linking costs $\zeta \in \{0.0075, 0.01, 0.02\}$. The parameters used are $n = 10$, $\lambda = 0.06$, $\rho = 0.02$ and $\boldsymbol{\eta} = (1.00, 0.71, 0.58, 0.50, 0.45, 0.41, 0.38, 0.35, 0.33, 0.32)^\top$.

starting from the initial vector $\mathbf{y}_0 = (0, \ldots, 0)^\top$. We then observe that the map $f_i : \mathbb{R}_+ \to \mathbb{R}_+$ is η-order preserving. That is, if $\eta_i > \eta_j$ and $y_{i,t} > y_{j,t}$, then also $y_{i,t+1} > y_{j,t+1}$. To show this, we proceed by induction. For the induction basis consider $t = 0$. Then $y_{i,1} = \eta_i$ for all $i = 1, \ldots, n$, and the claim follows. Next, consider the induction step, assuming that the claim holds for some $t > 0$. Then

$$y_{i,t+1} - y_{j,t+1} = f_i(\mathbf{y}_t) - f_j(\mathbf{y}_t) = \lambda(\eta_i - \eta_j) + \rho \sum_{k \in \mathcal{N}_{i,t} \backslash \mathcal{N}_{j,t}} y_{k,t} + \lambda(y_{i,t} - y_{j,t}) > 0,$$

where we have used the fact that the condition $y_{i,t} y_{j,t} > \frac{\zeta}{\rho}$ for i and j being linked represents a nested split graph, and for such a graph if $y_{i,t} > y_{j,t}$ (so that $d_{i,t} > d_{j,t}$) then $\mathcal{N}_{j,t} \subset \mathcal{N}_{i,t}$. Hence, for all t, the claim holds, and in particular, taking the limit as $t \to \infty$ it holds for the fixed point $\mathbf{y}^* = \lim_{t \to \infty} \mathbf{y}_t$ solving Equation (A.12).

□

Figure A.1 shows the R&D effort iteration of Eq. (A.14) over the firm fixed effects η_i together with the adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j,n}$ (see also Figure 4) whose elements are given by $a_{ij} = \mathbb{1}_{\{y_i y_j > \zeta/\rho\}}$ and the vector $\mathbf{y}$ is the fixed point of Equation (A.14). We observe that firms with higher η_i also have higher R&D effort levels. Moreover, the corresponding adjacency matrix is stepwise, characterizing a nested split graph (see also Supplementary Appendix B), and becomes increasingly sparse with increasing linking costs ζ.

We next provide the proof identifying the welfare maximizing state.

Proof of Proposition 4. Welfare can be written as

$$W(\mathbf{y}) = \sum_{i=1}^n \eta_i y_i - \frac{1-\psi}{2} \sum_{i=1}^n y_i^2 - \frac{\lambda}{2} \sum_{i=1}^n \sum_{j \neq i}^n y_i y_j + \sum_{i=1}^n \sum_{j \neq i}^n a_{ij}(\rho y_i y_j - \zeta).$$

The only network dependent part in $W(\mathbf{y}, G)$ is the last term $\sum_{i=1}^n \sum_{j \neq i}^n a_{ij}(\rho y_i y_j - \zeta)$. For a given R&D effort vector $\mathbf{y}$, the network that maximizes this term is a nested split graph G (see Definition A.1 and Supplementary Appendix B) where $ij \in G$ if and only if $\rho y_i y_j > \zeta$.In such a network, the optimal R&D effort levels satisfy the following necessary first-order conditions

$$y_i = f_i(\mathbf{y}) \equiv \max\left\{0, \frac{\eta_i}{1-\psi} - \frac{\lambda}{1-\psi} \sum_{j \neq i}^n y_j + \frac{2\rho}{1-\psi} \sum_{j \neq i}^n y_j \mathbb{1}_{\{\rho y_i y_j > \zeta\}}\right\}. \tag{A.15}$$

We can compare this to the equilibrium R&D effort levels of Equation (A.12), which were given by

$$y_i = g_i(\mathbf{y}) \equiv \max\left\{0, \eta_i - \lambda \sum_{j \neq i}^n y_j + \rho \sum_{j \neq i}^n y_j \mathbb{1}_{\{\rho y_i y_j > \zeta\}}\right\}. \tag{A.16}$$

We have for any $\mathbf{y} \in \mathcal{Y}^n$ that $f_i(\mathbf{y}) > g_i(\mathbf{y})$. This is because

$$f_i(\mathbf{y}) - g_i(\mathbf{y}) = \left(\frac{1}{1-\psi} - 1\right)\left(\eta_i - \lambda \sum_{j \neq i}^n y_j\right) + \left(\frac{2}{1-\psi} - 1\right) \rho \sum_{j \neq i}^n y_j \mathbb{1}_{\{\rho y_i y_j > \zeta\}} \geq 0.$$

Next, consider the differential equations $\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}) - \mathbf{y}$ and $\frac{d\mathbf{x}}{dt} = \mathbf{g}(\mathbf{x}) - \mathbf{x}$, both with initial condition $\mathbf{y}_0 = \mathbf{x}_0 = (0, \ldots, 0)^\top$. Because $\mathbf{f}(\mathbf{y}) > \mathbf{g}(\mathbf{y})$, the comparison lemma implies that $\mathbf{y}(t) > \mathbf{x}(t)$ for all $t \geq 0$ (see Khalil (2002), Lemma 3.4). In particular, we can conclude that the fixed point $\mathbf{f}(\mathbf{y}) = \mathbf{y}$ must be higher than the fixed point $\mathbf{g}(\mathbf{x}) = \mathbf{x}$. That is, in the stochastically stable equilibrium R&D effort levels are too low compared to the social optimum. Moreover, because a link is only present if $\rho y_i y_j > \zeta$ there are fewer links in the stochastically stable network than in the efficient network. $\square$

Supplementary Appendix for "Endogenous Technology Spillovers in R&D Collaboration Networks"

Chih-Sheng Hsieh[a], Michael D. König[b,c,d], Xiaodong Liu[e]

[a] *Department of Economics, National Taiwan University, Taipei 10617, Taiwan.*
[b] *Centre for Economic Policy Research (CEPR), London, EC1V 0DX, United Kingdom.*
[c] *KOF Swiss Economic Institute, ETH Zurich, 8092 Zurich, Switzerland.*
[d] *Tinbergen Institute and Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.*
[e] *Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309–0256, United States.*

Contents

Appendix B: Network Definitions and Characterizations

A *network* (*graph*) G 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. 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 $\mathcal{N}_i^{(2)} = \bigcup_{j \in \mathcal{N}_i} \mathcal{N}_j \backslash (\mathcal{N}_i \cup \{i\})$ 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 $\mathcal{N}_i^{(k)} = \bigcup_{j \in \mathcal{N}_i^{(k-1)}} \mathcal{N}_j \backslash \left(\bigcup_{l=0}^{k-1} \mathcal{N}_i^{(l)} \right)$. 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.

Let $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ 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 $(i, j) \notin \mathcal{E}$.

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.

Given a graph G and a set $\mathcal{S} \subseteq \mathcal{N}$, we say that $G_{\mathcal{S}}$ is the *subgraph* G *induced* S whenever the adjacency matrix of $G_{\mathcal{S}}$ is $\mathbf{A}_{\mathcal{S}}$. We write $G_{-\mathcal{S}}$ to denote the network $G_{\mathcal{N} \backslash \mathcal{S}}$, that is $G_{-\mathcal{S}}$ is the network that results after eliminating all the nodes in $\mathcal{S}$.

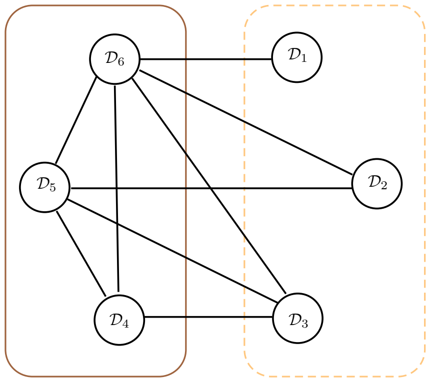
A *dominating set* for a graph $G = (\mathcal{N}, \mathcal{E})$ is a subset $\mathcal{S} \subseteq \mathcal{N}$ such that every node not in $\mathcal{S}$ is connected to at least one member of S by a link. An *independent set* is a set of nodes in a graph in which no two nodes are adjacent. For example 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). Further, 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. A *nested split graph* is a graph with a nested neighborhood structure such that the set of neighbors of each node is contained in the set of neighbors of each higher degree node (Cvetkovic and Rowlinson 1990; Mahadev and Peled 1995). Let $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ be the degree partition of a nested split graph $G = (\mathcal{N}, \mathcal{E})$. 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 $G' = (\mathcal{N} \backslash \mathcal{D}_0, \mathcal{E})$. Moreover, the neighborhoods of the nodes are nested. 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 $i = \left\lfloor \frac{k}{2} \right\rfloor + 1, \ldots, k$. See also the left panel of Figure B.1. An alternative definition can be found in Definition A.1.

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 $\mathcal{V}_1$ and $\mathcal{V}_2$ such that each link connects a node in $\mathcal{V}_1$ to a node in $\mathcal{V}_2$. $\mathcal{V}_1$ and $\mathcal{V}_2$ are independent sets with cardinalities n_1 and n_2, respectively. In a complete bipartite graph K_{n_1,n_2} each node in $\mathcal{V}_1$ is connected to each other node in $\mathcal{V}_2$. 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 $\overline{K}_n$.

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,



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

Figure B.1. (Left panel) Representation of nested split graphs and their degree partitions $\mathcal{D}$ with corresponding adjacency matrices $\mathbf{A}$. A line between $\mathcal{D}_i$ and $\mathcal{D}_j$ indicates that every node in $\mathcal{D}_i$ is adjacent to every node in $\mathcal{D}_j$. The partitions included in the solid frame ($\mathcal{D}_i$ with $\left\lfloor \frac{k}{2} \right\rfloor + 1 \leq i \leq k$) are the dominating subsets while the partitions in the dashed frame ($\mathcal{D}_i$ with $1 \leq i \leq \left\lfloor \frac{k}{2} \right\rfloor$) are the independent sets with $k = 6$. The illustration follows Mahadev and Peled (1995, p. 11). (Right panel) The corresponding stepwise adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ 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 $h \leq i$. The step function separating the zero entries in the adjacency matrix from the one entries is shown with a line.

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* $\lambda_{\max}(G)$ is the *largest real eigenvalue* of $\mathbf{A}$ associated with G, i.e., all eigenvalues λ_i of $\mathbf{A}$ satisfy $|\lambda_i| \leq \lambda_{\max}(G)$ for $i = 1, \ldots, n$ and there exists an associated nonnegative eigenvector $\mathbf{v}_{\max} \geq 0$ such that $\mathbf{A}\mathbf{v}_{\max} = \lambda_{\max}(G)\mathbf{v}_{\max}$. For a connected graph G the adjacency matrix $\mathbf{A}$ has a unique largest real eigenvalue $\lambda_{\max}(G)$ and a positive associated eigenvector $\mathbf{v}_{\max} > 0$. There 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 $\operatorname{tr}\left(\mathbf{A}^k\right) = \sum_{i=1}^n \left(\mathbf{A}^k\right)_{ii} = \sum_{i=1}^n \lambda_i^k$. We further have that $\operatorname{tr}(\mathbf{A}) = 0$, $\operatorname{tr}\left(\mathbf{A}^2\right)$ gives twice the number of links in G and $\operatorname{tr}\left(\mathbf{A}^3\right)$ 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 $h \leq i$. In a stepwise adjacency matrix, a step function separates the zero entries from the one entries in the matrix. See also the right panel in Figure B.1. 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 that maximize the largest eigenvalue, $\lambda_{\max}(G)$, for a given number of links (Brualdi and Solheid 1986), and they are the ones that maximize the degree variance (Peled et al. 1999). See e.g., König et al. (2014) for further properties.

Given n and $p \in [0, 1]$, the *random graph* $G(n, p)$ is generated by letting each pair of nodes be connected by an edge with probability p, independently. A natural generalization of $G(n, p)$ is obtained by replacing the single parameter p by a symmetric $n \times n$ matrix $(p_{ij})_{1 \leq i,j,n}$ with $0 \leq p_{ij} \leq 1$. We write $G(n, (p_{ij})_{1 \leq i,j,n})$ for the *inhomogeneous random graph* with nodes set $\mathcal{N}$ where i and j are connected by a link with probability p_{ij}, and these events are independent for all pairs (i, j) with $1 \leq i < j \leq n$ (Söderberg 2002; Bollobás et al. 2007).

Given a set of attributes $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$, a vertex v is associated with the set $\mathcal{S}(v)$ of attributes selected by v from $\mathcal{A}$. Let the vertices $v_1, \ldots, v_n$ choose their attribute sets $\mathcal{S}_i = \mathcal{S}(v_i)$, $1 \leq i \leq n$, independently at random, and make v_i and v_j adjacent whenever they have at least $s \geq 1$ attributes in common, that is, $|\mathcal{S}_i \cap \mathcal{S}_j| \geq s$. The graph on the vertex set $\mathcal{N} = \{v_1, \ldots, v_n\}$ defined by this adjacency relationship is then called the *random intersection graph* $G(n, m, s)$ (Deijfen and Kets 2009).

Appendix C: Cournot Oligopoly with Cost Reducing R&D Collaborations

We consider a Cournot oligopoly game in which a set $\mathcal{N} = \{1, \ldots, n\}$ of firms is competing in a product market with imperfectly substitutable goods.[1] Firms are not only competitors in the product market, but they can also form pairwise collaborative R&D agreements. These pairwise links involve a commitment to share R&D results and lead to lower marginal cost of production of the collaborating firms. The amount of this cost reduction depends on the effort the firms invest into R&D. Given the collaboration network $G \in \mathcal{G}^n$, where $\mathcal{G}^n$ denotes the set of all graphs with n nodes, and the firms' R&D effort levels, $\mathbf{e} = (e_i)_{i=1}^n$, the marginal cost of production for firm i is given by (Goyal and Moraga-González 2001):

$$c_i = \bar{c}_i - \alpha e_i - \beta \sum_{j=1}^{n} a_{ij} e_j. \tag{C.1}$$

where $a_{ij} \in \{0, 1\}$ in Equation (C.1) is the ij-th element of the adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ indicating whether firms i and j are collaborating (or not). The parameter $\alpha \geq 0$ measures the relative cost reduction due to a firm's own R&D effort while the parameter $\beta \geq 0$ measures the relative cost reduction due to the R&D effort of its collaboration partners.[2] We further allow for ex-ante heterogeneity among firms in the fixed marginal cost parameter $\bar{c}_i \geq 0$ expressing their different technological and organizational capabilities (Blundell et al. 1995).

We also assume that firms incur a direct cost γe_i^2, $\gamma \geq 0$, for their R&D effort and a fixed cost $\zeta \geq 0$ for each R&D collaboration.[3] The profit of firm i is then given by

$$\pi_i = (p_i - c_i)q_i - \gamma e_i^2 - \zeta d_i, \tag{C.2}$$

where $q_i \geq 0$ is the output level and $d_i = \sum_{j=1}^n a_{ij}$ counts the number of collaborations of firm i. Inserting marginal costs from Equation (C.1) gives

$$\pi_i = p_i q_i - \bar{c}_i q_i + \alpha q_i e_i + \beta q_i \sum_{j=1}^{n} a_{ij} e_j - \gamma e_i^2 - \zeta d_i.$$

The optimal R&D effort level follows from the first-order condition $\frac{\partial \pi_i}{\partial e_i} = \alpha q_i - 2\gamma e_i = 0$. Solving for e_i delivers $e_i^* = \tau q_i$, where we have denoted by $\tau \equiv \frac{\alpha}{2\gamma}$. This equation can be viewed as reflecting the learning-by-doing effects of production on R&D efforts. It is consistent with various empirical studies which have found that the R&D effort is correlated with firm size (Cohen and Klepper 1996a,b). We then can write marginal costs from Equation (C.1) as follows[4]

$$c_i = \bar{c}_i - \tau\alpha q_i - \tau\beta \sum_{j=1}^{n} a_{ij} q_j, \tag{C.3}$$

while profits can be written as

$$\pi_i = p_i q_i - \bar{c}_i q_i + \tau\alpha q_i^2 + \tau\beta q_i \sum_{j=1}^{n} a_{ij} q_j - \tau^2 \gamma q_i^2 - \zeta d_i. \tag{C.4}$$

1. Generalizations to Bertrand competition are straightforward (see e.g., König et al. 2019; Westbrock 2010). In the empirical application of the model in Section 3 we further generalize the model by allowing for competition in multiple product markets.

2. Note that in this model, firms are exposed to business stealing effects if their rivals increase their output via cost reducing R&D collaborations.

3. In Section 2.6 we discuss several extensions of the model including heterogeneous linking costs.

4. We assume that firms always implement the optimal R&D effort level. Since the optimal R&D effort decision only depends on a firm's own output, we assume that a firm does not face any uncertainty when implementing this strategy. See also Kamien et al. (1992) for a similar model of competitive research joint ventures in which firms unilaterally (and optimally) choose their R&D effort levels.

Next we consider the demand for goods produced by firm i. A representative consumer maximizes (Singh and Vives 1984)

$$U = I + \upsilon \sum_{i=1}^{n} q_i - \frac{\psi}{2} \sum_{i=1}^{n} q_i^2 - \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} q_i q_j, \tag{C.5}$$

with the budget constraint $I + \sum_{i=1}^{n} p_i q_i \leq E$, endowment E and a numeraire good I. The parameter $\upsilon > 0$ captures the total size of the market, whereas $\lambda \in (0, 1]$, measures the degree of substitutability between products. In particular, $\lambda = 1$ depicts a market of perfect substitutable goods, while $\lambda \to 0$ represents the case of almost independent markets. The parameter λ is therefore a measure of the degree of competition between firms. We assume that $\psi > \lambda > 0$ to ensure that the utility function is concave. The budget constraint is binding and the utility maximization of the representative consumer gives the inverse demand function for firm i:[5]

$$p_i = \upsilon - \psi q_i - \lambda \sum_{j \neq i}^{n} q_j. \tag{C.6}$$

Moreover, inserting Equation (C.6) into the utility (C.5) of the consumer gives:

$$\begin{aligned} U &= E - \sum_{i=1}^{n} p_i q_i + \upsilon \sum_{i=1}^{n} q_i - \frac{\psi}{2} \sum_{i=1}^{n} q_i^2 - \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} q_i q_j \\ &= E - \sum_{i=1}^{n} \left(\upsilon - \psi q_i - \lambda \sum_{j \neq i}^{n} q_j \right) q_i + \upsilon \sum_{i=1}^{n} q_i - \frac{\psi}{2} \sum_{i=1}^{n} q_i^2 - \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} q_i q_j \\ &= E + \frac{\psi}{2} \sum_{i=1}^{n} q_i^2 + \frac{\lambda}{2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} q_i q_j. \end{aligned} \tag{C.7}$$

This defines the consumer surplus and is relevant for the efficiency analysis in Section 2.5. Note that in the special case of non-substitutable goods, when $\lambda \to 0$, we obtain $U = E + \frac{\psi}{2} \sum_{i=1}^{n} q_i^2$, while in the case of perfectly substitutable goods, when $\lambda \to 1$, we get $U = E + \frac{1+\psi}{2} \left(\sum_{i=1}^{n} q_i \right)^2$.

Inserting marginal cost from Equation (C.3) and inverse demand from Equation (C.6) we can write firm i's profit as

$$\pi_i = (\upsilon - \bar{c}_i) q_i - (\psi - \tau\alpha + \tau^2\gamma) q_i^2 - \lambda q_i \sum_{j \neq i} q_j + \tau\beta \sum_{j=1}^{n} a_{ij} q_i q_j - \zeta d_i. \tag{C.8}$$

If we denote by $\eta_i \equiv \upsilon - \bar{c}_i$, $\nu \equiv \psi - \tau\alpha + \tau^2\gamma$ and $\rho \equiv \tau\beta$ then we can write Equation (C.8) more compactly as follows

$$\pi_i = \eta_i q_i - \nu q_i^2 - \lambda q_i \sum_{j \neq i} q_j + \rho \sum_{j=1}^{n} a_{ij} q_i q_j - \zeta d_i. \tag{C.9}$$

Further, if we normalize $\nu = 1/2$ then profits from Equation (C.9) can be written as in Equation (1).

Appendix D: Equilibrium Characterization for Homogeneous Firms

In the following we provide a complete equilibrium characterization in the special case of ex-ante homogeneous firms, that is, $\eta_i = \eta$ for all $i \in \mathcal{N}$ in Equation (1).

5. With the budget constraint $I = E - \sum_{i=1}^{n} p_i q_i$ in the consumer's utility in Equation (C.5), the FOC is given by $\frac{\partial U}{\partial q_i} = -p_i + \upsilon - \psi q_i - \lambda \sum_{j \neq i}^{n} y_j = 0$, from which we obtain Equation (C.6).

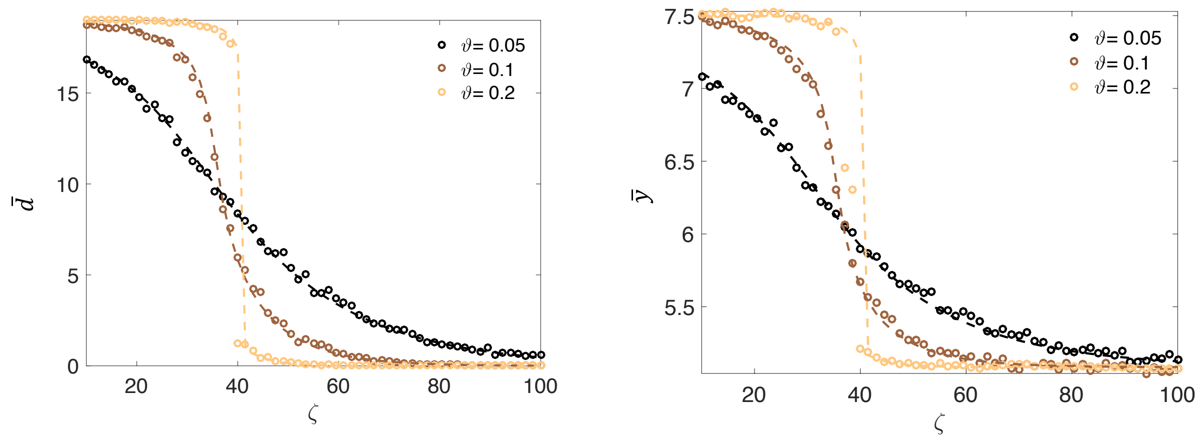


FIGURE D.1. The average degree $\bar{d}$ (left panel) and the average R&D effort $\bar{y}$ (right panel) as a function of the linking cost ζ for varying values of $\vartheta \in \{0.05, 0.1, 0.2\}$ with $n = 20$ firms and $\tau = \xi = \chi = 1$, $\eta = 300$, $\rho = 1$, $\lambda = 1$ and $\nu = 20$. Dashed lines indicate the theoretical predictions of Equations (D.1) and Equation (D.3) in Proposition D.1, respectively.

PROPOSITION D.1. *Consider homogeneous firms such that $\eta_i = \eta$ in the profit function of Equation (1) for all $i = 1, \ldots, n$, and let the evolution of the firms' R&D effort levels and collaborations be governed by the stochastic process in Definition 1. Denote by $\eta^* \equiv \eta/(n-1)$ and $\nu^* \equiv \nu/(n-1)$. Moreover, let the empirical average R&D effort level be denoted by $\bar{y} \equiv \frac{1}{n}\sum_{i=1}^n y_i$ and the average degree be $\bar{d} \equiv \frac{1}{n}\sum_{i=1}^n d_i$. Further, let the empirical degree distribution be given by $\bar{P}^\vartheta(k) \equiv \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{d_i = k\}}$, and denote by $P^\vartheta(k) \equiv \mathbb{E}_{\mu^\vartheta}\left(\bar{P}^\vartheta(k)\right)$.*

(i) Let $y^* \in \mathcal{Y}$ be the root of

$$(\lambda + 2\nu^*)y - \eta^* = \frac{\rho}{2}\left(1 + \tanh\left(\frac{\vartheta}{2}\left(\rho y^2 - \zeta\right)\right)\right)y, \tag{D.1}$$

which has at least one solution if $\lambda + 2\nu^* > \rho$. Then, $\bar{y} \xrightarrow{a.s.} y^*$. Moreover, for large n, the firms' R&D effort levels become independent Gaussian distributed random variables, $y_i \xrightarrow{d} \mathcal{N}(y^*, \sigma^2)$, with mean y^* and variance $\sigma^2 = n/\left(2\vartheta\nu^* + \vartheta^2(by^* - \eta^* + 2\nu^* y^*)(y^*(b + 2\nu^* - \rho) - \eta^*)\right)$. The degree d_i of firm i follows a mixed Poisson distribution with mixing parameter $\int_{\mathcal{Y}} p^\vartheta(y, y')\mu^\vartheta(dy')$, where $p^\vartheta(y, y') = e^{\vartheta(\rho y y' - \zeta)}/(1 + e^{\vartheta(\rho y y' - \zeta)})$, and for any $1 < m \leq n$ the degrees $d_1, \ldots, d_m$ are asymptotically independent. In particular,

$$P^\vartheta(k) = \mathbb{E}_{\mu^\vartheta}\left(\frac{e^{-\bar{d}(y_1)}\bar{d}(y_1)^k}{k!}\right)(1 + o(1)), \tag{D.2}$$

where the expected degree for large ϑ is given by

$$\mathbb{E}_{\mu^\vartheta}\left(\bar{d}\right) = \frac{n-1}{2}\left(1 + \tanh\left(\frac{\vartheta}{2}\left(\rho(y^*)^2 - \zeta\right)\right)\right) + O\left(\frac{1}{\vartheta}\right), \tag{D.3}$$

and y^* is given by the solution to Equation (D.1).

(ii) For $\vartheta \to \infty$, in the stochastically stable state, the probability measure μ^* is concentrated on

$$y^* = \begin{cases} \frac{\eta^*}{\lambda + 2\nu^* - \rho}, & \text{if } \zeta < \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*)^2}, \\ \left\{\frac{\eta^*}{\lambda+2\nu^*-\rho}, \frac{\eta^*}{\lambda+2\nu^*}\right\}, & \text{if } \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*)^2} < \zeta < \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*-\rho)^2}, \\ \frac{\eta^*}{\lambda+2\nu^*}, & \text{if } \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*-\rho)^2} < \zeta, \end{cases} \tag{D.4}$$

and we refer to the two possible R&D effort levels in Equation (D.4) as the high equilibrium and the low equilibrium, respectively. The expected average degree in the high equilibrium is $\mathbb{E}_{\mu^*}\left(\bar{d}\right) = \lim_{\vartheta\to\infty}\mathbb{E}_{\mu^\vartheta}\left(\bar{d}\right) = n - 1$, which corresponds to a complete graph, K_n, and $\mathbb{E}_{\mu^*}\left(\bar{d}\right) = \lim_{\vartheta\to\infty}\mathbb{E}_{\mu^\vartheta}\left(\bar{d}\right) = 0$ in the low equilibrium, which corresponds to an empty graph, $\overline{K}_n$.

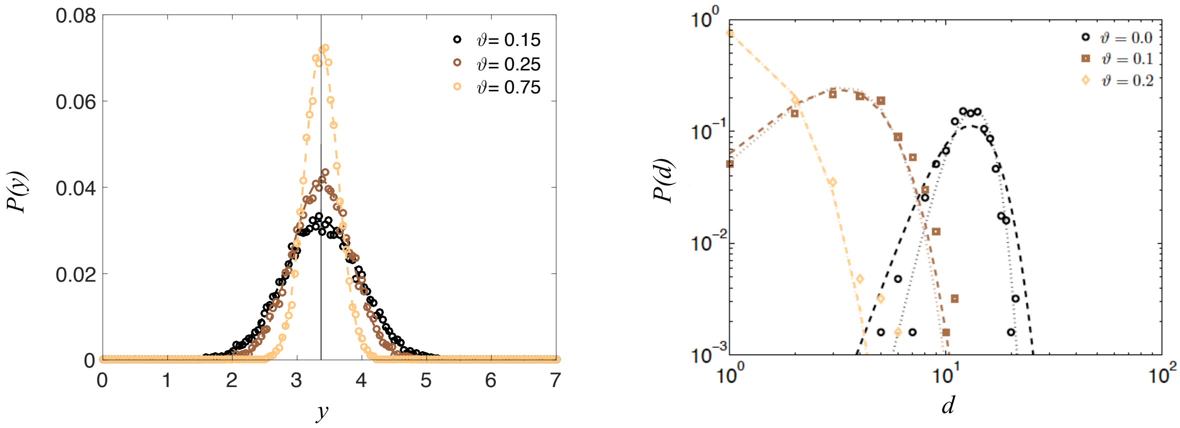


FIGURE D.2. (Left panel) The stationary R&D effort distribution, $P(y)$, for $n = 50$, $\eta = 150$, $\lambda = 0.5$, $\nu = 10$, $\rho = 1$, $\vartheta \in \{0.1, 0.25, 0.75\}$ and $\zeta = 60$. Dashed lines indicate the normal distribution $\mathcal{N}(y^*, \sigma^2)$ of part (i) of Proposition D.1. (Right panel) The stationary degree distribution, $P(d)$, for the same parameter values. The dashed lines indicate the solution in Equation (D.2) of Proposition D.1.

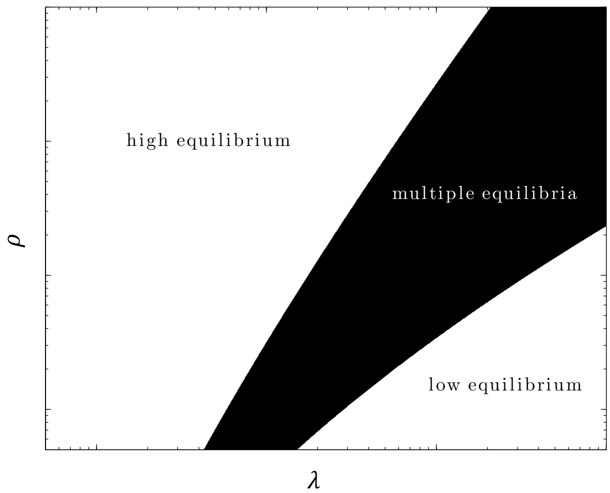


FIGURE D.3. A phase diagram illustrating the regions with a unique and with multiple equilibria according to Equation (D.1) in Proposition D.1 for varying values of the competition parameter $b \in \{0, \ldots, 0.01\}$ and the spillover parameter $\rho \in \{0, \ldots, 0.01\}$ with $n = 100$, $\nu = 0.5$, $\eta = 100$ $\vartheta = 1$ and $\zeta = 50$.

Figure D.1 shows the average R&D effort $\bar{y}$ of Equation (D.1) and the average degree $\bar{d}$ of Equation (D.3) in part (i) Proposition D.1 as a function of the linking cost ζ. With increasing cost, ζ, both network connectivity and R&D effort are decreasing. The transition from an economy with high R&D effort and collaboration intensity (high equilibrium) to an economy with low R&D effort and collaboration intensity (low equilibrium) is becoming sharper as ϑ increases, consistent with the limit of part (ii) in Proposition D.1. An illustration of the Gaussian R&D effort distribution, $\mathcal{N}(y^*, \sigma^2)$, in part (i) of Proposition D.1 together with the results of numerical simulations can be seen in the left panel in Figure D.2. The figure shows that the analytic prediction reproduces the simulation results fairly well even for small values of ϑ, and that the distribution becomes more concentrated on y^* as ϑ increases. Further note that in the limit of large n, the variance-covariance matrix of the firms' effort distribution becomes a diagonal matrix, implying that the output levels become independent with mean y^* given by the solution to Equation (D.4). This independence for large networks has similarly been used in the mean-field approximation proposed in Mele and Zhu (2022) for estimating exponential random graph models. Moreover, in the absence of network spillovers, when we set ρ to zero, total output ($Y^* = ny^*$) converges to total output in the perfect competition equilibrium ($Y^* \underset{n\to\infty}{\longrightarrow} (\upsilon - \bar{c})/\lambda$; cf. Supplementary Appendix C).

A phase diagram illustrating the regions with a unique and with multiple equilibria according to Equation (D.4) in part (ii) Proposition D.1 can be seen in Figure D.3. Multiple equilibria arise

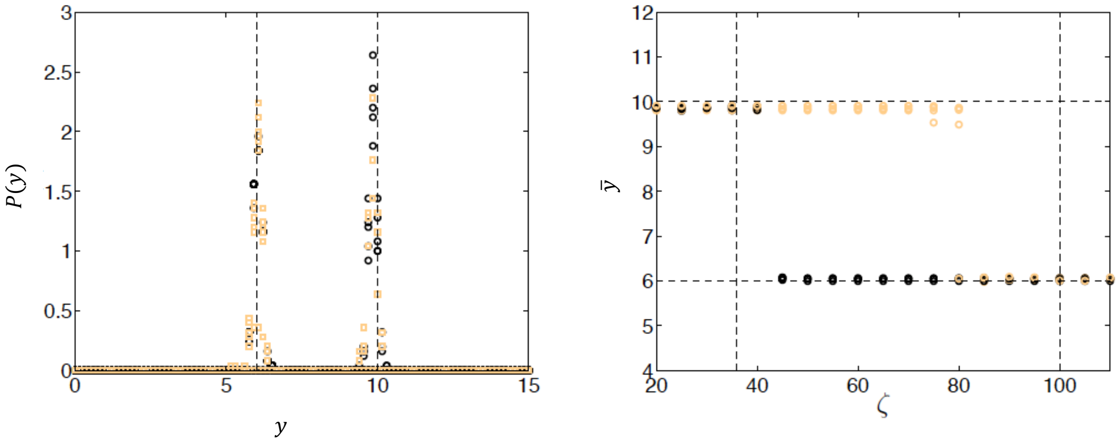


FIGURE D.4. (Left panel) The stationary R&D effort distribution. The vertical dashed lines indicate the theoretical predictions from Equation (D.4). (Right panel) The average R&D effort level from numerical simulations with $\vartheta = 1$ starting with different initial conditions (indicated with different colors). The horizontal dashed lines indicate the equilibrium quantities and the vertical dashed lines are the threshold cost levels from Equation (D.4). In the region of the cost ζ between the lower and upper thresholds two equilibria exist.

in the intermediate range of technology spillovers (ρ) and market stealing effects from competition (λ). In contrast, when spillovers are much stronger (weaker) than competition, then the unique high (low) equilibrium is selected. Note that the equilibrium R&D effort levels in Equations (D.1) and (D.4) are increasing in ρ (spillovers) and η (market size), and decreasing in ζ (linking cost) and λ (competition).[6] This implies that both higher collaboration costs (or weaker spillovers) and more intense competition (or smaller market size and/or higher production costs) decrease overall production and collaboration intensity. An illustration with the average R&D effort level from numerical simulations starting with different initial conditions and a comparison with the predictions of Equation (D.4) can be seen in Figure D.4.

Proof of Proposition D.1. We start with the proof of the first part of the proposition. Observe that the potential of Equation (2) can be written as

$$\Phi(\mathbf{y}, G) = \underbrace{\sum_{i=1}^{n}\left(\eta - \nu y_i - \frac{\lambda}{2}\sum_{j\neq i}^{n} y_j\right) y_i}_{\varphi(\mathbf{y})} + \sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}\underbrace{(\rho y_i y_j - \zeta)}_{\sigma_{ij}} = \varphi(\mathbf{y}) + \sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}\sigma_{ij}. \quad \text{(D.5)}$$

We then have that $e^{\vartheta\Phi(\mathbf{y},G)} = e^{\vartheta\varphi(\mathbf{y})}e^{\vartheta\sum_{i<j}^{n} a_{ij}\sigma_{ij}}$, where only the last factor on the RHS is network dependent. Observing that the sequence $(a_{ij})_{1\leq i<j\leq n} = (a_{12}, a_{13}, \ldots, a_{n-1,n})$ is a binary sequence as in Lemma A.1, we then can use the fact that for any constant, symmetric $\sigma_{ij} = \sigma_{ji}$, $1 \leq i, j \leq n$, we can write

$$\sum_{G\in\mathcal{G}^n} e^{\vartheta\sum_{i<j}^{n} a_{ij}\sigma_{ij}} = \prod_{i=1}^{n}\prod_{j=i+1}^{n}\left(1 + e^{\vartheta\sigma_{ij}}\right). \quad \text{(D.6)}$$

From Equation (D.6) we then obtain

$$\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)} = e^{\vartheta\varphi(\mathbf{y})}\prod_{i<j}^{n}\left(1 + e^{\vartheta\sigma_{ij}}\right) = \prod_{i=1}^{n} e^{\vartheta\left(\eta - \nu y_i - \frac{\lambda}{2}\sum_{j\neq i}^{n} y_j\right)y_i}\prod_{i<j}^{n}\left(1 + e^{\vartheta(\rho y_i y_j - \zeta)}\right). \quad \text{(D.7)}$$

6. See also Figure D.5.

We can use Equation (D.7) to compute the marginal distribution

$$\begin{aligned}
\mu^\vartheta(\mathbf{y}) &= \frac{1}{\mathscr{Z}_\vartheta}\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)} \\
&= \frac{1}{\mathscr{Z}_\vartheta}\prod_{i=1}^n e^{\vartheta\left(\eta-\nu y_i-\frac{\lambda}{2}\sum_{j\neq i}y_j\right)y_i}\prod_{i<j}^n\left(1+e^{\vartheta(\rho y_i y_j-\zeta)}\right) \\
&= \frac{1}{\mathscr{Z}_\vartheta}e^{\vartheta\sum_{i=1}^n\left(\eta-\nu y_i-\frac{\lambda}{2}\sum_{j\neq i}y_j\right)y_i}e^{\sum_{i<j}^n\ln\left(1+e^{\vartheta(\rho y_i y_j-\zeta)}\right)} \\
&= \frac{1}{\mathscr{Z}_\vartheta}e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y})},
\end{aligned} \tag{D.8}$$

where we have introduced the Hamiltonian

$$\mathscr{H}_\vartheta(\mathbf{y}) \equiv \sum_{i=1}^n\left(\eta y_i-\nu y_i^2+\sum_{j>i}^n\left(\frac{1}{\vartheta}\ln\left(1+e^{\vartheta(\rho y_i y_j-\zeta)}\right)-\lambda y_i y_j\right)\right). \tag{D.9}$$

Using the fact that $\int_{\mathcal{Y}^n}\mu^\vartheta(\mathbf{y})d\mathbf{y}=1$, it follows from Equation (D.8) that we can write the partition function as

$$\mathscr{Z}_\vartheta = \int_{\mathcal{Y}^n} e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y})}d\mathbf{y}.$$

We next make the Laplace approximation (Wong 2001, Theorem 3, p. 495)

$$\mathscr{Z}_\vartheta \sim \left(\frac{2\pi}{\vartheta}\right)^{\frac{n}{2}}\left|\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)_{y_i=y^*}\right|^{-\frac{1}{2}}e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y}^*)}, \tag{D.10}$$

for large ϑ, where $\mathbf{y}^*=\text{argmax}_{\mathbf{y}\in[0,\bar{y}]^n}\mathscr{H}_\vartheta(\mathbf{y})$, and the Hessian is given by $\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}$ for $1\leq i,j\leq n$. From Equation (D.9) we find that

$$\frac{\partial\mathscr{H}_\vartheta}{\partial y_i} = \eta-2\nu y_i+\sum_{j\neq i}^n\left(\frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}(\rho y_i y_j-\zeta)\right)\right)-\lambda\right)y_j. \tag{D.11}$$

The first order conditions $\frac{\partial\mathscr{H}_\vartheta}{\partial y_i}=0$ in Equation (D.11) imply that

$$\eta-2\nu y_i = \sum_{j\neq i}^n\left(\lambda-\frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}(\rho y_i y_j-\zeta)\right)\right)\right)y_j.$$

This system of equations has a symmetric solution, $y_i=y$ for all $i=1,\ldots,n$, where

$$(\lambda(n-1)+2\nu)y-\eta = \frac{(n-1)\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}(\rho y^2-\zeta)\right)\right)y.$$

Introducing the variables $\eta^*\equiv\eta/(n-1)$ and $\nu^*\equiv\nu/(n-1)$, this can be written as

$$(\lambda+2\nu^*)y-\eta^* = \frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}(\rho y^2-\zeta)\right)\right)y. \tag{D.12}$$

Let the RHS of Equation (D.12) be denoted by $F(y)$ so that we can write it as $(\lambda+2\nu^*)y-\eta^*=F(y)$. Then we have that $F(0)=0$, $F'(y)\geq 0$ and $F(y)\sim\rho y$ for $y\to\infty$. It follows that $(\lambda+2\nu^*)y-\eta^*=F(y)$ has at least one solution when $\lambda+2\nu^*>\rho$.[7] Moreover, any iteration $(y_t)_{t=0}^\infty$ with updates $(\lambda+2\nu^*)y_{t+1}-\eta^*=F(y_t)$ starting at $y_0=0$ converges to the smallest fixed point y^* such that $(\lambda+2\nu^*)y^*-\eta^*=F(y^*)$.

7. Since the RHS, $F(y)$, of Equation (D.12) is increasing (one can see this from taking the derivative), is zero at $y=0$, i.e., $F(0)=0$, and asymptotically grows linearly as ρy, it follows that when $\lambda+2\nu^*>\rho$ there must exist at least one fixed point. This is because the LHS, $(\lambda+2\nu^*)y-\eta^*$, of Equation (D.12) starts below zero at $y=0$ (where it is $-\eta^*$), both LHS and RHS are increasing, and the RHS approaches asymptotically a line with a slope smaller than the slope $\lambda+2\nu^*$ of the LHS. Hence they must intersect at some $y\geq 0$.

We next compute the average R&D effort level $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$. We have that

$$\begin{aligned}\mathbb{E}_{\mu^\vartheta}\left(\sum_{i=1}^{n} y_i\right) &= \sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n} d\mathbf{y}\left(\sum_{i=1}^{n} y_i\right)\mu^\vartheta(\mathbf{y},G) = \frac{1}{\mathscr{Z}_\vartheta}\sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n} d\mathbf{y}\left(\sum_{i=1}^{n} y_i\right)e^{\vartheta\Phi(\mathbf{y},G)} \\ &= \frac{1}{\mathscr{Z}_\vartheta}\sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n} d\mathbf{y}\frac{1}{\vartheta}\frac{\partial}{\partial\eta}e^{\vartheta\Phi(\mathbf{y},G)} = \frac{1}{\vartheta}\frac{1}{\mathscr{Z}_\vartheta}\frac{\partial\mathscr{Z}_\vartheta}{\partial\eta} = \frac{1}{\vartheta}\frac{\partial\ln\mathscr{Z}_\vartheta}{\partial\eta} = -\frac{1}{\vartheta}\frac{\partial\mathscr{F}_\vartheta}{\partial\eta},\end{aligned}$$

where we have denoted $\mathscr{F}_\vartheta \equiv -\ln\mathscr{Z}_\vartheta$. The average R&D effort is then given by

$$\mathbb{E}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = -\frac{1}{n\vartheta}\frac{\partial\mathscr{F}_\vartheta}{\partial\eta}.$$

With Equation (D.10) we get

$$\mathscr{F}_\vartheta \sim -\frac{n}{2}\ln\left(\frac{2\pi}{\vartheta}\right) + \frac{1}{2}\ln\left|\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)_{y_i=y^*}\right| - \vartheta\mathscr{H}_\vartheta(\mathbf{y}^*).$$

We then find that

$$\begin{aligned}\frac{\partial\mathscr{F}_\vartheta}{\partial\eta} &= -\vartheta\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\eta} + \frac{1}{2}\frac{\partial}{\partial\eta}\ln\left|\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)_{y_i=y^*}\right| \\ &= -\vartheta\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\eta} + \frac{1}{2}\mathrm{tr}\left(\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)^{-1}\frac{\partial}{\partial\eta}\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)\right)_{y_i=y^*},\end{aligned}$$

where we have used Jacobi's formula (Horn and Johnson 1990).[8] From Equation (D.11) we further have that

$$\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i^2} = -2\nu + \frac{\vartheta\rho^2}{4}\sum_{j\neq i}^{n} y_j^2\left(1 - \tanh\left(\frac{\vartheta}{2}(\rho y_i y_j - \zeta)\right)^2\right), \tag{D.13}$$

and for $j \neq i$ we have that

$$\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j} = -\lambda + \frac{\rho}{2}\left(1 + \tanh\left(\frac{\vartheta}{2}(\rho y_i y_j - \zeta)\right)\right)\left(1 + \frac{\vartheta\rho}{2}y_i y_j\left(1 - \tanh\left(\frac{\vartheta}{2}(\rho y_i y_j - \zeta)\right)\right)\right). \tag{D.14}$$

This shows that $\frac{\partial}{\partial\eta}\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right) = 0$, so that $\frac{\partial\mathscr{F}_\vartheta}{\partial\eta} = -\vartheta\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\eta}$, and the expected average R&D effort level is then given by

$$\mathbb{E}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\eta}.$$

Using the fact that $\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\eta} = \sum_{i=1}^{n} y_i = ny^*$, we then get in leading order terms for large ϑ that

$$\mathbb{E}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \lim_{\vartheta\to\infty}\mathbb{E}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = y^*.$$

Next, we compute the R&D effort distribution. It can be written as follows

$$\mu^\vartheta(\mathbf{y}) = \frac{1}{\mathscr{Z}_\vartheta}\sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)} = \frac{1}{\mathscr{Z}_n^\vartheta}e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y})},$$

where the Hamiltonian is implicitly defined by $e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y})} = \sum_{G\in\mathcal{G}^n} e^{\vartheta\Phi(\mathbf{y},G)}$. From a Taylor expansion around $\mathbf{y}^*$ we have that

$$\mathscr{H}_\vartheta(\mathbf{y}) = \mathscr{H}_\vartheta(\mathbf{y}^*) + (\mathbf{y}-\mathbf{y}^*)D^1\mathscr{H}_\vartheta(\mathbf{y}^*) + \frac{1}{2}(\mathbf{y}-\mathbf{y}^*)^\top D^2\mathscr{H}_\vartheta(\mathbf{y}^*)(\mathbf{y}-\mathbf{y}^*) + o\left(\|\mathbf{y}-\mathbf{y}^*\|^2\right),$$

8. For any invertible matrix $\mathbf{M}(y)$ for all y, Jacobi's formula states that $\frac{d}{dy}|\mathbf{M}(y)| = |\mathbf{M}(y)|\,\mathrm{tr}\left(\mathbf{M}(y)^{-1}\frac{d}{dy}\mathbf{M}(y)\right)$, which can be written more compactly as $\frac{d}{dy}\ln|\mathbf{M}(y)| = \mathrm{tr}\left(\mathbf{M}(y)^{-1}\frac{d}{dy}\mathbf{M}(y)\right)$.

as $\vartheta \to \infty$, where $\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y}\in[0,\bar{y}]^n} \mathscr{H}_\vartheta(\mathbf{y})$, the gradient is $D^1\mathscr{H}_\vartheta(\mathbf{y}) = \left(\frac{\partial \mathscr{H}_\vartheta}{\partial y_i}\right)_{i=1,\dots,n}$, and the Hessian is $D^2\mathscr{H}_\vartheta(\mathbf{y}) = \left(\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right)_{i,j=1,\dots,n}$. As the gradient $D^1\mathscr{H}_\vartheta(\mathbf{y})$ vanishes at $\mathbf{y}^*$, we have that

$$\mathscr{H}_\vartheta(\mathbf{y}) = \mathscr{H}_\vartheta(\mathbf{y}^*) + \frac{1}{2}(\mathbf{y}-\mathbf{y}^*)^\top D^2\mathscr{H}_\vartheta(\mathbf{y}^*)(\mathbf{y}-\mathbf{y}^*) + o\left(\|\mathbf{y}-\mathbf{y}^*\|^2\right).$$

We then can write

$$\mu^\vartheta(\mathbf{y}) = \frac{1}{\mathscr{Z}_n^\vartheta} e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y}^*)} \exp\left\{-\frac{1}{2}\vartheta(\mathbf{y}-\mathbf{y}^*)^\top(-D^2\mathscr{H}_\vartheta(\mathbf{y}^*))(\mathbf{y}-\mathbf{y}^*)\right\} + o\left(\|\mathbf{y}-\mathbf{y}^*\|^2\right).$$

Normalization implies that

$$\begin{aligned}\mathscr{Z}_n^\vartheta = \int_{\mathcal{Y}^n} d\mathbf{y} e^{\mathscr{H}_\vartheta(\mathbf{y})} &= e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y}^*)} \int_{\mathcal{Y}^n} e^{-\frac{1}{2}\vartheta(\mathbf{y}-\mathbf{y}^*)^\top(-D^2\mathscr{H}_\vartheta(\mathbf{y}^*))(\mathbf{y}-\mathbf{y}^*)} d\mathbf{y} + o\left(\|\mathbf{y}-\mathbf{y}^*\|^2\right) \\ &= e^{\vartheta\mathscr{H}_\vartheta(\mathbf{y}^*)}(2\pi)^{\frac{n}{2}} \left|-D^2\mathscr{H}_\vartheta(\mathbf{y}^*)\right|^{-\frac{1}{2}} + o\left(\|\mathbf{y}-\mathbf{y}^*\|^2\right).\end{aligned}$$

The Laplace approximation of $\mu^\vartheta(\mathbf{y})$ is then given by

$$\mu^\vartheta(\mathbf{y}) = \left(\frac{2\pi}{\vartheta}\right)^{-\frac{n}{2}} \left|-D^2\mathscr{H}_\vartheta(\mathbf{y}^*)\right|^{\frac{1}{2}} e^{-\frac{1}{2}\vartheta(\mathbf{y}-\mathbf{y}^*)^\top(-D^2\mathscr{H}_\vartheta(\mathbf{y}^*))(\mathbf{y}-\mathbf{y}^*)} + o\left(\|\mathbf{y}-\mathbf{y}^*\|^2\right). \tag{D.15}$$

That is, in the limit of large ϑ, $\mathbf{y}$ is asymptotically normally distributed with mean $\mathbf{y}^*$ and variance $-\frac{1}{\vartheta}D^2\mathscr{H}_\vartheta(\mathbf{y}^*)^{-1}$.

Imposing symmetry, $y_i = y$ for all $i = 1,\dots,n$, in Equation (D.13) we can write

$$\left.\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i^2}\right|_{y_i=y} = -2\nu + (n-1)\frac{\vartheta\rho^2}{4}y^2\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right)\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right),$$

and for $j \neq i$ we have from Equation (D.13) that

$$\left.\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right|_{y_i=y_j=y} = -\lambda + \frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right)\left(1+\frac{\vartheta\rho}{2}y^2\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right)\right).$$

Using Equation (D.12), from which we get

$$\frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right) = \frac{((n-1)\lambda+2\nu)y-\eta}{(n-1)y},$$

and

$$\frac{\rho}{2}\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y_i y_j-\zeta\right)\right)\right) = \frac{((n-1)(\rho-\lambda)-2\nu)y+\eta}{(n-1)y},$$

we then can write

$$\left.\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i^2}\right|_{y_i=y} = -2\nu + \frac{\vartheta(((n-1)(\rho-\lambda)-2\nu)y+\eta)(((n-1)\lambda+2\nu)y-\eta)}{n-1},$$

and

$$\left.\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right|_{y_i=y_j=y} = -b + \frac{((n-1)\lambda+2\nu)y-\eta}{(n-1)y}\left(1+\frac{\vartheta x(((n-1)(\rho-\lambda)-2\nu)y+\eta)}{n-1}\right).$$

Denoting by $\nu^* \equiv \nu/(n-1)$ and $\eta^* \equiv \eta/(n-1)$ we can further write

$$\left.\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i^2}\right|_{y_i=y} = (n-1)\left(-2\nu^* + \vartheta y^2\left(\rho-\lambda-2\nu^*+\frac{\eta^*}{y}\right)\left(\lambda+2\nu^*-\frac{\eta^*}{y}\right)\right), \tag{D.16}$$

and

$$\left.\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right|_{y_i=y_j=y} = -b + \left(\lambda+2\nu^*-\frac{\eta^*}{y}\right)\left(1+\vartheta y^2\left(\rho-\lambda-2\nu^*+\frac{\eta^*}{y}\right)\right). \tag{D.17}$$

Note that due to symmetry, the Hessian $D^2\mathscr{H}_\vartheta(\mathbf{y}^*)$ with components in Equations (D.16) and (D.17) is a special case of a circulant matrix. Denoting by a the diagonal elements of $D^2\mathscr{H}_\vartheta(\mathbf{y}^*)$ and by λ the off-diagonal elements, the determinant in Equation (D.15) follows from the general formula (Horn and Johnson 1990):

$$\left|-D^2\mathscr{H}_\vartheta(\mathbf{y}^*)\right| = \begin{vmatrix} a & b & b & \dots \\ b & a & b & \dots \\ b & b & a & \\ \vdots & \vdots & & \ddots \end{vmatrix} = (a-b)^{n-1}(a+(n-1)b).$$

Similarly, for a circulant matrix (by applying the Sherman-Morrison formula; see e.g., Horn and Johnson (1990)) we get for the inverse in Equation (D.15) that

$$\begin{aligned} -D^2\mathscr{H}_\vartheta(\mathbf{y}^*)^{-1} &= \begin{pmatrix} a & b & b & \dots \\ b & a & b & \dots \\ b & b & a & \\ \vdots & \vdots & & \ddots \end{pmatrix}^{-1} \\ &= \frac{1}{a^2+(n-2)ab-(n-1)b^2} \begin{pmatrix} a+(n-2)b & -b & -b & \dots \\ -b & a+(n-2)b & -b & \dots \\ -b & -b & a+(n-2)b & \\ \vdots & \vdots & & \ddots \end{pmatrix} \end{aligned}$$

For large n we see from Equations (D.16) and (D.17) that the off-diagonal elements vanish relatively to the diagonal elements. As $\mathbf{y}$ is asymptotically normally distributed with mean $\mathbf{y}^*$ and variance $-\frac{1}{\vartheta}D^2\mathscr{H}_\vartheta(\mathbf{y}^*)^{-1}$, this implies that, in the limit of $n \to \infty$, the individual firms' R&D effort levels become independent. The diagonal entries are given by

$$-\frac{1}{\vartheta}\left(D^2\mathscr{H}_\vartheta(\mathbf{y}^*)\right)^{-1}_{ii} \sim \frac{1}{\vartheta}\frac{n}{2\nu^*+\vartheta(\lambda y-\eta^*+2\nu^* y)(y(\lambda+2\nu^*-\rho)-\eta^*)} \equiv \sigma^2.$$

Next, we compute the expected average degree $\bar{d}$. The expected number of links can be obtained as follows

$$\mathbb{E}_{\mu^\vartheta}(m) = \sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n} m\mu^\vartheta(\mathbf{y},G)d\mathbf{y} = \frac{1}{\mathscr{Z}_\vartheta}\sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n}\underbrace{me^{\vartheta\Phi(\mathbf{y},G)}}_{-\frac{1}{\vartheta}\frac{\partial}{\partial\zeta}e^{\vartheta\Phi(\mathbf{y},G)}}d\mathbf{y} = -\frac{1}{\vartheta}\frac{1}{\mathscr{Z}_\vartheta}\frac{\partial\mathscr{Z}_\vartheta}{\partial\zeta} = \frac{1}{\vartheta}\frac{\partial\mathscr{F}_\vartheta}{\partial\zeta},$$

where we have denoted $\mathscr{F}_\vartheta \equiv -\ln\mathscr{Z}_\vartheta$. From the Laplace approximation in Equation (D.10) we find that

$$\begin{aligned} \frac{\partial\mathscr{F}_\vartheta}{\partial\zeta} &= -\vartheta\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\zeta} + \frac{1}{2}\frac{\partial}{\partial\zeta}\ln\left|\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)_{y_i=y^*}\right| \\ &= -\vartheta\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\zeta} + \frac{1}{2}\mathrm{tr}\left(\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)^{-1}\frac{\partial}{\partial\zeta}\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)\right)_{y_i=y^*}, \end{aligned}$$

where we have used Jacobi's formula (see e.g., Horn and Johnson 1990). Consequently, the expected number of links is

$$\mathbb{E}_{\mu^\vartheta}(m) = -\frac{\partial\mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial\zeta} + \frac{1}{2\vartheta}\mathrm{tr}\left(\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)^{-1}\frac{\partial}{\partial\zeta}\left(\frac{\partial^2\mathscr{H}_\vartheta}{\partial y_i\partial y_j}\right)\right)_{y_i=y^*}.$$

Further, we have that

$$\frac{\partial\mathscr{H}_\vartheta}{\partial\zeta} = -\frac{1}{2}\sum_{i=1}^n\sum_{j>i}^n\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y_i y_j-\zeta\right)\right)\right),$$

and in the symmetric equilibrium this is

$$\left.\frac{\partial\mathscr{H}_\vartheta}{\partial\zeta}\right|_{y_i=y} = -\frac{n(n-1)}{4}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right).$$

The expected number of links can then be written as

$$\mathbb{E}_{\mu^\vartheta}(m) = \frac{n(n-1)}{2}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right) + \frac{1}{2\vartheta}\operatorname{tr}\left(\left(\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right)^{-1}\frac{\partial}{\partial \zeta}\left(\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right)\right)_{y_i=y^*}.$$

Using the fact that

$$\frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)^2\right) = \lambda + 2\nu^* - \frac{\eta^*}{y},$$

where $\nu^* = \frac{\nu}{n-1}$ and $\eta^* = \frac{\eta}{n-1}$, we can write

$$\left.\frac{\partial \mathscr{H}_\vartheta}{\partial \zeta}\right|_{y_i=y} = -\frac{n(n-1)}{2\rho}\left(\lambda + 2\nu^* - \frac{\eta^*}{y}\right).$$

In the limit of $\vartheta \to \infty$ in the low equilibrium, where $y = \frac{\eta^*}{\lambda+2\nu^*}$ and therefore $\frac{\eta^*}{y} = \lambda + 2\nu^*$, we then get

$$\left.\frac{\partial \mathscr{H}_\vartheta}{\partial \zeta}\right|_{y_i=y} = 0.$$

In contrast, in the limit of $\vartheta \to \infty$ in the high equilibrium, where $y = \frac{\eta^*}{\lambda+2\nu^*-\rho}$, and $\frac{\eta^*}{y} = \lambda + 2\nu^* - \rho$ we find that

$$\left.\frac{\partial \mathscr{H}_\vartheta}{\partial \zeta}\right|_{y_i=y} = -\frac{n(n-1)}{2}.$$

Further, the derivatives with respect to the linking cost ζ in Equation (D.13) are given by

$$\frac{\partial}{\partial \zeta}\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i^2} = \frac{\vartheta^2\rho^2}{4}\sum_{j\neq i}^{n}\tanh\left(\frac{\vartheta}{2}\left(\rho y_i y_j-\zeta\right)\right)\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y_i y_j-\zeta\right)\right)^2\right),$$

and for $j \neq i$ from Equation (D.14) we get that

$$\frac{\partial}{\partial \zeta}\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i \partial y_j} = -\frac{\vartheta\rho}{4}\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y_i y_j-\zeta\right)\right)^2\right)\left(1-\vartheta\rho y_i y_j \tanh\left(\frac{\vartheta}{2}\left(\rho y_i y_j-\zeta\right)\right)\right).$$

Imposing symmetry, $y_i = y$ for all $i = 1, \ldots, n$, we then can write

$$\left.\frac{\partial}{\partial \zeta}\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i^2}\right|_{y_i=y} = \frac{(n-1)\vartheta^2\rho^2}{4}\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)^2\right), \tag{D.18}$$

and

$$\left.\frac{\partial}{\partial \zeta}\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i \partial y_j}\right|_{y_i=y_j=y} = -\frac{\vartheta\rho}{4}\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)^2\right)\left(1-\vartheta\rho y^2\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right). \tag{D.19}$$

For a circulant matrix (by applying the Sherman-Morrison formula) we have that

$$\begin{pmatrix} a & b & b & \ldots \\ b & a & b & \ldots \\ b & b & a & \\ \vdots & \vdots & & \ddots \end{pmatrix}^{-1} = \frac{1}{a^2+(n-2)ab-(n-1)b^2}\begin{pmatrix} a+(n-2)b & -b & -b & \ldots \\ -b & a+(n-2)b & -b & \ldots \\ -b & -b & a+(n-2)b & \\ \vdots & \vdots & & \ddots \end{pmatrix}$$

and

$$\operatorname{tr}\begin{pmatrix} c & d & d & \ldots \\ d & c & d & \ldots \\ d & d & c & \\ \vdots & \vdots & & \ddots \end{pmatrix}\begin{pmatrix} e & f & f & \ldots \\ f & e & f & \ldots \\ f & f & e & \\ \vdots & \vdots & & \ddots \end{pmatrix} = n(ce+(n-1)df),$$

so that

$$\operatorname{tr}\begin{pmatrix} a & b & b & \cdots \\ b & a & b & \cdots \\ b & b & a & \\ \vdots & \vdots & & \ddots \end{pmatrix}^{-1}\begin{pmatrix} e & f & f & \cdots \\ f & e & f & \cdots \\ f & f & e & \\ \vdots & \vdots & & \ddots \end{pmatrix} = \frac{n((a+(n-2)b)e-(n-1)bf)}{a^2+(n-2)ab-(n-1)b^2}.$$

The expected number of links can then be written as follows

$$\mathbb{E}_{\mu^\vartheta}(m) = \frac{n(n-1)}{2}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right)+\frac{1}{2\vartheta}\mathscr{R}_\vartheta,$$

where

$$\mathscr{R}_\vartheta \equiv \frac{n((c_1+(n-2)c_2)c_3-(n-1)c_2c_4)}{c_1^2+(n-2)c_1c_2-(n-1)c_2^2},$$

with

$$\begin{aligned}
c_1 &\equiv -2\nu+(n-1)\frac{\vartheta\rho^2y^2}{4}\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)^2\right), \\
c_2 &\equiv -\lambda+\frac{\rho}{2}\left(1+\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right)\left(1+\frac{\vartheta\rho y^2}{2}\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right)\right), \\
c_3 &\equiv \frac{(n-1)\vartheta^2\rho^2}{4}\tanh\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)^2\right), \\
c_4 &\equiv -\frac{\rho\vartheta}{4}\left(1-\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)^2\right)\left(1-\vartheta\rho y^2\tanh\left(\frac{\vartheta}{2}\left(\rho y^2-\zeta\right)\right)\right).
\end{aligned}$$

In the following we compute the degree distribution. From our previous discussion we know that each firm i has an R&D effort level y distributed identically and independently with density $\mu^\vartheta(y)$ given by $\mathcal{N}(y^*,\sigma^2)$ and converging to $\delta(y-y^*)$ in the limit $\vartheta\to\infty$. With the marginal distribution from Equation (D.8) and the potential in Equation (D.5) we then can write the conditional distribution as

$$\begin{aligned}
\mu^\vartheta(G|\mathbf{y}) = \frac{\mu^\vartheta(\mathbf{y},G)}{\mu^\vartheta(\mathbf{y})} = \frac{e^{\vartheta\Phi(\mathbf{y},G)}}{\sum_{G'\in\mathcal{G}^n}e^{\vartheta\Phi(\mathbf{y},G')}} &= \frac{e^{\varphi(\mathbf{y})}e^{\vartheta\sum_{i<j}^n a_{ij}(\rho y_iy_j-\zeta)}}{e^{\varphi(\mathbf{y})}\prod_{i<j}\left(1+e^{\vartheta(\rho y_iy_j-\zeta)}\right)} \\
&= \frac{e^{\vartheta\sum_{i<j}^n a_{ij}(\rho y_iy_j-\zeta)}}{\prod_{i<j}\left(1+e^{\vartheta(\rho y_iy_j-\zeta)}\right)} \\
&= \prod_{i<j}\frac{e^{\vartheta a_{ij}(\rho y_iy_j-\zeta)}}{1+e^{\vartheta(\rho y_iy_j-\zeta)}} \\
&= \prod_{i<j}\left(\frac{e^{\vartheta(\rho y_iy_j-\zeta)}}{1+e^{\vartheta(\rho y_iy_j-\zeta)}}\right)^{a_{ij}}\left(1-\frac{e^{\vartheta(\rho y_iy_j-\zeta)}}{1+e^{\vartheta(\rho y_iy_j-\zeta)}}\right)^{1-a_{ij}} \\
&= \prod_{i<j}p^\vartheta(y_i,y_j)^{a_{ij}}\left(1-p^\vartheta(y_i,y_j)\right)^{1-a_{ij}}. \qquad \text{(D.20)}
\end{aligned}$$

Hence, we obtain the likelihood of an inhomogeneous random graph with link probability[9]

$$p^\vartheta(y_i,y_j) = \frac{e^{\vartheta(\rho y_iy_j-\zeta)}}{1+e^{\vartheta(\rho y_iy_j-\zeta)}} = \frac{g^\vartheta(y_i,y_j)}{1+g^\vartheta(y_i,y_j)}, \qquad \text{(D.21)}$$

where we have denoted $g^\vartheta(y,y')\equiv e^{\vartheta(\rho yy'-\zeta)}$. The probability of observing the network G, given the R&D effort levels $\mathbf{y}$ can then be written as follows

$$\mu^\vartheta(G|\mathbf{y}) = \prod_{i=1}^n\prod_{j=i+1}^n\left(\frac{g^\vartheta(y_i,y_j)}{1+g^\vartheta(y_i,y_j)}\right)^{a_{ij}}\left(\frac{1}{1+g^\vartheta(y_i,y_j)}\right)^{1-a_{ij}}$$

9. See also Supplementary Appendix B for further details about inhomogeneous random graphs.

which can be written as

$$\mu^{\vartheta}(G|\mathbf{y}) = \mathscr{C}_{\vartheta}(\mathbf{y}) \prod_{i=1}^{n} \prod_{j=i+1}^{n} g^{\vartheta}(y_i, y_j)^{a_{ij}},$$

with the normalizing constant

$$\mathscr{C}_{\vartheta}(\mathbf{y}) \equiv \prod_{i=1}^{n} \prod_{j=i+1}^{n} \left(1 + g^{\vartheta}(y_i, y_j)\right).$$

Since $\sum_{G \in \mathcal{G}^n} \mathbb{P}\left(G|\, \mathbf{y}\right) = 1$, the constant $\mathscr{C}_{\vartheta}(\mathbf{y})$ can also be written as

$$\mathscr{C}_{\vartheta}(\mathbf{y}) = \sum_{G \in \mathcal{G}^n} \prod_{i=1}^{n} \prod_{j=i+1}^{n} g^{\vartheta}(y_i, y_j)^{a_{ij}}.$$

Next, we consider the probability generating function of the vector of degrees, $(d_i(G))_{i=1}^n$, given by

$$\begin{aligned}
\mathbb{E}_{\mu^{\vartheta}}\left(\prod_{i=1}^{n} y_i^{d_i(G)} \middle| \mathbf{y}\right) &= \mathbb{E}\left(\prod_{i=1}^{n} \prod_{j=i+1}^{n} (y_i y_j)^{a_{ij}} \middle| \mathbf{y}\right) \\
&= \sum_{G \in \mathcal{G}^n} \mathbb{P}\left(G|\, \mathbf{y}\right) \prod_{i=1}^{n} \prod_{j=i+1}^{n} (y_i y_j)^{a_{ij}} \\
&= \frac{1}{\mathscr{C}_{\vartheta}(\mathbf{y})} \sum_{G \in \mathcal{G}^n} \prod_{i=1}^{n} \prod_{j=i+1}^{n} g^{\vartheta}(y_i, y_j)^{a_{ij}} \prod_{i=1}^{n} \prod_{j=i+1}^{n} (y_i y_j)^{a_{ij}} \\
&= \frac{1}{\mathscr{C}_{\vartheta}(\mathbf{y})} \sum_{G \in \mathcal{G}^n} \prod_{i=1}^{n} \prod_{j=i+1}^{n} \left(g^{\vartheta}(y_i, y_j) y_i y_j\right)^{a_{ij}} \\
&= \frac{\sum_{G \in \mathcal{G}^n} \prod_{i=1}^{n} \prod_{j=i+1}^{n} \left(g^{\vartheta}(y_i, y_j) y_i y_j\right)^{a_{ij}}}{\prod_{i=1}^{n} \prod_{j=i+1}^{n} (1 + g^{\vartheta}(y_i, y_j))} \\
&= \prod_{i=1}^{n} \prod_{j=i+1}^{n} \frac{1 + g^{\vartheta}(y_i, y_j) y_i y_j}{1 + g^{\vartheta}(y_i, y_j)},
\end{aligned} \tag{D.22}$$

where we have used the fact that

$$\sum_{G \in \mathcal{G}^n} \prod_{i=1}^{n} \prod_{j=i+1}^{n} \left(g^{\vartheta}(y_i, y_j) y_i y_j\right)^{a_{ij}} = \prod_{i=1}^{n} \prod_{j=i+1}^{n} \left(1 + g^{\vartheta}(y_i, y_j) y_i y_j\right).$$

To compute the generating function of $d_1(G)$, we simply set $y_i = 1$ for all $i > 1$. Then

$$\begin{aligned}
\mathbb{E}_{\mu^{\vartheta}}\left(y_1^{d_1(G)}\right) &= \mathbb{E}_{\mu^{\vartheta}}\left(\mathbb{E}_{\mu^{\vartheta}}\left(y_1^{d_1(G)} \middle| y_1\right)\right) \\
&= \mathbb{E}_{\mu^{\vartheta}}\left(\mathbb{E}_{\mu^{\vartheta}}\left(\prod_{j=2}^{n} \frac{1 + g^{\vartheta}(y_1, y_j) y_1}{1 + g^{\vartheta}(y_1, y_j)} \middle| y_1\right)\right) \\
&= \mathbb{E}_{\mu^{\vartheta}}\left(\left(\mathbb{E}_{\mu^{\vartheta}}\left(\frac{1 + g^{\vartheta}(y_1, y_2) y_1}{1 + g^{\vartheta}(y_1, y_2)} \middle| y_1\right)\right)^{n-1}\right),
\end{aligned}$$

where we have used symmetry and the independence of $y_1, \ldots, y_n$. Further, note that

$$\frac{1 + xy}{1 + x} = 1 + (y - 1)x + O\left(x^2\right).$$

Hence, for $g^{\vartheta}(y_1, y_2)$ small in the sparse graph limit, we can write

$$\begin{aligned}
\mathbb{E}_{\mu^{\vartheta}}\left(\frac{1 + g^{\vartheta}(y_1, y_2) y_1}{1 + g^{\vartheta}(y_1, y_2)} \middle| y_1\right) &= \int_{\mathcal{Y}} \frac{1 + g^{\vartheta}(y_1, y_2) y_1}{1 + g^{\vartheta}(y_1, y_j)} \mu^{\vartheta}(dy_2) \\
&= 1 + (y_1 - 1) \int_{\mathcal{Y}} g^{\vartheta}(y_1, y_2) \mu^{\vartheta}(dy_2) + o(1) \\
&= 1 + (y_1 - 1)\nu^{\vartheta}(y_1) + o(1),
\end{aligned}$$

where we have denoted $\nu^{\vartheta}(y) \equiv \int_{\mathcal{Y}} g^{\vartheta}(y,y')\mu^{\vartheta}(dy')$. It then follows that

$$\mathbb{E}_{\mu^{\vartheta}}\left(y_1^{d_1(G)}\right) = \mathbb{E}_{\mu^{\vartheta}}\left(\left(1+(y_1-1)\nu^{\vartheta}(y_1)\right)^{n-1}\right)(1+o(1)) = \mathbb{E}_{\mu^{\vartheta}}\left(e^{(y_1-1)(n-1)\nu^{\vartheta}(y_1)}\right)(1+o(1)),$$

where we have used the fact that $e^{(y_1-1)\nu^{\vartheta}(y)} = 1+(y_1-1)\nu^{\vartheta}(y)+o(1)$. This is the probability generating function of a mixed Poisson random variable with mixing parameter $\nu^{\vartheta}(y)$. In particular, since $p^{\vartheta}(y,y') = g^{\vartheta}(y,y')+o(1)$, we can write $n\nu^{\vartheta}(y) = n\int_{\mathcal{Y}} p(y,y')\mu^{\vartheta}(dy') = \sum_{j=1}^{n}\int_{\mathcal{Y}} p^{\vartheta}(y,y_j)\mu^{\vartheta}(dy_j) = \sum_{j=1}^{n}\mathbb{P}\left(a_{1j}=1\middle| y_1=y\right) = \mathbb{E}_{\mu^{\vartheta}}\left(d_1(G)\middle| y_1=y\right)$, which is the expected degree of a firm with R&D effort y, and we denote it by $\bar{d}(y)$. Further, it then follows that

$$\begin{aligned}
\mathbb{E}_{\mu^{\vartheta}}\left(y_1^{d_1(G)}\right) &= \sum_{k=0}^{n} y_1^k \mathbb{P}\left(d_1(G)=k\right) \\
&= \mathbb{E}_{\mu^{\vartheta}}\left(e^{(y_1-1)\bar{d}(y_1)}\right)(1+o(1)) \\
&= \mathbb{E}_{\mu^{\vartheta}}\left(e^{-\bar{d}(y_1)}\sum_{k=0}^{n}\frac{(y_1\bar{d}(y_1))^k}{k!}\right)(1+o(1)) \\
&= \sum_{k=0}^{n} y_1^k \mathbb{E}_{\mu^{\vartheta}}\left(\frac{e^{-\bar{d}(y_1)}\bar{d}(y_1)^k}{k!}\right)(1+o(1)).
\end{aligned}$$

Let the empirical degree distribution be given by $\bar{P}^{\vartheta}(k) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{d_i(G)=k\}}$, and denoted by $P^{\vartheta}(k) \equiv \mathbb{E}_{\mu^{\vartheta}}\left(\bar{P}^{\vartheta}(k)\right)$. Then we have that

$$P^{\vartheta}(k) = \mathbb{P}\left(d_1(G)=k\right) = \mathbb{E}_{\mu^{\vartheta}}\left(\frac{e^{-\bar{d}(y_1)}\bar{d}(y_1)^k}{k!}\right)(1+o(1)).$$

We now give a proof of part (ii) of the proposition. In the limit of $\vartheta \to \infty$ we obtain from the FOC in Equation (D.12) that

$$(\lambda+2\nu^*)y-\eta^* = \begin{cases} \rho y, & \text{if } \zeta < \rho y^2, \\ 0, & \text{if } \rho y^2 < \zeta. \end{cases}$$

This shows that the right hand side of Equation (D.1) has a point of discontinuity at $\sqrt{\frac{\zeta}{\rho}}$ (see Figure D.5). It then follows that, in the limit of $\vartheta \to \infty$ (for the stochastically stable equilibrium), we have

$$y^* = \begin{cases} \frac{\eta}{\lambda+2\nu^*-\rho}, & \text{if } \zeta < \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*)^2}, \\ \left\{\frac{\eta^*}{\lambda+2\nu^*-\rho}, \frac{\eta^*}{\lambda+2\nu^*}\right\}, & \text{if } \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*)^2} < \zeta < \frac{\rho\eta^2}{(\lambda+2\nu^*-\rho)^2}, \\ \frac{\eta^*}{\lambda+2\nu^*}, & \text{if } \frac{\rho(\eta^*)^2}{(\lambda+2\nu^*-\rho)^2} < \zeta, \end{cases} \tag{D.23}$$

which is increasing in ρ and η^*, and decreasing in ζ and λ (see Figure D.5). Next, note that

$$\begin{aligned}
\mathbb{E}_{\mu^{\vartheta}}\left(\sum_{i=1}^{n} y_i^2\right) &= \sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n}\sum_{i=1}^{n} y_i^2\mu^{\vartheta}(\mathbf{y},G)d\mathbf{y} = \frac{1}{\mathscr{Z}_{\vartheta}}\sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n}\sum_{i=1}^{n} y_i^2 e^{\vartheta\Phi(\mathbf{y},G)}d\mathbf{y} \\
&= \frac{1}{\mathscr{Z}_{\vartheta}}\sum_{G\in\mathcal{G}^n}\int_{\mathcal{Y}^n}\frac{1}{\vartheta^2}\frac{\partial^2}{\partial\eta^2}e^{\vartheta\Phi(\mathbf{y},G)}d\mathbf{y} = \frac{1}{\mathscr{Z}_{\vartheta}}\frac{1}{\vartheta^2}\frac{\partial^2\mathscr{Z}_{\vartheta}}{\partial\eta^2},
\end{aligned}$$

where we have denoted $\mathscr{F}_{\vartheta} \equiv -\ln\mathscr{Z}_{\vartheta}$. We further have that

$$\begin{aligned}
\frac{\partial^2\ln\mathscr{Z}_{\vartheta}}{\partial\eta^2} &= \frac{1}{\mathscr{Z}_{\vartheta}}\frac{\partial^2\mathscr{Z}_{\vartheta}}{\partial\eta^2} - \frac{1}{\mathscr{Z}_{\vartheta}^2}\left(\frac{\partial\mathscr{Z}_{\vartheta}}{\partial\eta}\right)^2 = \frac{1}{\mathscr{Z}_{\vartheta}}\frac{\partial^2\mathscr{Z}_{\vartheta}}{\partial\eta^2} - \left(\frac{\partial\ln\mathscr{Z}_{\vartheta}}{\partial\eta}\right)^2 \\
&= \vartheta^2\mathbb{E}_{\mu^{\vartheta}}\left(\sum_{i=1}^{n} y_i^2\right) - \vartheta^2\mathbb{E}_{\mu^{\vartheta}}\left(\sum_{i=1}^{n} y_i\right)^2.
\end{aligned}$$

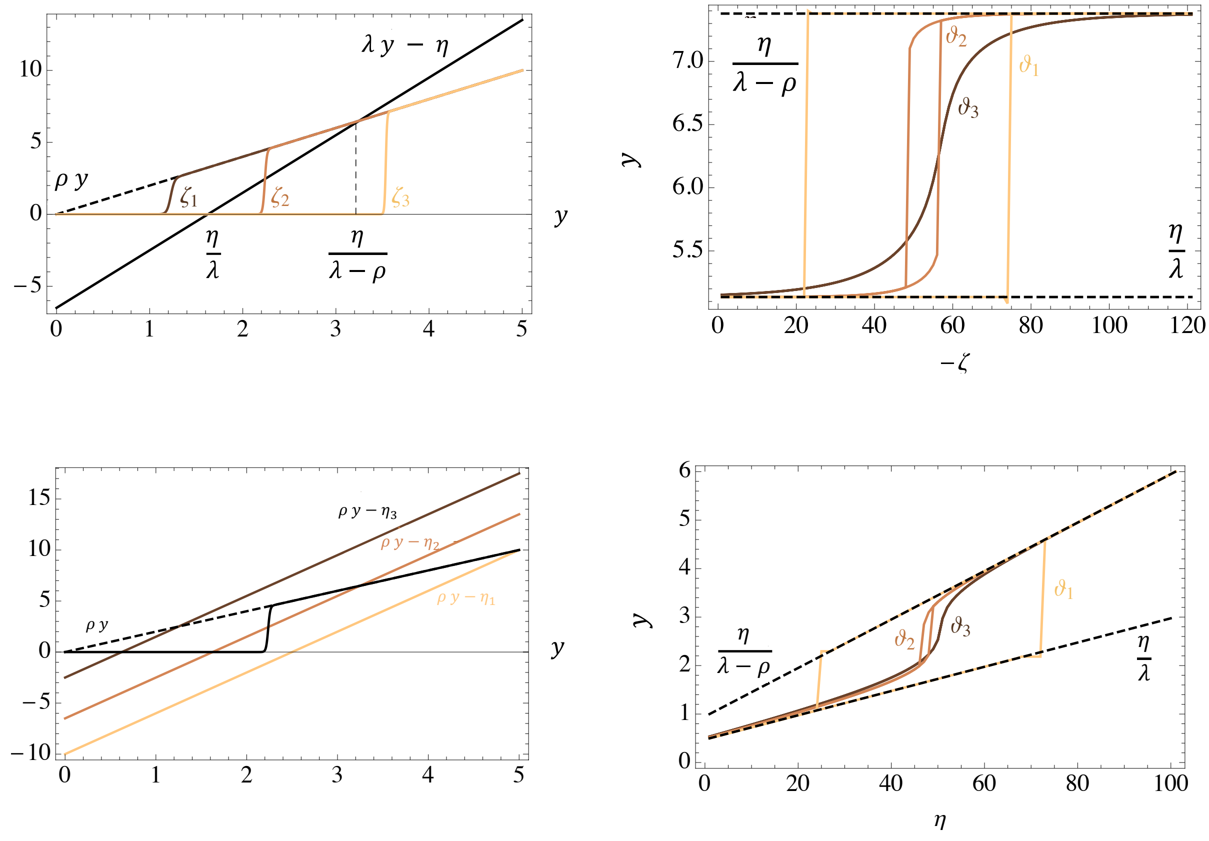


FIGURE D.5. (Top left panel) The right hand side of Equation (D.1) for different values of $\zeta_1 = 25$, $\zeta_2 = 10$, $\zeta_3 = 3$ and $\lambda = 4$, $\rho = 2$, $\eta = 6.5$, $\nu = 0$ and $\vartheta = 10$. (Top right panel) The values of y solving Equation (D.1) for different values of ζ with $\lambda = 1.48$, $\rho = 0.45$ and $\vartheta_1 = 49.5$, $\vartheta_2 = 0.495$, $\vartheta_3 = 0.2475$. (Bottom left panel) The right hand side of Equation (D.1) for different values of $\eta_1 = 2.5$, $\eta_2 = 6.5$, $\eta_3 = 10$ and $b = 4$, $\rho = 2$, $\zeta = 10$ and $\vartheta = 10$. (Bottom right panel) The values of y solving Equation (D.1) for different values of η with $\lambda = 4$, $\rho = 2$ and $\vartheta_1 = 10$, $\vartheta_2 = 0.26$, $\vartheta_3 = 0.2$.

We then get

$$\mathrm{Var}_{\mu^\vartheta}\left(\sum_{i=1}^n y_i\right) = \mathbb{E}_{\mu^\vartheta}\left(\sum_{i=1}^n y_i^2\right) - \mathbb{E}_{\mu^\vartheta}\left(\sum_{i=1}^n y_i\right)^2 = \frac{1}{\vartheta^2}\frac{\partial^2 \ln \mathscr{Z}_\vartheta}{\partial \eta^2} = -\frac{1}{\vartheta^2}\frac{\partial^2 \mathscr{F}_\vartheta}{\partial \eta^2}.$$

The variance of the mean is then given by

$$\mathrm{Var}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^n y_i\right) = -\frac{1}{n^2\vartheta^2}\frac{\partial^2 \mathscr{F}_\vartheta}{\partial \eta^2}.$$

We have that

$$\frac{\partial^2 \mathscr{F}_\vartheta}{\partial \eta^2} = -\vartheta \frac{\partial^2 \mathscr{H}_\vartheta(\mathbf{y}^*)}{\partial \eta^2} = 0,$$

and we get

$$\mathrm{Var}_{\mu^*}\left(\frac{1}{n}\sum_{i=1}^n y_i\right) = \lim_{\vartheta\to\infty} \mathrm{Var}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^n y_i\right) = 0.$$

Note that the variance of the average R&D effort can be equal to zero only if it is equal to its expectation in all of its support. This can only happen if the average R&D effort is equal to y^* with probability one in the large ϑ limit.

Further, in the limit of $\vartheta \to \infty$, for both, the low equilibrium, where $y = \frac{\eta^*}{\lambda+2\nu^*}$ and therefore $\frac{\eta^*}{y} = \lambda + 2\nu^*$, as well as the high equilibrium, where $y = \frac{\eta^*}{\lambda+2\nu^*-\rho}$, and $\frac{\eta^*}{y} = \lambda + 2\nu^* - \rho$ we find from Equation (D.18) that

$$\frac{\partial}{\partial \zeta}\left.\frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i^2}\right|_{y_i=y} = 0,$$

and from Equation (D.19) we get

$$\frac{\partial}{\partial \zeta} \left. \frac{\partial^2 \mathscr{H}_\vartheta}{\partial y_i \partial y_j} \right|_{y_i=y_j=y} = 0.$$

Hence, we find that in the high equilibrium $\mathbb{E}_{\mu^*}(m) = \lim_{\vartheta\to\infty} \mathbb{E}_{\mu^\vartheta}(m) = \frac{n(n-1)}{2}$, while in the low equilibrium $\mathbb{E}_{\mu^*}(m) = \lim_{\vartheta\to\infty} \mathbb{E}_{\mu^\vartheta}(m) = 0$. Consequently, the expected average degree in the high equilibrium is $\mathbb{E}_{\mu^*}\left(\frac{1}{n}\sum_{i=1}^n d_i\right) = \lim_{\vartheta\to\infty} \mathbb{E}_{\mu^\vartheta}\left(\frac{1}{n}\sum_{i=1}^n d_i\right) = n-1$, where we have a complete graph, K_n, and zero in the low equilibrium where we obtain an empty graph, $\overline{K}_n$. □

Appendix E: Additional Sources of Firm Heterogeneity

In the following sections we will discuss two possible extensions of the model that incorporate firm heterogeneity (see also the empirical application in Section 3.2). First, in Section E.1 we allow for heterogeneous collaboration costs. Second, in Section E.2 we incorporate heterogeneous spillovers between collaborating firms.

E.1. Heterogeneous Marginal Collaboration Costs

In the following we assume that the marginal cost of collaboration between firms i and j can be written as an additively separable function: $\zeta_{ij} = z_i + z_j$ (cf. Equation (12) in Section 3.2). The probability of a link between firms i and j is then given by Equation (8).[10] It can be written as

$$p^\vartheta(y_i, z_i, y_j, z_j) = \frac{e^{\vartheta(\rho y_i y_j - z_i - z_j)}}{1 + e^{\vartheta(\rho y_i y_j - z_i - z_j)}}. \tag{E.1}$$

Similar to e.g., Melitz (2003); Melitz et al. (2008) we further assume that the cost parameter $z_i \geq 0$ is given by the inverse of the firm's productivity, $z_i = \frac{1}{\varphi_i}$, where $\varphi_i > 0$ is the productivity (or efficiency) of firm i. Then firms with higher productivity incur lower collaboration costs. Moreover, assume that the firms' productivities, $\varphi_i \geq c > 0$, are Pareto distributed (Melitz et al. 2008; König et al. 2016), with density $f(\varphi) = \frac{\gamma}{c}\left(\frac{c}{\varphi}\right)^{\gamma+1}$ and $\varphi > c$, where $c > 0$ is a lower cut-off, and $\gamma > 0$ is a positive parameter. The complementary distribution function is then given by $F(\varphi) = 1 - \left(\frac{c}{\varphi}\right)^\gamma$. It follows that the cost $z = \frac{1}{\varphi}$ has the density $f(z) = \gamma c^\gamma z^{\gamma-1}$ for $z \in \left(0, \frac{1}{c}\right)$, and the cumulative distribution function $F(z) = (cz)^\gamma$. For this specification of the cost, we can derive various network statistics that are consistent with real-world networks. In particular, we show below that the degree distribution is highly skewed and follows a power law (Proposition E.1), the average degree of the neighbors of a node is correlated with the degree of the node (Proposition E.2), and the clustering coefficient of a node (a measure for the tendency of the nodes in a graph to cluster together) is also correlated with the degree of a node (Proposition E.3).[11] An illustration is given in Figure E.1.[12]

10. Similar specifications can be found in the empirical literature on network formation (see e.g., Graham 2015, for a review). For example, Graham (2017) and Fafchamps and Gubert (2007) consider an econometric network formation model in which the probability of a link between agents i and j is given by $\mathbb{P}(a_{ij} = 1) = e^{Y_i + Y_j + \mathbf{Z}_{ij}^\top \boldsymbol{\beta}}/(1 + e^{Y_i + Y_j + \mathbf{Z}_{ij}^\top \boldsymbol{\beta}})$ where Y_i is an agent specific fixed effect and $\mathbf{Z}_{ij}$ is a vector of pair-specific covariates. Similarly, Chatterjee et al. (2011) analyze a network formation model with linking probability $\mathbb{P}(a_{ij} = 1) = e^{Y_i + Y_j}/(1 + e^{Y_i + Y_j})$.

11. In the following propositions, we will assume that the firms' R&D effort levels are concentrated on y^* in the limit of $\vartheta \to \infty$, and the assumption that is typically satisfied in the simulation studies that we did. Moreover, concentration can be shown to hold in the basic model with homogeneous firms in the limit of vanishing noise.

12. See e.g., König (2016) for a comparison and additional details about these statistics.

Proposition E.1. *Assume that the firms' R&D effort levels are concentrated on y^* in the limit of $\vartheta \to \infty$. Then under a continuum approximation,*[13] *the degree distribution is given by*

$$P(k) = \frac{c^\gamma}{k}\left(\rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right)^{\gamma-1}\left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}, \tag{E.2}$$

and for large k the degree distribution $P(k)$ decays as $O\left(k^{-\frac{\gamma-1}{\gamma}}\right)$.

Proof of Proposition E.1. The generating function of the degree $d_1(G)$ is given by

$$\begin{aligned}\mathbb{E}\left(y_1^{d_1(G)}\right) &= \mathbb{E}\left(\mathbb{E}\left(y_1^{d_1(G)}\middle| y_1, z_1\right)\right)\\ &= \mathbb{E}\left(\left(\mathbb{E}\left(\frac{1+p(y_1,z_1,y_2,z_2)y_1}{1+p(y_1,z_1,y_2,z_2)}\middle| y_1, z_1\right)\right)^{n-1}\right),\end{aligned}$$

With the cost distributed as $f(z) = \gamma c^\gamma z^{\gamma-1}$ for $z \in \left(0, \frac{1}{c}\right)$, we can write

$$\begin{aligned}\mathbb{E}\left(y_1^{d_1(G)}\middle| y_1 = y, z_1 = z\right) &= \mathbb{E}\left(\frac{1+p(y,z,y_2,z_2)y_1}{1+p(y,z,y_2,z_2)}\middle| x, z\right)^{n-1}\\ &= \left(1 + (y_1-1)\mathbb{E}\left(p(y,z,y_2,z_2)\right| y_1 = y, z_1 = z)\right)^{n-1}\\ &= \left(1 + (y_1-1)\int_{\mathcal{Y}} dy'\mu^\vartheta(y')\int dz'\gamma c^\gamma (z')^{\gamma-1} p(y,z,y',z')\right)^{n-1}\\ &= \left(1 + (y_1-1)\int_{\mathcal{Y}} dy'\mu^\vartheta(y')\int dz'\gamma c^\gamma (z')^{\gamma-1}\frac{e^{\vartheta(\rho y y' - z - z')}}{1+e^{\vartheta(\rho y y' - z - z')}}\right)^{n-1}.\end{aligned}$$

In the limit of $\vartheta \to \infty$ in Equation (E.1) we obtain

$$\lim_{\vartheta\to\infty} p^\vartheta(y,z,y',z') = \lim_{\vartheta\to\infty}\frac{e^{\vartheta(\rho y y' - z - z')}}{1+e^{\vartheta(\rho y y' - z - z')}} = \mathbb{1}_{\{\rho y y' > z + z'\}},$$

so that we can write

$$\begin{aligned}\mathbb{E}\left(y_1^{d_1(G)}\middle| y_1 = y, z_1 = z\right) &= \left(1 + (y_1-1)\int_{\mathcal{Y}} dy'\mu^\vartheta(y')\int dz'\gamma c^\gamma (z')^{\gamma-1}\mathbb{1}_{\{\rho y y' > z + z'\}}\right)^{n-1}\\ &= \left(1 + (y_1-1)\gamma c^\gamma\int_0^{\rho y y^* - z} dz'(z')^{\gamma-1}\right)^{n-1}\\ &= \left(1 + (y_1-1)c^\gamma\left(\rho y y^* - z\right)^\gamma\right)^{n-1}\\ &= e^{(y_1-1)(n-1)c^\gamma(\rho y y^* - z)^\gamma}.\end{aligned}$$

This is the generating function of a Poisson random variable with expectation and variance given by $\bar{d}(y,z) \equiv (n-1)c^\gamma\left(\rho y y^* - z\right)^\gamma$. When the cut-off c is small, the variance becomes small, and we can approximate the Poisson random variable with a constant random variable at the expected value. Making further a continuum approximation, where we treat the degree as a continuous variable, we can write

$$\mathbb{P}\left(d_1(G) = k\right| y_1 = y, z_1 = z) = \delta\left(k - \bar{d}(y,z)\right) = \delta\left(k - (n-1)c^\gamma\left(\rho y y^* - z\right)^\gamma\right).$$

Note that under the continuum approximation there exists a one-to-one mapping from the degree k to the cost z, where for a given k and R&D effort y, the cost s is given by

$$z = \rho y y^* - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}.$$

13. This is an approximation that has shown to be accurate in various network formation models as the network size become large (Dorogovtsev and Mendes 2013, pp. 117).

Using the fact that[14]

$$\delta\left(k-(n-1)c^{\gamma}\left(\rho(y^*)^2-z\right)^{\gamma}\right)=\delta\left(z-\left(\rho(y^*)^2-\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\right)\right)\frac{1}{\gamma k}\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}, \quad \text{(E.3)}$$

and assuming that the R&D effort distribution concentrates on y^*, the degree distribution is given by

$$\begin{aligned}
P(k) &= \int dz\mathbb{P}\left(d_1(G)=k|\, y_1=y^*, z_1=z\right)f(z)\\
&= \gamma c^{\gamma}\int dz\delta\left(z-(n-1)c^{\gamma}\left(\rho(y^*)^2-z\right)^{\gamma}\right)z^{\gamma-1}\\
&= \gamma c^{\gamma}\int dz\delta\left(k-\left(\rho(y^*)^2-\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\right)\right)\frac{1}{\gamma k}\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}z^{\gamma-1}\\
&= \frac{c^{\gamma}}{k}\left(\rho(y^*)^2-\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\right)^{\gamma-1}\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\\
&= O\left(k^{-\frac{\gamma-1}{\gamma}}\right).
\end{aligned}$$

□

Hence, we obtain a power law degree distribution with parameter $\frac{\gamma-1}{\gamma}$, consistent with previous empirical studies which have found power law degree distributions in R&D alliance networks (e.g., Powell et al. 2005). An illustration can be seen in Figure E.1 for the case of $\gamma=2$ and $n=200$ firms.

We next analyze correlations between the average degree of a firm and its neighbors in the network.

Proposition E.2. *Assume that the firms R&D effort levels are concentrated on y^* in the limit of $\vartheta\to\infty$. Then under the continuum approximation, the average nearest neighbor degree distribution is given by*

$$k_{\mathrm{nn}}(k)=1+\frac{(n-1)^2\gamma c^{2\gamma}}{k}\int_0^{\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}}dz'(z')^{\gamma-1}(\rho(y^*)^2-z')^{\gamma}. \quad \text{(E.4)}$$

Proof of Proposition E.2. Next, we compute the average nearest neighbor degree distribution

$$k_{\mathrm{nn}}(k)=1+\frac{1}{P(k)}\int dz\int_{\mathcal{Y}}dyf(z)\mu^{\vartheta}(y)g(k|y,z)\tilde{k}_{\mathrm{nn}}(y,z),$$

where

$$\begin{aligned}
\tilde{k}_{\mathrm{nn}}(y,z) &= \int dz'\int_{\mathcal{Y}}dy'p(y',z'|y,z)\bar{d}(y',z'),\\
g(k|y,z) &= \mathbb{P}\left(d_1(G)=k|\,y_1=y,z_1=z\right)=\delta\left(k-\bar{d}(y,z)\right),\\
\bar{d}(y,z) &= (n-1)c^{\gamma}\left(\rho yy^*-z\right)^{\gamma}\\
p(y',z'|y,z) &= \frac{(n-1)p(y,z,y',z')f(z')\mu^{\vartheta}(y')}{\bar{d}(y,z)},\\
\lim_{\vartheta\to\infty}p^{\vartheta}(y,z,y',z') &= \lim_{\vartheta\to\infty}\frac{e^{\vartheta(\rho yy'-z-z')}}{1+e^{\vartheta(\rho yy'-z-z')}}=\mathbb{1}_{\{\rho yy'>z+z'\}}\\
\mu^{\vartheta}(y) &= \delta(y-y^*)\\
f(z) &= \gamma c^{\gamma}z^{\gamma-1}.
\end{aligned}$$

14. When $g(y)$ is a continuously differentiable function in $\mathbb{R}$ it holds that $\delta(g(y))=\sum_{i=1}^{m}\frac{\delta(y-y_i)}{|g'(y_i)|}$ where the m roots y_i satisfy $g(y_i)=0$ for all $i=1,\ldots,m$.

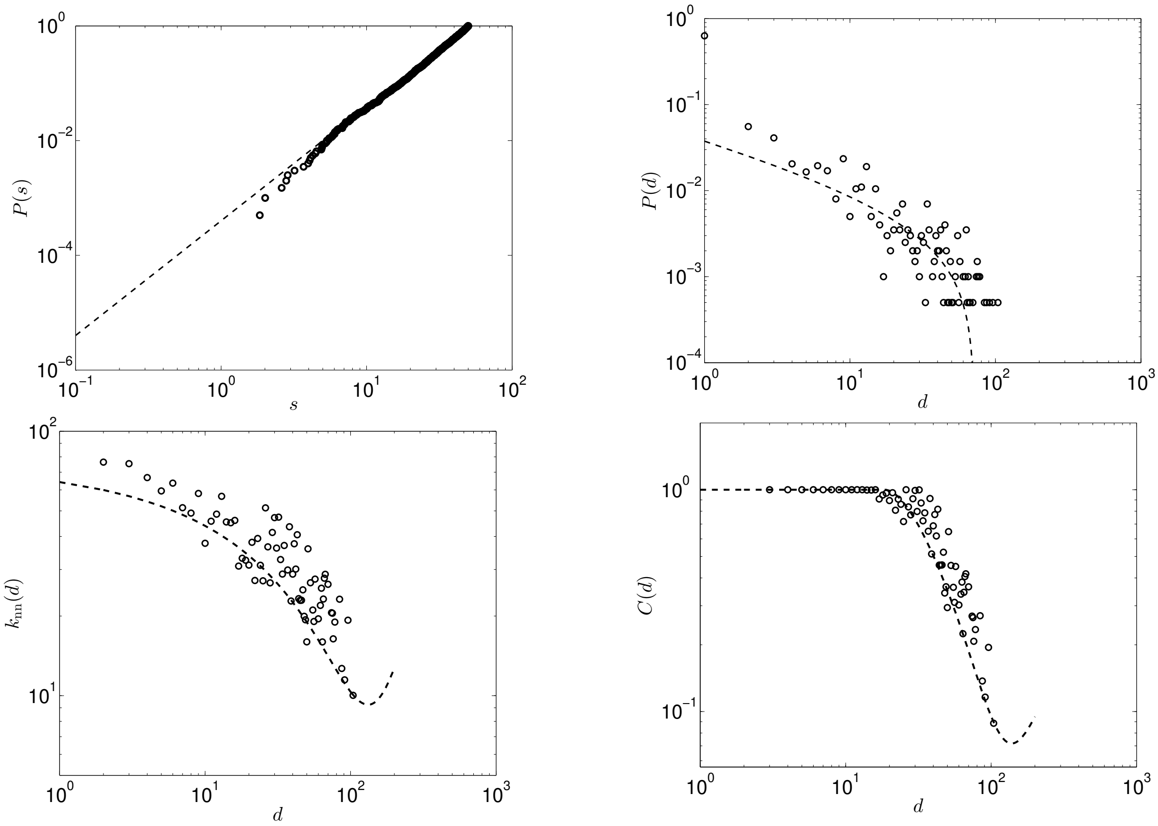


Figure E.1. (Top left panel) The empirical and the theoretical cumulative cost distribution $F(s) = (cs)^\gamma$ with $\gamma = 2$ and $c = 0.02$. The empirical distribution is indicated with circles and the theoretical distribution with a dashed line. (Top right panel) The degree distribution $P(d)$. The dashed line indicates the theoretical prediction of Equation (E.2). (Bottom left panel) The average nearest neighbor degree distribution $k_{\rm nn}(d)$, decreasing with increasing degrees d and thus indicating a disassortative network. The dashed line indicates the theoretical prediction of Equation (E.4). (Bottom right panel) The clustering degree distribution $C(d)$, decreasing with increasing degree d. The parameters used are $b = 0.75$, $\nu = 1$ and $\rho = 1$. The distributions are computed across 10 independent simulation runs with $n = 200$ firms.

It then follows that

$$p(y', z'|y, z) = \frac{(n-1)\mathbb{1}_{\{\rho y y' > z + z'\}} \gamma c^\gamma (z')^{\gamma-1} \delta(y' - y^*)}{\bar{d}(y, z)},$$

and therefore

$$\begin{aligned}
\tilde{k}_{\rm nn}(y, z) &= \frac{n-1}{\bar{d}(y, z)} \int dz' f(z') \int dy' \delta(y - y^*) \mathbb{1}_{\{\rho y y' > z + z'\}} \bar{d}(y', z') \\
&= \frac{n-1}{\bar{d}(y, z)} \int dz' f(z') \mathbb{1}_{\{\rho y y^* > z + z'\}} \bar{d}(y^*, z') \\
&= \frac{n-1}{\bar{d}(y, z)} \int_0^{\rho y y^* - z} dz' f(z') \bar{d}(y^*, z') \\
&= c^{-\gamma} (\rho y y^* - z)^{-\gamma} \int_0^{\rho y y^* - z} dz' \gamma c^\gamma (z')^{\gamma-1} (n-1) c^\gamma (\rho (y^*)^2 - z')^{-\gamma} \\
&= \frac{(n-1)\gamma c^\gamma J(y, z)}{(\rho y y^* - z)^\gamma},
\end{aligned}$$

where we have denoted

$$J(y, z) \equiv \int_0^{\rho y y^* - z} dz' (\rho (y^*)^2 - z')^\gamma (z')^{\gamma-1}.$$

We then get

$$
\begin{aligned}
k_{\mathrm{nn}}(k) &= 1 + \frac{1}{P(k)} \int dz \int_{\mathcal{Y}} dy f(z) \delta(y - y^*) \delta\left(k - \bar{d}(y, z)\right) \tilde{k}_{\mathrm{nn}}(y, z) \\
&= 1 + \frac{1}{P(k)} \int dz \gamma c^{\gamma} z^{\gamma - 1} \delta\left(k - \bar{d}(y^*, z)\right) \tilde{k}_{\mathrm{nn}}(y^*, z) \\
&= 1 + \frac{1}{P(k)} \int dz f(z) \delta\left(k - \bar{d}(y^*, z)\right) \frac{(n-1)\gamma c^{\gamma} J(y^*, z)}{(\rho y y^* - z)^{\gamma}}.
\end{aligned}
$$

Using Equation (E.3) we can write this as

$$
\begin{aligned}
k_{\mathrm{nn}}(k) &= 1 + \frac{1}{P(k)} \gamma c^{\gamma} \left(\rho (y^*)^2 - \left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\right)^{\gamma - 1} \frac{1}{\gamma k} \left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}} \\
&\quad \times (n-1) c^{\gamma} \gamma J\left(y^*, \rho (y^*)^2 - \left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\right) \\
&= 1 + \frac{(n-1)^2 \gamma c^{2\gamma}}{k} J\left(y^*, \rho (y^*)^2 - \left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}\right) \\
&= 1 + \frac{(n-1)^2 \gamma c^{2\gamma}}{k} \int_0^{\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}} dz' (z')^{\gamma - 1} (\rho (y^*)^2 - z')^{\gamma}.
\end{aligned}
$$

□

Figure E.1 shows the results from numerical simulations compared with the theoretical prediction of Equation (E.4). Heterogeneous collaboration costs thus induce degree correlations between a node and its neighbors, consistent with real-world networks.

The following proposition derives the clustering coefficient as a function of the degree of a node in the network.

Proposition E.3. *Assume that the firms R&D effort levels are concentrated on y^* in the limit of $\vartheta \to \infty$. Then under the continuum approximation, the clustering coefficient is given by*

$$
\begin{aligned}
C(k) &= \mathbb{1}_{\left\{k < (n-1)\left(\frac{\rho (y^*)^2 c}{2}\right)^{\gamma}\right\}} + \mathbb{1}_{\left\{k > (n-1)\left(\frac{\rho (y^*)^2 c}{2}\right)^{\gamma}\right\}} \frac{(n-1)c^{\gamma}}{k} \\
&\quad \times \left(1 + \gamma \frac{(n-1)c^{\gamma}}{k} \int_{\rho (y^*)^2 - \left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}}^{\left(\frac{k}{(n-1)c^{\gamma}}\right)^{\frac{1}{\gamma}}} dz' (z')^{\gamma - 1} \left(\rho (y^*)^2 - z'\right)^{\gamma}\right),
\end{aligned}
\tag{E.5}
$$

and for large k the clustering coefficient $C(k)$ decays as $O\left(\frac{1}{k}\right)$.

Proof of Proposition E.3. Next we analyze the clustering coefficient of a firm with degree k, which can be written as

$$
\begin{aligned}
C(k) &= \frac{1}{P(k)} \int dz \int_{\mathcal{Y}} dy f(z) \delta(y - y^*) g(k|y, z) \tilde{C}(y, z) \\
&= \frac{1}{P(k)} \int dz f(z) g(k|y^*, z) \tilde{C}(y^*, z),
\end{aligned}
$$

where

$$
\tilde{C}(y^*, z) = \int dz' \int dz'' \int_{\mathcal{Y}} dy' \int_{\mathcal{Y}} dy'' p(y', z', y'', z'') p(y', z'|y^*, z) p(y'', z''|y^*, z).
$$

This can further be written as follows

$$
\begin{aligned}
\tilde{C}(y^*,z) &= \int dz' \int dz'' \int_{\mathcal{Y}} dy' \int_{\mathcal{Y}} dy'' \mathbb{1}_{\{\rho y' y'' > z' + z''\}} \\
&\quad \times \frac{(n-1)\mathbb{1}_{\{\rho y^* y' > z+z'\}} f(z')\delta(y'-y^*)}{\bar{d}(y^*,z)} \frac{(n-1)\mathbb{1}_{\{\rho y^* y'' > z+z''\}} f(z'')\delta(y''-y^*)}{\bar{d}(y^*,z)} \\
&= \frac{(n-1)^2}{\bar{d}(y^*,z)^2} \int dz' f(z') \int dz'' f(z'') \mathbb{1}_{\{\rho(y^*)^2 > z'+z''\}} \mathbb{1}_{\{\rho(y^*)^2 > z+z'\}} \mathbb{1}_{\{\rho(y^*)^2 > z+z''\}} \\
&= \frac{(n-1)^2}{\bar{d}(y^*,z)^2} \Bigg(\mathbb{1}_{\left\{z > \frac{\rho(y^*)^2}{2}\right\}} \int_0^{\rho(y^*)^2 - z} dz' f(z') \int_0^{\rho(y^*)^2 - z} dz'' f(z'') \\
&\quad + \mathbb{1}_{\left\{z < \frac{\rho(y^*)^2}{2}\right\}} \left(\int_0^z dz' f(z') \int_0^{\rho(y^*)^2 - z} dz'' f(z'') + \int_z^{\rho(y^*)^2 - z} dz' f(z') \int_0^{\rho(y^*)^2 - z'} dz'' f(z'') \right) \Bigg) .
\end{aligned}
$$

We then get (see also Figure E.2)

$$
\begin{aligned}
\tilde{C}(y^*,z) &= \frac{\gamma^2 c^{2\gamma}(n-1)^2}{\bar{d}(y^*,z)} \Bigg(\mathbb{1}_{\left\{z < \frac{\rho(y^*)^2}{2}\right\}} \int_0^{\rho(y^*)^2 - z} dz'(z')^{\gamma-1} \int_0^{\rho(y^*)^2 - z} dz''(z'')^{\gamma-1} \\
&\quad + \mathbb{1}_{\left\{z > \frac{\rho(y^*)^2}{2}\right\}} \Bigg(\int_0^z dz'(z')^{\gamma-1} \int_0^{\rho(y^*)^2 - z} dz''(z'')^{\gamma-1} \\
&\quad + \int_z^{\rho(y^*)^2 - z} dz'(z')^{\gamma-1} \int_0^{\rho(y^*)^2 - z'} dz''(z'')^{\gamma-1} \Bigg) \Bigg) \\
&= \frac{\gamma c^{2\gamma}(n-1)^2}{\bar{d}(y^*,z)} \Bigg(\mathbb{1}_{\left\{z < \frac{\rho(y^*)^2}{2}\right\}} \int_0^{\rho(y^*)^2 - z} dz'(z')^{\gamma-1} \left(\rho(y^*)^2 - z\right)^\gamma \\
&\quad + \mathbb{1}_{\left\{z > \frac{\rho(y^*)^2}{2}\right\}} \left(\int_0^z dz'(z')^{\gamma-1} \left(\rho(y^*)^2 - z\right)^\gamma + \int_z^{\rho(y^*)^2 - z} dz'(z')^{\gamma-1} \left(\rho(y^*)^2 - z'\right)^\gamma \right) \Bigg) \\
&= \frac{\gamma c^{2\gamma}(n-1)^2}{\bar{d}(y^*,z)} \Bigg(\mathbb{1}_{\left\{z < \frac{\rho(y^*)^2}{2}\right\}} \frac{1}{\gamma} \left(\rho(y^*)^2 - z\right)^{2\gamma} \\
&\quad + \mathbb{1}_{\left\{z > \frac{\rho(y^*)^2}{2}\right\}} \left(\frac{1}{\gamma} z^\gamma \left(\rho(y^*)^2 - z\right)^\gamma + \int_z^{\rho(y^*)^2 - z} dz'(z')^{\gamma-1} \left(\rho(y^*)^2 - z'\right)^\gamma \right) \Bigg) \\
&= \frac{c^{2\gamma}(n-1)^2}{\bar{d}(y^*,z)} \left(\mathbb{1}_{\left\{z < \frac{\rho(y^*)^2}{2}\right\}} \left(\rho(y^*)^2 - z\right)^{2\gamma} + \mathbb{1}_{\left\{z > \frac{\rho(y^*)^2}{2}\right\}} \left(z^\gamma \left(\rho(y^*)^2 - z\right)^\gamma + \gamma J(z) \right) \right),
\end{aligned}
\tag{E.6}
$$

where we have denoted by

$$
J(z) \equiv \int_z^{\rho(y^*)^2 - z} dz'(z')^{\gamma-1} \left(\rho(y^*)^2 - z'\right)^\gamma .
$$

Using the fact that $\bar{d}(y,z) \equiv (n-1)c^\gamma \left(\rho y y^* - z\right)^\gamma$ this can be written as

$$
\tilde{C}(y^*,z) = \frac{1}{\left(\rho(y^*)^2 - z\right)^{2\gamma}} \left(\mathbb{1}_{\left\{z < \frac{\rho(y^*)^2}{2}\right\}} \left(\rho(y^*)^2 - z\right)^{2\gamma} + \mathbb{1}_{\left\{z > \frac{\rho(y^*)^2}{2}\right\}} \left(z^\gamma \left(\rho(y^*)^2 - z\right)^\gamma + \gamma J(z) \right) \right) .
\tag{E.7}
$$

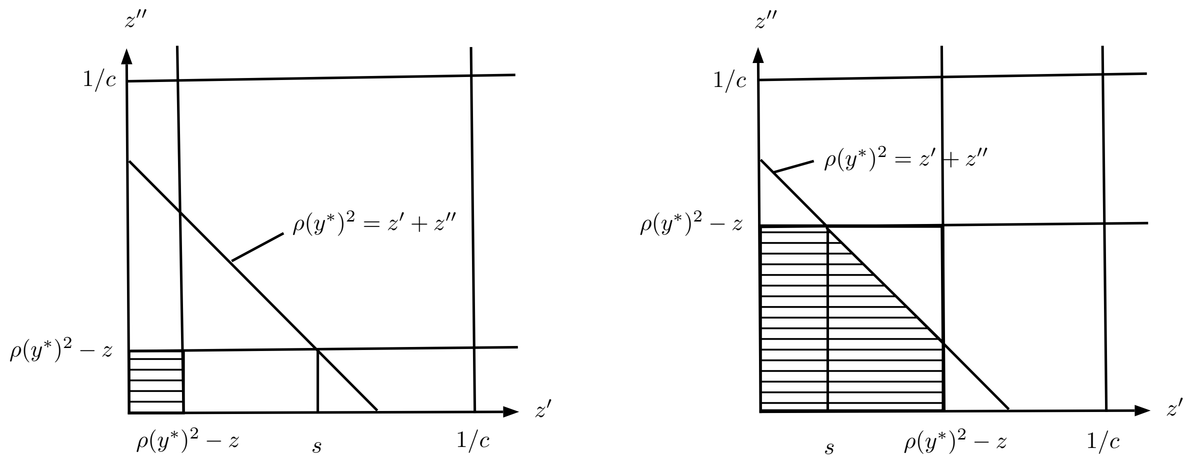


FIGURE E.2. The area of integration in Equation (E.6) for the case of $z < \frac{\rho(y^*)^2}{2}$ (left panel) and for the case of $z > \frac{\rho(y^*)^2}{2}$ (right panel).

Hence we get

$$
\begin{aligned}
C(k) &= \frac{1}{P(k)} \int dz f(z) \delta\left(k - \bar{d}(y^*, z)\right) \tilde{C}(y^*, z) \\
&= \frac{1}{P(k)} \int dz \gamma c^\gamma z^{\gamma-1} \delta\left(z - \left(\rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right)\right) \frac{1}{\gamma k} \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}} \tilde{C}(y^*, z) \\
&= \frac{1}{P(k)} \gamma c^\gamma \left(\rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right)^{\gamma-1} \frac{1}{\gamma k} \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}} \tilde{C}\left(y^*, \rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right) \\
&= \tilde{C}\left(y^*, \rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right).
\end{aligned}
$$

Inserting Equation (E.7) this gives

$$
\begin{aligned}
C(k) = \mathbb{1}_{\left\{k<(n-1)\left(\frac{\rho(y^*)^2 c}{2}\right)^\gamma\right\}} &+ \mathbb{1}_{\left\{k>(n-1)\left(\frac{\rho(y^*)^2 c}{2}\right)^\gamma\right\}} \left(\frac{(n-1)c^\gamma}{k}\right. \\
&+ \gamma \left(\frac{(n-1)c^\gamma}{k}\right)^2 J\left(\rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right)\Bigg) \\
= \mathbb{1}_{\left\{k<(n-1)\left(\frac{\rho(y^*)^2 c}{2}\right)^\gamma\right\}} &+ \mathbb{1}_{\left\{k>(n-1)\left(\frac{\rho(y^*)^2 c}{2}\right)^\gamma\right\}} \left(\frac{(n-1)c^\gamma}{k}\right. \\
&+ \gamma \left(\frac{(n-1)c^\gamma}{k}\right)^2 \int_{\rho(y^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}}^{\left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}} dz' (z')^{\gamma-1} \left(\rho(y^*)^2 - z'\right)^\gamma \Bigg).
\end{aligned}
$$

□

Figure E.1 shows the results from numerical simulations compared with the theoretical prediction of Equation (E.5). The figure further illustrates that the model can generate two-node and three-node degree correlations, such as a decreasing average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, indicating a disassortative network, as well as a decreasing clustering degree distribution, $C(d)$, with the degree d.

E.2. Heterogeneous Technology Spillovers

In this section we allow for heterogeneity among firms in terms of their technological abilities and the spillovers they can generate (Griffith et al. 2003). We assume that the technologies embodied in a firm $i \in \mathcal{N} = \{1, \ldots, n\}$ can be represented as an N-dimensional vector $\mathbf{h}_i$ in the technology space $\mathcal{H}^N = \{0, 1\}^N$, which consists of all binary sequences with elements in $\{0, 1\}$ of length N. The number of such sequences is 2^N. The technology vector $\mathbf{h}_i$, with components $h_{ik} \in \{0,1\}$, indicates whether firm i knows idea $k \in \{1,\ldots,N\}$ or not. We introduce a spillover function $f : \mathcal{H}^N \times \mathcal{H}^N \to \mathbb{R}$ capturing the potential technology transfer between any pairs of firms (cf. Equation (12). A possible specification is one in which $f(\mathbf{h}_i, \mathbf{h}_j) = \mathbb{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle > \tau\}}$, where $\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $\mathbb{R}^n$, so that $\langle \mathbf{h}_i, \mathbf{h}_j \rangle$ counts the number of technologies known to both i and j, and $f(\mathbf{h}_i, \mathbf{h}_j)$ is one iff i and j have at least $\tau > 0$ technologies in common. This is an instance of a *random intersection graph* (Deijfen and Kets 2009) (see also Supplementary Appendix B).[15]

Given the spillover function $f(\mathbf{h}_i, \mathbf{h}_j)$, the marginal cost of production of a firm i becomes (cf. Equation (C.1) in Supplementary Appendix C)

$$c_i = \bar{c} - \alpha e_i - \beta \sum_{j=1}^{n} a_{ij} f(\mathbf{h}_i, \mathbf{h}_j) e_j,$$

and profits of firm i (from Equation (C.2) in Supplementary Appendix C) are given by

$$\pi_i = (a - \bar{c}) q_i - q_i^2 - b q_i \sum_{j \neq i} q_j + \alpha q_i e_i + \beta q_i \sum_{j=1}^{n} a_{ij} f(\mathbf{h}_i, \mathbf{h}_j) e_j - \gamma e_i^2 - \zeta d_i.$$

The optimal effort levels are given by $e_i = \frac{\alpha}{2\gamma} q_i = \tau q_i$. Inserting into profits and denoting by $\eta_i = a - \bar{c}_i$, $\nu = 1 - \tau\alpha + \tau^2\gamma$, $\rho = \tau\beta$ and $\lambda = b$ yields

$$\pi_i = \eta q_i - \nu q_i^2 - \lambda q_i \sum_{j \neq i} q_j + \rho q_i \sum_{j=1}^{n} a_{ij} f(\mathbf{h}_i, \mathbf{h}_j) q_j - \zeta d_i.$$

We can then obtain a potential function (see Proposition 1) given by

$$\Phi(\mathbf{q}, G, \mathbf{h}) = \sum_{i=1}^{n} ((a - \bar{c}) q_i - \nu q_i^2) - \frac{\lambda}{2} \sum_{i=1}^{n} q_i \sum_{j \neq i} q_j + \sum_{i=1}^{n} q_i \sum_{j=1}^{n} a_{ij} f(\mathbf{h}_i, \mathbf{h}_j) q_j - \zeta m.$$

The stationary distribution (see Theorem 1) is given by

$$\mu^{\vartheta}(\mathbf{q}, G, \mathbf{h}) = \frac{e^{\vartheta \Phi(\mathbf{q}, G, \mathbf{h})}}{\sum_{\mathbf{h}' \in \mathcal{H}^N} \sum_{G' \in \mathcal{G}^n} \int e^{\vartheta \Phi(\mathbf{s}, G', \mathbf{h}')} d\mathbf{s}}.$$

The probability of observing a network $G \in \mathcal{G}^n$, given an R&D effort distribution $\mathbf{q} \in [0, \bar{q}]^n$ and technology portfolios $\mathbf{h} \in \mathcal{H}^N$ is determined by the conditional distribution

$$\mu^{\vartheta}(G | \mathbf{q}, \mathbf{h}) = \prod_{i<j} \frac{e^{\vartheta a_{ij} (\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}{1 + e^{\vartheta (\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}, \tag{E.8}$$

which is equivalent to the probability of observing an inhomogeneous random graph with link probability

$$p^{\vartheta}(q_i, \mathbf{h}_i, q_j, \mathbf{h}_j) \equiv \frac{e^{\vartheta (\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}{1 + e^{\vartheta (\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}. \tag{E.9}$$

In the following we consider a particularly simple specification in which each firm i is assigned a technology $k \in \{1,\ldots,N\}$ uniformly at random so that $h_{ik} = 1$ and $h_{il} = 0$ for all $l \neq k$. Moreover, let $f(\mathbf{h}_i, \mathbf{h}_j) = \mathbb{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle \geq 1\}}$, that is, firms i an j can only benefit from a collaboration if they have a technology in common. For this model we can derive analytically various network statistics of interest, as the following proposition illustrates.

15. There is a variety of other functional forms that can be incorporated in our model. For example, a simple choice for the function f could be $f(\mathbf{h}_i, \mathbf{h}_j) = a|\mathbf{h}_i \cap \mathbf{h}_j|$, where $a \in \mathbb{R}_+$ and $|\mathbf{h}_i \cap \mathbf{h}_j| = \mathbf{h}_i^\top \mathbf{h}_j = \sum_{k=1}^{N} h_{ik} h_{jk}$ denotes the common knowledge of i and j. Alternative specifications for similarity can be found in Liben-Nowell and Kleinberg (2007) and Bloom et al. (2013); Jaffe (1989). Alternatively, following Berliant and Fujita (2008, 2009), a possible parametric specification for f would be $f(\mathbf{h}_i, \mathbf{h}_j) = |\mathbf{h}_i \cap \mathbf{h}_j|^{\kappa} d(\mathbf{h}_i, \mathbf{h}_j)^{\frac{1-\kappa}{2}}$ for some $\kappa \in (0,1)$. The distance is the product of the total number of ideas known by agent i but not by j times the total number of ideas known by j but not by i, i.e., $d(\mathbf{h}_i, \mathbf{h}_j) = |\mathbf{h}_i \backslash \mathbf{h}_j| \times |\mathbf{h}_j \backslash \mathbf{h}_i| = |\mathbf{h}_i \cap \mathbf{h}_j^c| \times |\mathbf{h}_i^c \cap \mathbf{h}_j| = \sum_{k=1}^{N} h_{ik}(1 - h_{jk}) \sum_{k=1}^{N} (1 - h_{ik}) h_{jk}$, where $\mathbf{u} = (1,\ldots,1)^\top$ and $\mathbf{h}_i^c = \mathbf{u} - \mathbf{h}_i$. Other functional forms have been suggested in the literature (see e.g., Baum et al. 2009; Nooteboom et al. 2007), such as $f(\mathbf{h}_i, \mathbf{h}_j) = a_1 |\mathbf{h}_i \cap \mathbf{h}_j| - a_2 |\mathbf{h}_i \cap \mathbf{h}_j|^2$, with constants $a_1, a_2 \geq 0$.

Proposition E.4. *Assume that each firm i is assigned a technology $k \in \{1, \ldots, N\}$ uniformly at random and let $f(\mathbf{h}_i, \mathbf{h}_j) = \mathbb{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle \geq 1\}}$.*

(i) The degree distribution is given by

$$P(k) = \binom{n}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{n-k}.$$

(ii) The average nearest neighbor degree distribution is given by

$$k_{nn}(k) = \frac{k(1 - \frac{1}{N})(1 + n\frac{1}{N} - (n+1)\left(\frac{1}{N}\right)^n)}{\frac{1}{N}(1 + n - k)}, \tag{E.10}$$

and for large n the average nearest neighbor degree distribution, $k_{nn}(k)$, grows linearly as $O(k)$.

(iii) The clustering coefficient is given by $C(k) = 1$.

Proof of Proposition E.4. We first prove part (i) of the proposition. If technologies are assigned uniformly at random then

$$\mathbb{P}\left(\langle \mathbf{h}_i, \mathbf{h}_j \rangle \geq 1 \middle| q_i = q, q_j = q'\right) = \frac{1}{N} \mathbb{1}_{\{\rho q q' > \zeta\}}.$$

Due to symmetry, the firms' quantities in the stationary state when $\vartheta \to \infty$ are identical and given by q^*. In the case of $\rho(q^*)^2 > \zeta > 0$ we then we have that

$$\mathbb{P}(a_{ij} = 1) = \frac{1}{N},$$

and the degree distribution is given by

$$P(k) = \mathbb{P}(d_1(G) = k) = \binom{n}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{n-k}.$$

We next give proof of part (ii) of the proposition. The average nearest neighbor degree distribution is then given by

$$k_{\mathrm{nn}}(k) = \sum_{k'=1}^{n} k' \mathbb{P}\left(d_2(G) = k' - 1 \middle| a_{12} = 1, d_1(G) = k\right)$$

where

$$\begin{aligned}
\mathbb{P}\left(d_2(G) = k' - 1 \middle| a_{12} = 1, d_1(G) = k\right) &= \frac{\mathbb{P}\left(d_2(G) = k' - 1, d_1(G) = k \middle| a_{12} = 1\right)}{P(k)} \\
&= \frac{1}{P(k)} \left(\frac{1}{N}\right)^{k'-1} \left(1 - \frac{1}{N}\right)^{n-k'+1} \\
&\times \left(\frac{1}{N}\right)^{k-1} \left(1 - \frac{1}{N}\right)^{n-k+1} \\
&= \frac{P(k'-1)P(k-1)}{P(k)}.
\end{aligned}$$

We then get

$$\begin{aligned}
k_{\mathrm{nn}}(k) &= \sum_{k'=1}^{n} k' \mathbb{P}\left(d_2(G) = k' - 1 \middle| a_{12} = 1, d_1(G) = k\right) \\
&= \sum_{k'=1}^{n} k' \frac{P(k'-1)P(k-1)}{P(k)} \\
&= \frac{k(1 - \frac{1}{N})(1 + n\frac{1}{N} - (n+1)\left(\frac{1}{N}\right)^n)}{\frac{1}{N}(1 + n - k)} \\
&= O(k),
\end{aligned}$$

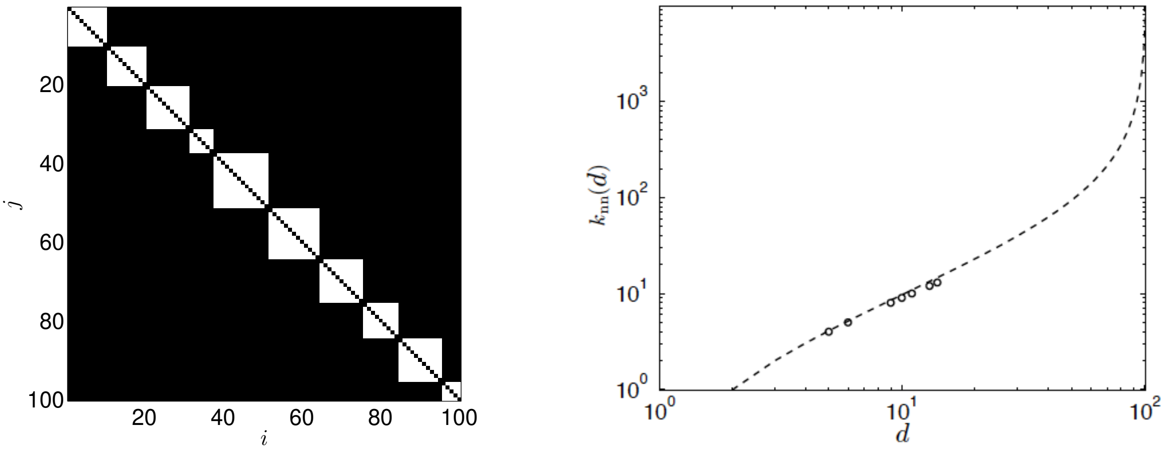


FIGURE E.3. (Left panel) Illustration of the matrix with elements $\mathbb{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle > 0\}}$ (with ones indicated with white squares and zeros indicated with black squares) for $n = 100$ firms and $N = 10$ technologies. (Right panel) The average nearest neighbor degree distribution, $k_{\mathrm{nn}}(d)$, for the same parameters. The dashed line represents the solution from Equation (E.10) while the circles correspond to a numerical simulation.

as $n \to \infty$. That is, the average nearest neighbor degree $k_{\mathrm{nn}}(k)$ is asymptotically linearly increasing with the degree k, and thus we have an assortative network.

Finally, we give proof of part (iii) of the proposition. The clustering coefficient is simply given by $C(k) = \mathbb{P}\left(a_{23} = 1 \middle| a_{12} = 1, a_{23} = 1, d_1(G) = k\right) = 1$. This is because if firm 1 is connected to firm 2 then they must have the same technology. Similarly, if firm 1 is connected to firm 3 then they also must have the same technology. Due to transitivity, firms 2 and 3 then must have the same technology and thus must be connected.

□

An illustration of the average nearest neighbor degree $k_{\mathrm{nn}}(k)$ can be seen in Figure E.3. Different to Figure E.1 this specification gives rise to positive degree correlations between a node and its neighbors in the network, a feature that is characteristic for many real-world networks (cf. König 2016).

Appendix F: Data

In the following we provide a detailed description of the data used for our empirical analysis in Section 3.

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 (Schilling 2009). The first is the Cooperative Agreements and Technology Indicators (CATI) database (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 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 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). We could match 21% of the firms appearing in both databases. Considering only firms without missing observations on R&D expenditures and industry classifications (see also Appendix F.2 below on how we obtained balance sheet and R&D expenditures

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

Sector	2-dig SIC	# firms	% of tot.	Rank
Chemical and Allied Products	28	524	30.15	1
Electronic and Other Electric Equipment	36	283	16.28	2
Instruments and Related Products	38	195	11.22	3
Business Services	73	193	11.10	4
Industrial Machinery and Equipment	35	172	9.90	5
Transportation Equipment	37	81	4.66	6
Engineering and Management Services	87	46	2.65	7
Primary Metal Industries	33	34	1.96	8
Food and Kindred Products	20	29	1.67	9
Communications	48	24	1.38	10
Electric Gas and Sanitary Services	49	17	0.98	11
Health Services	80	16	0.92	12
Miscellaneous Manufacturing Industries	39	12	0.69	13
Nonclassifiable Establishments	99	11	0.63	14
Fabricated Metal Products	34	11	0.63	15
Petroleum and Coal Products	29	11	0.63	16
Rubber and Miscellaneous Plastics Products	30	10	0.58	17
Heavy Construction other than Building	16	10	0.58	18
Paper and Allied Products	26	9	0.52	19
Textile Mill Products	22	9	0.52	20

Table F.2. The 20 largest sectors at the 3-digit SIC level.

Sector	3-dig SIC	# firms	% of tot.	Rank
Drugs	283	410	23.59	1
Computer and Data Processing Services	737	184	10.59	2
Electronic Components and Accessories	367	144	8.29	3
Medical Instruments and Supplies	384	97	5.58	4
Computer and Office Equipment	357	70	4.03	5
Measuring and Controlling Devices	382	69	3.97	6
Motor Vehicles and Equipment	371	64	3.68	7
Communications Equipment	366	56	3.22	8
Special Industry Machinery	355	38	2.19	9
Research and Testing Services	873	35	2.01	10
Chemicals & Allied Products	280	26	1.50	11
Misc. Electrical Equipment and Supplies	369	25	1.44	12
Plastics Materials and Synthetic	282	25	1.44	13
General Industrial Machinery	356	23	1.32	14
Electrical Industrial Apparatus	362	16	0.92	15
Blast Furnace and Basic Steel Products	331	15	0.86	16
Aircraft and Parts	372	14	0.81	17
Metalworking Machinery	354	14	0.81	18
Agricultural Chemicals	287	14	0.81	19
Medical and Dental Laboratories	807	13	0.75	20

information), it gives us a sample of $1,738$ firms and a total of 632 collaborations in the year 2006. The average degree of the firms in this sample is 0.73 with a standard deviation of 2.06, and the maximum degree is 25 attained by *Pfizer Inc.*

Tables F.1 and F.2 show, respectively, 20 largest sectors at the 2-digit and 3-digit SIC levels. The largest sector at the SIC-28 level is chemical and allied products, with 524 firms (30.15 % of the total), followed by the sector of electronic and other electric equipment, with 283 firms (16.28 % of the total). At the 3-digit SIC level the largest sector is the drugs development sector, with 410 firms (23.59 % of the total), and the second largest sector is computer and data processing services with 184 firms (10.59% of the total).

Figure F.1 shows the degree distribution, $P(d)$, the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, the clustering degree distribution, $C(d)$,[16] and the component size distribution, $P(s)$, across different levels of sectoral aggregation, considering all firms in all sectors, firms in the SIC-28 sector only, or firms in the SIC-283 sector only. The degree distribution, $P(d)$, decays as a power law across all datasets considered. The clustering degree distribution, $C(d)$, is also decreasing with increasing degrees d across all datasets. These networks tend to be moderately clustered. The average clustering coefficient considering all firms is $C = 0.074$, for the firms in the SIC-28 sector it is $c = 0.043$ and for the firms in the SIC-283 sector it is $C = 0.038967$. Further, the component size distribution, $P(s)$, indicates a large connected component (see also Figure 2) with smaller components decaying as a power law. This pattern is also consistent across datasets. The largest connected component comprises 21.20% of all firms across sectors, $24,07\%$ of all firms in the SIC-28 sector, and 29.91% of all firms in the SIC-283 sector. While the level or sectoral aggregation does not matter much for the degree distribution, the clustering degree distribution and the component size distribution, a different pattern can be observed for the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$. While the average nearest neighbor connectivity $k_{\mathrm{nn}}(d)$ is decreasing with the increasing degree for the firms restricted to the SIC-28 or SIC-283 sectors, this monotonicity behavior is less pronounced when considering all firms across sectors. This pattern can also be observed in the assortativity coefficient, which is $\gamma = -0.031399$ for all firms, $\gamma = -0.25322$ restricting the sample to firms in the SIC-28 sector, and $\gamma = -0.27464$ for the firms in the SIC-283 sector. That is, while the network is weakly disassortative considering all firms, it becomes strongly disassortative when considering only a single sector. This observation is even more extreme when we consider all firms in the collaboration network without dropping those for which R&D expenditures are missing. In this case we find $\gamma = 0.03343$ for all firms, $\gamma = -0.11703$ restricting the sample to firms in the SIC-28 sector, and $\gamma = -0.14886$ for the firms in the SIC-283 sector. The explanation for this observation can be easily given when considering the extension of our model introduced in Section 2.6 and Appendix E.2. There we showed that when the spillovers from collaborations depend on the technological characteristics of the firms involved in a collaboration, and firms from different sectors have different characteristics, then the emerging network of cross-industry collaborations can be assortative, while the network of intra-industry collaborations is disassortative.

F.1. 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 (Hanaki et al. 2010).

To get a comprehensive picture of the M&A activities of the firms in our dataset, we use two extensive data sources 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 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 aforementioned name-matching algorithm; see also 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.

16. See Supplementary Appendix E for a discussion of these network statistics.

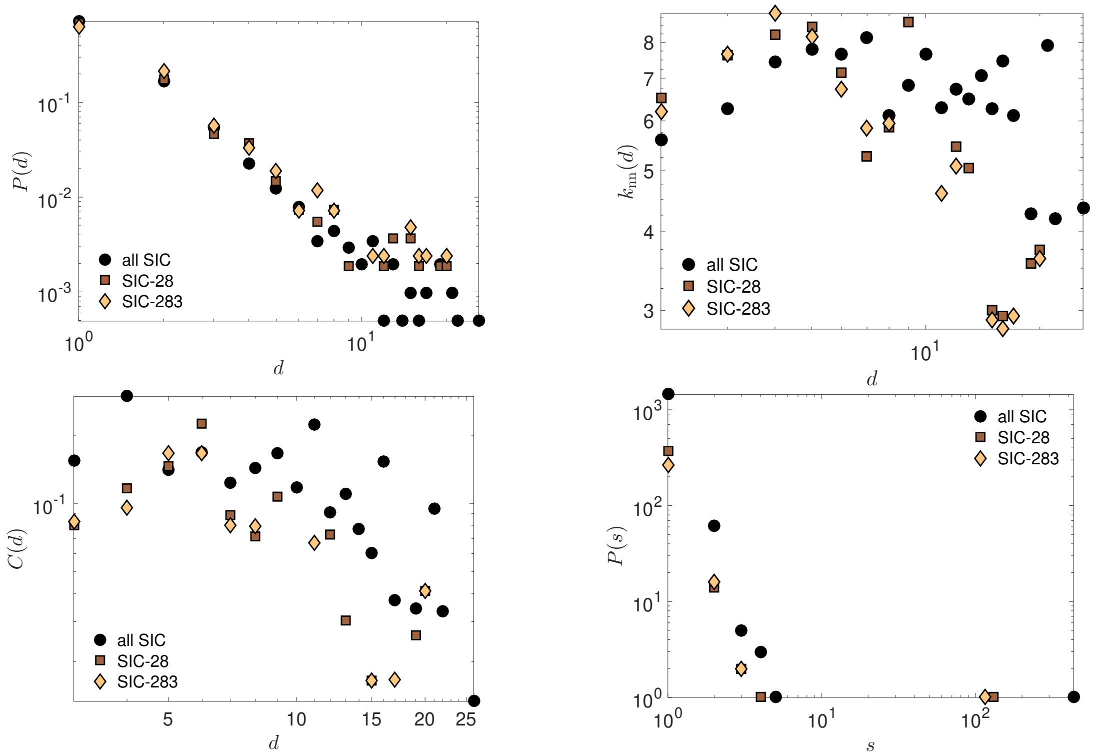


FIGURE F.1. The degree distribution, $P(d)$, the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, the clustering degree distribution, $C(d)$, and the component size distribution, $P(s)$.

F.2. Balance Sheet Statements, R&D and Productivity

The combined CATI-SDC alliance database provides the names of 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. and Global fundamentals annual databases and Bureau van Dijk (BvD)'s Orbis database, to obtain information about their balance sheets and income statements.

Compustat North America is a database of U.S. and Canadian active and inactive publicly held companies extracted from company filings. It provides more than 300 annual and 100 quarterly income statements, balance sheets, and statements of cash flows. Compustat Global is a database of non-U.S. and non-Canadian companies and contains market information on more than 33,900 active and inactive publicly held companies with annual data history from 1987. The Compustat databases cover 99% of the world's total market capitalization with annual company data history available back to 1950. The databases contain only firms listed on the stock market, so it typically excludes 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 it thus covers most of the R&D activities in the economy (Bloom et al. 2013).

The Orbis database is owned by Bureau van Dijk (BvD). It is a commercial dataset, which contains administrative data on 130 million firms worldwide. Orbis is an umbrella product that provides firm level data covering over 120 countries, both developed and emerging, since 2005. The financial and balance-sheet information in Orbis comes from business registers collected by the local Chambers of Commerce to fulfill legal and administrative requirements and are relayed to BvD via over 40 different information providers. Different from Compustat, Orbis contains not only information about publicly listed firms, but provides also information about private firms.[17]

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). We could match 25.53% of the firms in the combined CATI-SDC database with the combined Compustat-Orbis database. For

17. For a detailed comparison and further discussion of the Compustat and Orbis databases see Dai (2012), Bloom et al. (2013) and Papadopoulos (2012).

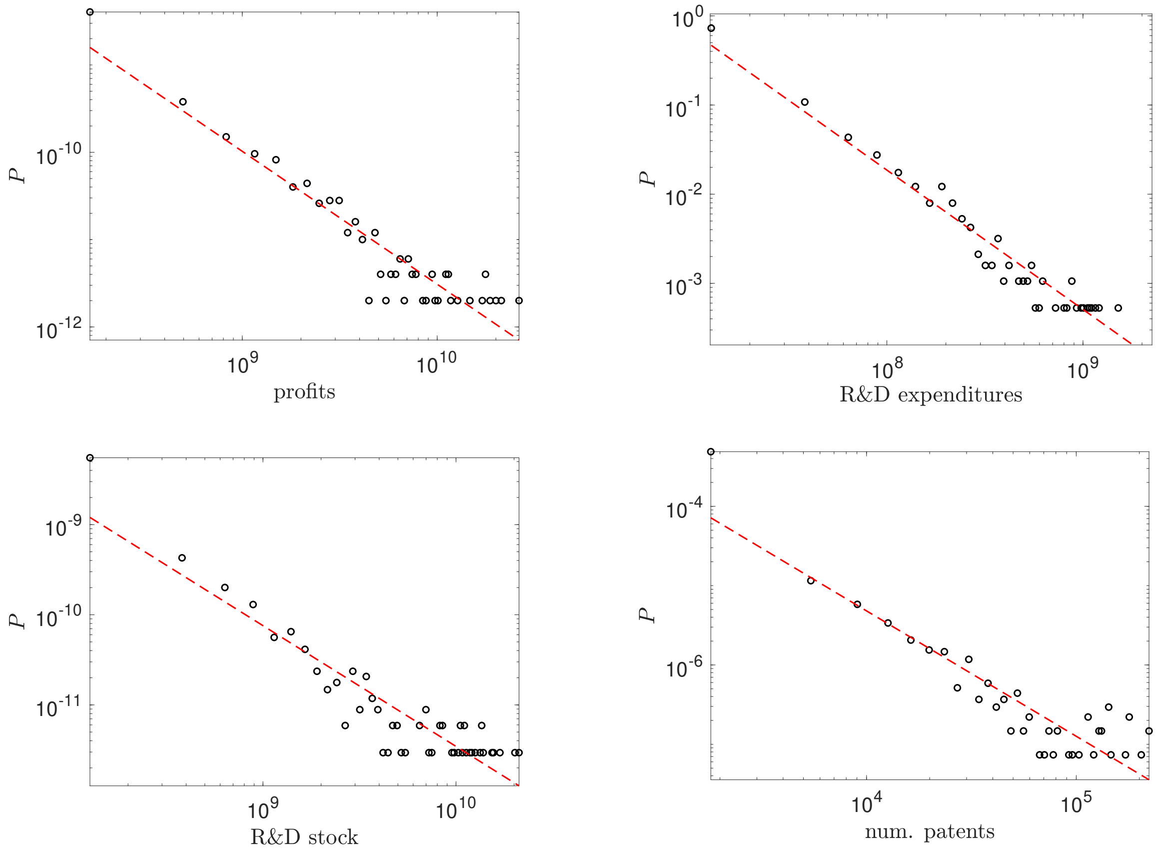


Figure F.2. The profits distribution (top left panel), the R&D expenditures distribution (top right panel), the R&D stocks distribution (bottom left panel), and the patent stock distribution (bottom right panel). Dashed lines indicate a power law fit.

the matched firms we obtained their sales, R&D expenditures, sales, employment, primary industry codes, and location. U.S. dollar translation rates for foreign currencies have been taken directly from the Compustat yearly averaged exchange rates. We adjusted for inflation using the consumer price index of the Bureau of Labor Statistics (BLS), averaged annually, with 1983 as the base year. From a firm's sales and employment, we then computed its labor productivity as sales relative to the number of employees. We then dropped all firms with missing information on R&D expenditures and industry codes. This pruning procedure left us with a subsample of $1,738$ firms, on which the empirical analysis in Section 3 is based.

The empirical profits distribution, the R&D expenditures distribution, the R&D stocks distribution and the patent stock distribution are shown in Figure F.2. All distributions are highly skewed, indicating a large degree of inequality (indicated with a power law fit) in firms' sizes, R&D expenditures, productivity (R&D stocks), and patent stocks. Moreover, Figure F.3 shows a correlation scatter plot for sales, productivity, R&D expenditures, and patent stocks. All are highly correlated, with a Spearman correlation coefficient between sales and R&D expenditures of $\varrho = 0.69$ between sales and productivity of $\varrho = 0.54$, and sales and the number of patents of $\varrho = 0.53$. The correlation between R&D expenditure and productivity is $\varrho = 0.29$ and R&D expenditures and the number of patents is $\varrho = 0.56$. Finally, the correlation between productivity and the number of patents is $\varrho = 0.22$.

F.3. Geographic Location and Distance

Table F.3 shows the 25 countries with the largest numbers of firms. The dominant role of the U.S. with 854 firms making up 49.14% of the total number of firms is clearly visible. The second largest country is Japan with 338 firms. The U.S. and Japan then together account for 68.59%, that is, more than two-thirds of all firms in the data.

In order to determine the precise locations of the firms in our data we have further added the longitude and latitude coordinates associated with the city of residence of each firm. Among the matched cities in our dataset 93.67% could be geo-localized using ArcGIS (see e.g., Dell 2009) and

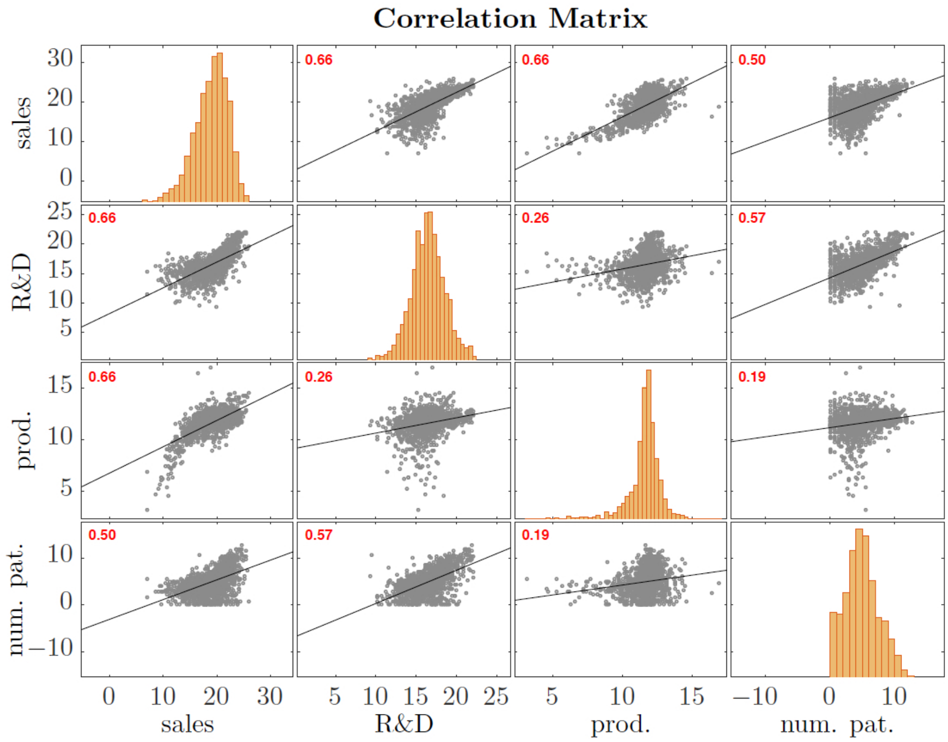


FIGURE F.3. Correlation scatter plot for sales, productivity, R&D expenditures and the patent stocks.

TABLE F.3. The 25 countries with the largest numbers of firms.

Name	Code	# firms	% of tot.	Rank
United States	USA	854	49.14	1
Japan	JPN	338	19.45	2
United Kingdom	GBR	93	5.35	3
Canada	CAN	62	3.57	4
Australia	AUS	50	2.88	5
Germany	DEU	48	2.76	6
Taiwan	TWN	45	2.59	7
France	FRA	35	2.01	8
Switzerland	CHE	34	1.96	9
Sweden	SWE	31	1.78	10
India	IND	21	1.21	11
Finland	FIN	16	0.92	12
Netherlands	NLD	14	0.81	13
Iceland	ISL	13	0.75	14
Slovakia	SVK	13	0.75	15
Denmark	DNK	12	0.69	16
Belgium	BEL	11	0.63	17
Italy	ITA	11	0.63	18
Israel	ISR	10	0.58	19
Morocco	MAR	9	0.52	20
Norway	NOR	8	0.46	21
China	CHN	7	0.40	22
Singapore	SGP	7	0.40	23
Spain	ESP	4	0.23	24
Hong Kong	HKG	4	0.23	25

the Google Maps Geocoding API.[18] We then used Vincenty's algorithm to compute the distances between pairs of geo-localized firms (Vincenty 1975). The mean distance between collaborating firms is $5,227$ km. The distance distribution, $P(d)$, across collaborating firms is shown in Figure F.5, while

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

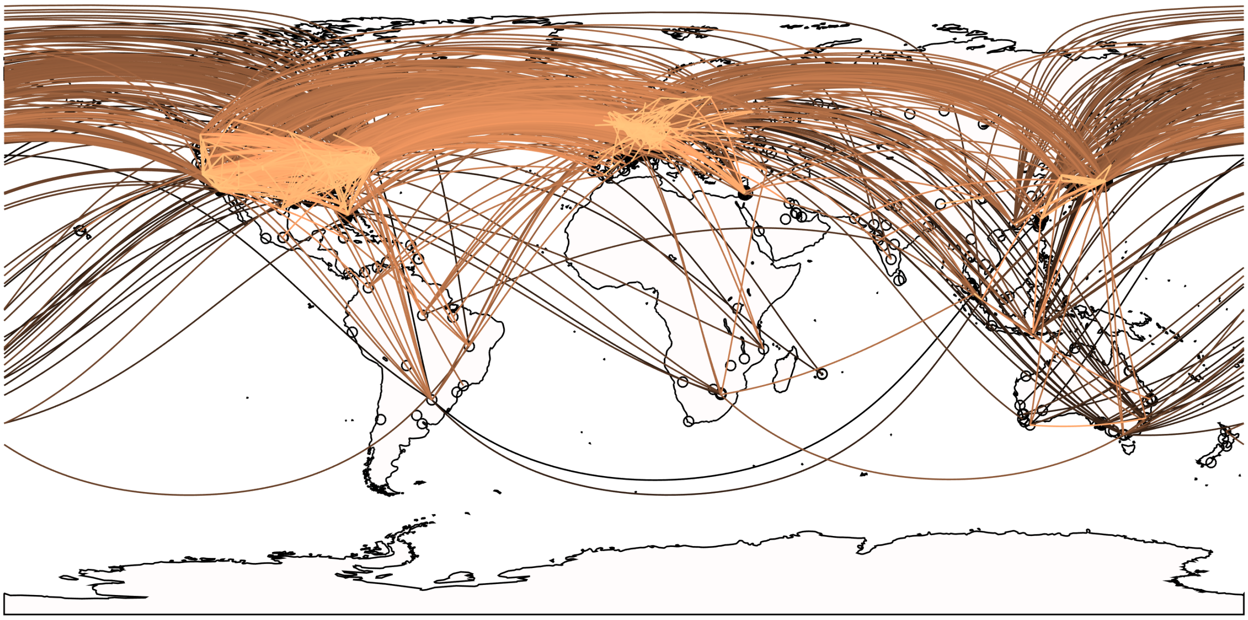


FIGURE F.4. The locations (at the city level) and collaborations of the firms in the combined CATI-SDC database. Link colors scale with the collaboration distance.

Figure F.4 shows the locations (at the city level) and collaborations of the firms in the database. The distance distribution, $P(d)$, 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 (Lychagin et al. 2016).

F.4. Technological Similarity

The technological proximity $f_{ij}^k = f_{ji}^k$, $k \in \{J, M\}$ (see Equation (12)), between firms i and j is measured with two alternative metrics. The first, f_{ij}^J, is based on Jaffe (1989). Let $\mathbf{P}_i$ represents the patent portfolio of firm i, 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 (see also Bloom et al. 2013). The technological proximity of firm i and j is then given by

$$f_{ij}^J = \frac{\mathbf{P}_i^\top \mathbf{P}_j}{\sqrt{\mathbf{P}_i^\top \mathbf{P}_i}\sqrt{\mathbf{P}_j^\top \mathbf{P}_j}}. \tag{F.1}$$

We denote $\mathbf{F}^J$ the $(n \times n)$ matrix with elements $(f_{ij}^J)_{1 \le i,j \le n}$.

As an alternative measure for technological similarity we also consider the Mahalanobis technology proximity measure, f_{ij}^M, introduced by Bloom et al. (2013). To construct this metric, 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 $T_{ji} = P_{ji} / \sum_{k=1}^n P_{ki}$, for all $1 \le i \le n$ and $1 \le j \le N$. Further, we construct the $(N \times n)$ normalized patent shares matrix $\tilde{\mathbf{T}}$ with elements $\tilde{T}_{ji} = T_{ji} / \sqrt{\sum_{k=1}^N T_{ki}^2}$, and the $(n \times N)$ normalized patent shares matrix across firms is defined by $\tilde{\mathbf{X}}$ with elements $\tilde{X}_{ik} = T_{ki} / \sqrt{\sum_{i=1}^N T_{ki}^2}$. Let $\mathbf{\Omega} = \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$. Then the $(n \times n)$ Mahalanobis technology similarity matrix $\mathbf{F}^M = (f_{ij}^M)_{1 \le i,j \le n}$ is defined as

$$\mathbf{F}^M = \tilde{\mathbf{T}}^\top \mathbf{\Omega} \tilde{\mathbf{T}}. \tag{F.2}$$

We then use either f_{ij}^J or f_{ij}^M as a measure for the potential technology spillovers between collaborating firms in the profit function of Equation (12). Both measures are highly correlated. The Spearman correlation coefficient between the Jaffe and the Mahalanobis proximity metrics is 0.92, and a correlation plot can be seen in the right panel of Figure F.5.

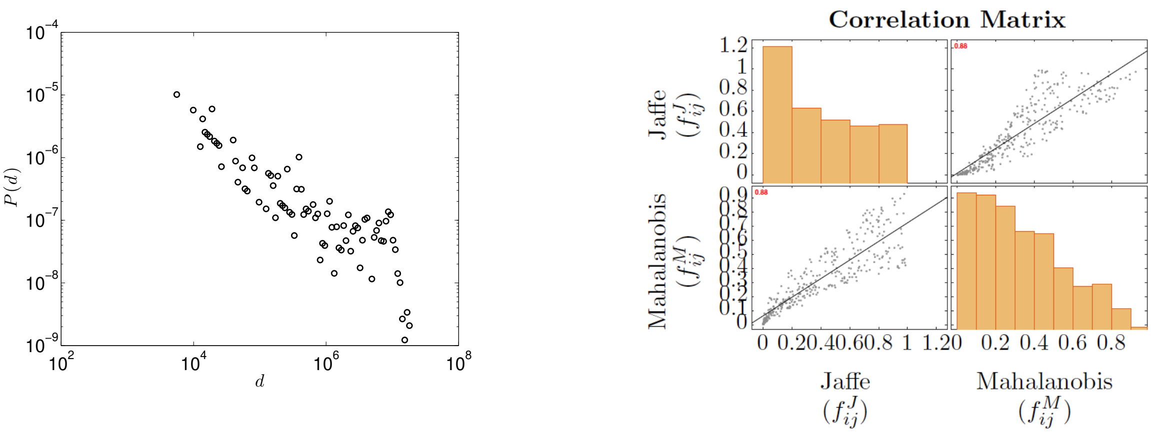


FIGURE F.5. (Left panel) The distance distribution, $P(d)$, across collaborating firms in the combined CATI-SDC database. (Right panel) Correlation plot for the Jaffe (f_{ij}^J) and the Mahalanobis (f_{ij}^M) technology proximity metrics across pairs of firms $1 \leq i, j \leq n$.

F.5. Nestedness and Modularity

We can identify clusters – or so-called modules – of densely connected firms in the network (with only sparser connections across clusters) using the *modularity* algorithm proposed by Newman (2006). This algorithm seeks a partition of the nodes in the network into non-overlapping clusters. Such clusters might emerge because of similarities of firms in terms of the industry in which they are operating, technological similarity, or geographic proximity (lowering collaboration costs and thus making certain pairs of firms more likely to be connected; see Section 3.4).

After having identified these modules, we can compute the nestedness of each module similar to Figure 3. The top left panel in Figure F.6 shows the adjacency matrix of the R&D network where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices of each module (sub-network) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Anderson et al. 2008).[19] The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n). We observe that all modules are highly nested and that 7 out of the 10 modules considered are also significantly nested. The same information but for the largest connected component (see Figure 2) is shown in Figure F.7, where a similarly high level of nestedness can be observed.

Appendix G: Bayesian MCMC Estimation Procedure

In this section we provide the details for the implementation of the Bayesian MCMC estimation algorithm. We divide the parameter vector $\boldsymbol{\theta}$ and unobserved latent variables z into blocks and

19. Statistical significance is computed via comparison to a null model generated from 100 random networks with the same link density as the original network.

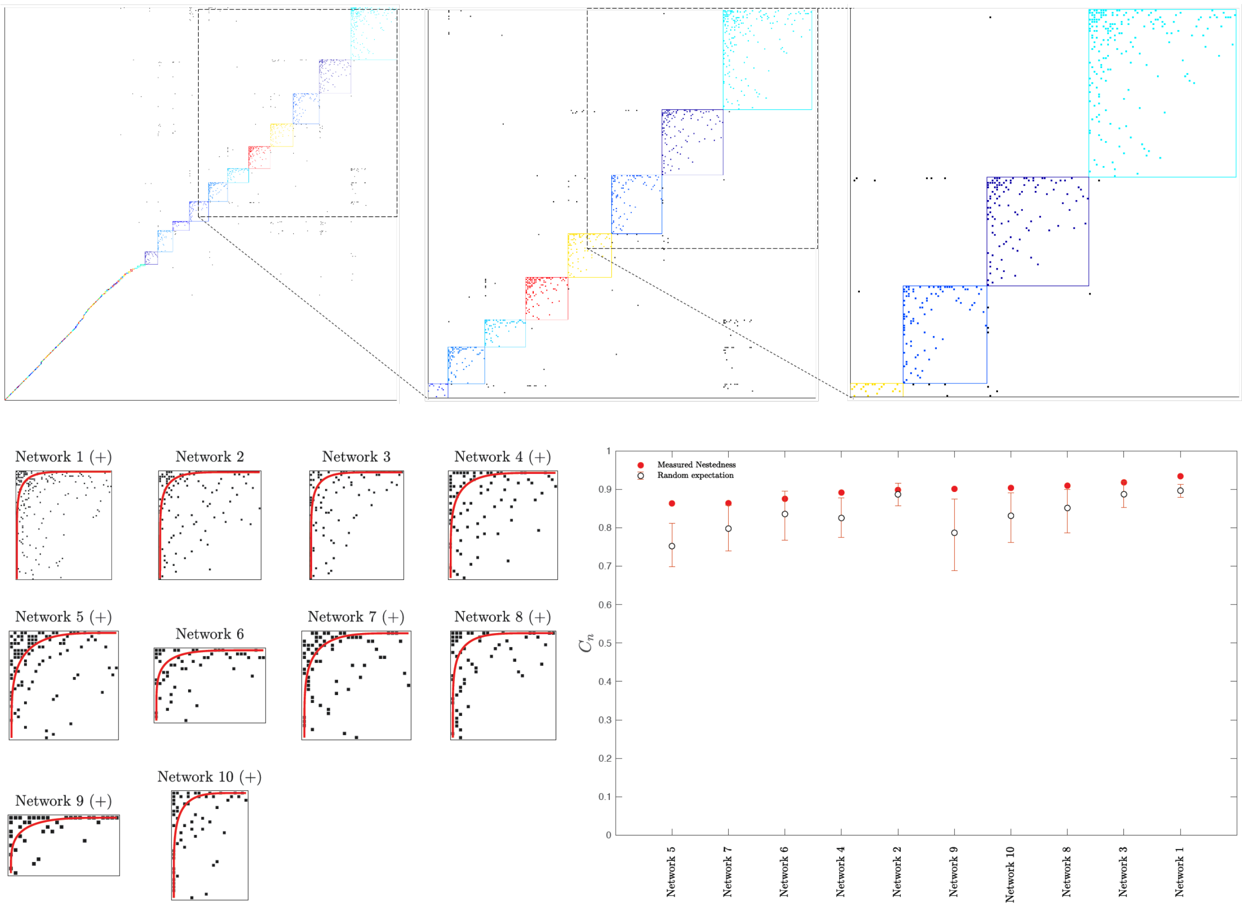


FIGURE F.6. The top left panel shows the adjacency matrix of the R&D network where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices (with ones indicated with black squares and zeros indicated with white squares) of each module (subnetwork) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Anderson et al. 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n). Figures have been created using the `BiMat` Matlab library (https://bimat.github.io/).

assign the prior distributions as follows:

$$
\begin{aligned}
z_i &\sim \mathcal{N}(0, \sigma_z^2), \quad \text{for } i = 1, \ldots, n, \\
\rho &\sim \mathcal{N}(0, \sigma_\rho^2), \\
\lambda &\sim \mathcal{N}(0, \sigma_\lambda^2), \\
\boldsymbol{\delta} &\sim \mathcal{N}(0, \Sigma_{\boldsymbol{\delta}}), \\
\kappa &\sim \mathcal{N}(0, \sigma_\kappa^2), \\
\boldsymbol{\gamma} &\sim \mathcal{N}(0, \Sigma_{\boldsymbol{\gamma}}), \\
\vartheta &\sim \mathcal{N}(0, \sigma_\vartheta^2), \\
\sigma_z^2 &\sim \mathcal{IG}\left(\frac{\varpi_1}{2}, \frac{\varpi_2}{2}\right).
\end{aligned}
$$

We consider the normal ($\mathcal{N}$) and inverse gamma ($\mathcal{IG}$) conjugate priors, which are widely used in the Bayesian literature (Koop et al. 2007). The hyperparameters are chosen to make the prior distributions relatively flat and cover a wide range of the parameter space, i.e., we set $\sigma_\rho^2 = \sigma_\lambda^2 = 10$, $\Sigma_{\boldsymbol{\delta}} = 10 I_{130}$, $\Sigma_{\boldsymbol{\gamma}} = 100 I_4$, $\sigma_\vartheta^2 = 10$, $\varpi_1 = 2.2$, and $\varpi_2 = 0.1$, where I_n denotes the identity matrix of dimension n.

Given the priors specified above and the composite likelihood discussed in Section 3.3, we can derive the joint posterior distribution for the parameter $\boldsymbol{\theta}$. The MCMC sampling procedure combines Gibbs sampling and the Metropolis-Hastings (M-H) algorithm. It consists of the following steps:

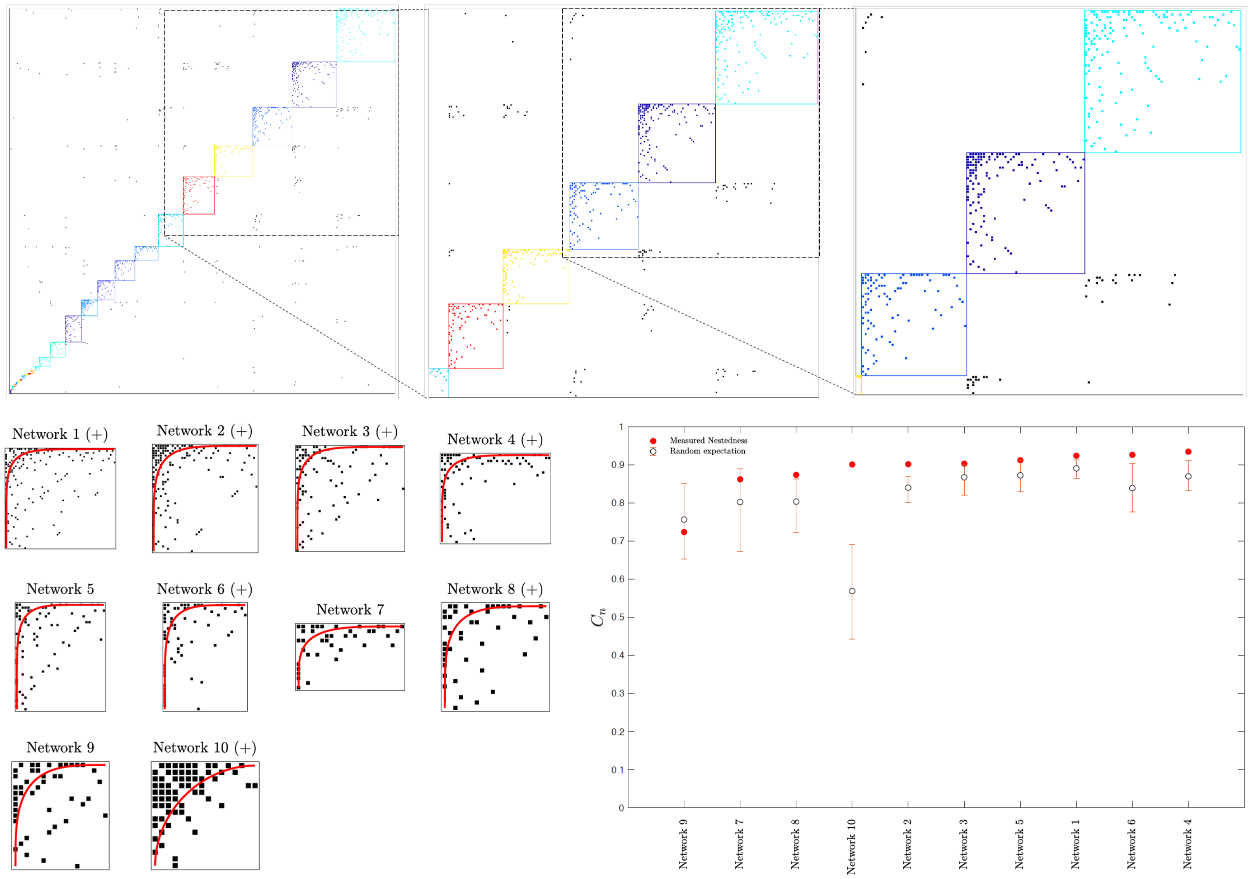


FIGURE F.7. The top left panel shows the adjacency matrix of the largest connected component of the R&D network (see Fig. 2) where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices (with ones indicated with black squares and zeros indicated with white squares) of each module (subnetwork) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Anderson et al. 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n). Figures have been created using the `BiMat` Matlab library (https://bimat.github.io/).

1. Draw the latent variable z_i using the M-H algorithm based on $f(z_i|\mathbf{y}, G, \boldsymbol{\theta}, z_{-i})$, for $i = 1, \ldots, n$.
2. Draw ρ using the M-H algorithm based on $f(\rho|\mathbf{y}, G, \theta\backslash\rho, z)$.
3. Draw λ using the M-H algorithm based on $f(\lambda|\mathbf{y}, G, \theta\backslash\lambda, z)$.
4. Draw δ using the M-H algorithm based on $f(\boldsymbol{\delta}|\mathbf{y}, G, \theta\backslash\boldsymbol{\delta}, z)$.
5. Draw κ using the M-H algorithm based on $f(\kappa|\mathbf{y}, G, \theta\backslash\kappa, z)$.
6. Draw $\boldsymbol{\gamma}$ using the M-H algorithm based on $f(\boldsymbol{\gamma}|\mathbf{y}, G, \theta\backslash\boldsymbol{\gamma}, z)$.
7. Draw ϑ using the M-H algorithm based on $f(\vartheta|\mathbf{y}, G, \theta\backslash\vartheta, z)$.
8. Draw σ_z^2 from the conjugate inverse gamma distribution, $\mathcal{IG}\left(\frac{\varpi_1+n}{2}, \frac{\varpi_2+\sum_{i=1}^n z_i}{2}\right)$.

Appendix H: Monte Carlo Simulation

We estimate the empirical model using a Bayesian MCMC algorithm based on the composite likelihood function discussed in Section 3.3. In order to demonstrate the performance of this estimation approach, we conduct a Monte Carlo simulation study to show that the true parameter values can be obtained with this method even for small samples.

H.1. Absence of Latent Variables

We first consider the case ignoring unobserved heterogeneity where we do not include the latent variables. We generate an artificial R&D network (G) and effort data ($\mathbf{y}$) from the data generating

Table H.1. MCMC simulation results.

Parameter	True Value	Model (A) $\mu^{\vartheta}(G\|\mathbf{y})\mu^{\vartheta}(\mathbf{y}\|G)$		Model (B) $\mu^{\vartheta}(G\|\mathbf{y})$		Model (C) $\mu^{\vartheta}(\mathbf{y}\|G)$	
Profits							
ρ	0.0400	0.0403	(0.0027)	0.0398	(0.0041)	0.0392	(0.0049)
λ	0.0200	0.0201	(0.0010)			0.0202	(0.0011)
δ	1.0000	0.9983	(0.0192)			1.0040	(0.0271)
Linking Cost							
γ_0	1.0000	1.0385	(0.3093)	0.9947	(0.0567)		
γ_1	2.5000	2.5762	(0.3366)	2.5096	(0.1744)		
Noise							
ϑ	1.5000	1.4802	(0.1509)	$1.5000^{\dagger}$		1.5122	(0.0271)
Ave. Degree				4.6245			
Sample Size				200			

Notes: This simulation study performs 300 repetitions and the values reported in this table are the mean and standard deviation (in parenthesis) of parameter estimates calculated across repetitions. For each repetition, we estimate the parameters using Bayesian MCMC sampling with 50,000 iterations and drop the first 10,000 iterations during a burn-in phase. † We fix ϑ at its true value because it is not identified in Model (B).

process (DGP) based on the stationary distribution of Equation (5) with the potential function $\Phi(\mathbf{y}, G)$ given by Equation (15). We set the network sizes (n) to 200 and the number of Monte Carlo simulations to 300. In the DGP, we capture individual exogenous heterogeneity in the effort process by $\eta_i = \delta x_i$, where the variable x_i is generated from a log-normal distribution, i.e., $\ln(x) \sim \mathcal{N}(1.5, 0.5)$. The coefficient δ is set to one.

The symmetric competition matrix $\mathbf{B} = (b_{ij})_{1\leq i,j\leq n}$ (cf. Equation (12) in Section 3.2) is generated exogenously with each component $b_{ij} \in \{0,1\}$ drawn from a Bernoulli distribution with linking probability 0.25. The linking cost for each potential link, $a_{ij} \in \{0,1\}$, in the R&D collaboration network matrix, $\mathbf{A} = (a_{ij})_{1\leq i,j\leq n}$, is captured by $\zeta_{ij} = \gamma_0 + \gamma_1 w_{ij}$, where $w_{ij} = |x_i - x_j|$. The parameters are set to $\boldsymbol{\gamma} = (\gamma_0, \gamma_1) = (1.0, 2.5)$. The true values of the spillover (complementarity) effect ρ and the substitutability effect λ are set to 0.04 and 0.02 respectively. We set the noise parameter ϑ to 1.5, which is close to our empirical estimate, and set $f_{ij} = 1$ for all i, j. The average network degree of the simulated R&D network is 4.6245.

We implement the Bayesian MCMC estimation with 50,000 iterations and drop the first 10,000 iterations during a burn-in phase. The simulation results are summarized in Table H.1 in which the values reported are the mean and standard deviation (in parenthesis) of parameter estimates calculated across repetitions. Table H.1 shows that our proposed Bayesian MCMC estimation based on the composite likelihood $\mu^{\vartheta}(G|\mathbf{y})\mu^{\vartheta}(\mathbf{y}|G)$ in Model (A) recovers all true parameters. As discussed in Section 3.3, when the latent variables ($\mathbf{z}$) are absent, the spillover effect ρ and the market competition effect λ can be identified and thus be consistently estimated from the conditional likelihood $\mu^{\vartheta}(\mathbf{y}|G)$ alone. The simulation result in Model (C) confirms this argument. For Model (B), we estimate parameters in the linking cost function based on the conditional likelihood $\mu^{\vartheta}(G|\mathbf{y})$. Because Model (B) alone does not identify the noise parameter ϑ, we fix it at its true value, and the simulation result shows that other parameters are correctly estimated. We also note an efficiency gain in estimating ρ using the composite likelihood.

H.2. Presence of Latent Variables

We next consider the case with the latent variables (incorporating unobserved heterogeneity). We follow the DGP described in Section H.1 and add the latent variable into η_i and ζ_{ij}, i.e., $\eta_i = \delta x_i + \kappa z_i$ and $\zeta_{ij} = \gamma_0 + \gamma_1 w_{ij} - z_i - z_j$, where the latent variable z_i is generated from a normal distribution $\mathcal{N}(0,1)$ and the true value of κ is set to one. We further change γ_0 from 1.0 to 2.0. Based on this DGP, the average network degree of the simulated R&D network is 6.758.

The simulation results are shown in Table H.2. From Model (B), we see that this estimation approach can successfully recover the true model parameters when considering the true DGP model (with unobserved heterogeneity). Moreover, similar to the pattern that we observe in the

Table H.2. MCMC simulation results.

Parameter	True Value	Model (A) W/o Latent Variable		Model (B) With Latent Variable	
Profits					
ρ	0.0400	0.0576	(0.0074)	0.0428	(0.0030)
λ	0.0200	0.0143	(0.0022)	0.0195	(0.0012)
δ	1.0000	0.8916	(0.0596)	0.9859	(0.0246)
κ	1.0000			0.9294	(0.1424)
Linking Cost					
γ_0	2.0000	2.0144	(0.3399)	1.9496	(0.2572)
γ_1	2.5000	3.4318	(0.5049)	2.6555	(0.2835)
Noise/Uncertainty					
ϑ	1.5000	0.8916	(0.0596)	1.4383	(0.1427)
σ_z^2	1.0000			1.1399	(0.2444)
Ave. Degree		6.7580			
Sample Size		200			

Notes: This simulation study performs 300 repetitions and the values reported in this table are the mean and standard deviation (in parenthesis) of parameter estimates calculated across repetitions. For each repetition, we estimate the parameters using Bayesian MCMC sampling with 50,000 iterations and drop the first 10,000 iterations during a burn-in phase.

empirical study, the estimated coefficient of complementarity effect ρ will be upward biased when the unobserved latent variables are ignored in Model (A).

Appendix I: Robustness Checks and Additional Goodness-of-Fit Statistics

In the following Section I.1 we provide additional estimation results illustrating the robustness of the estimates reported in Section 3.4 in the main text. Moreover, in Section I.2 we provide an alternative illustration of the goodness-of-fit of our estimated model complementing the discussion in Section 3.5 in the main text.

I.1. Robustness Checks

First, in Table I.1 we use value-added per employee as an alternative measure of a firm's productivity. We note, however, that since many firms have missing information on value-added or employment, when using value-added per employee to measure productivity the sample size reduces by 25% (from 1,738 to 1,291 firms). Nevertheless, even with this smaller sample, we find that the estimation results remain qualitatively unchanged.

Second, in Table I.2 we use different depreciation rates (10% and 20%) to calculate the R&D capital stock. We find the estimation results are qualitatively similar.

Third, in Tables I.3 and I.4 we take into account heterogeneous technology spillovers among collaborating firms. In specifying heterogeneous technology spillovers, R&D collaborations are weighted by two alternative technology proximity measures. More precisely, in Table I.3 we use weights f_{ij} based on the technological proximity measure introduced by Jaffe (1989), and in Table I.4 we use weights f_{ij} based on the Mahalanobis technological proximity metric (see Bloom et al. 2013, and Supplementary Appendix F.4 for more details). We find that the estimated spillover coefficient based on the Jaffe weights ($\widehat{\rho} = 0.0250$ in Table I.3) is larger than the one based on the Mahalanobis weights ($\widehat{\rho} = 0.0125$ in Table I.4), and both are larger than the homogeneous spillover coefficient ($\widehat{\rho} = 0.0099$ in Table 3). This is because the average of the Jaffe weights is lower than the average of the Mahalanobis weights. Furthermore, both weights are less than one (see Supplementary Appendix F.4). Next, observe that $f_{ij} = 1$ in the benchmark specification in Table 3. In contrast, in Table I.3 f_{ij} is based on Jaffe's measure, while in Table I.4 f_{ij} is based on the Mahalanobis measure. Taking these observations together we find that the magnitude of the combined spillover effect, ρf_{ij}, remains comparable under these three different specifications.

Table I.1. Estimation results with value-added per employee as an additional productivity measure.

		Model (A)		Model (B)	
Profits					
R&D Spillover	(ρ)	0.019***	(0.001)	0.019***	(0.001)
Substitutability	(λ)	6.18e-5**	(3.11e-5)	9.21e-5***	(3.34e-5)
Productivity					
R&D Capital Stock	(δ_1)	0.832***	(0.003)	0.808***	(0.009)
Value Added per Employee	(δ_2)			0.025***	(0.009)
Sector Dummies		Yes		Yes	
Linking Cost					
Constant	(γ_0)	6.716***	(0.224)	6.820***	(0.213)
Same Sector	(γ_1)	-1.073***	(0.072)	-1.069***	(0.066)
Same Country	(γ_2)	-0.393***	(0.056)	-0.399***	(0.060)
Diff-in-R&D Capital Stock	(γ_3)	-0.077***	(0.015)	-0.089***	(0.031)
Diff-in-Value Added per Emp.	(γ_4)			-0.071***	(0.014)
Noise					
Noise in Decisions	(ϑ)	1.849***	(0.064)	1.853***	(0.062)
Sample Size	(n)	1,291			

Notes: The dependent variables are log-R&D expenditures and (presence or absence of) R&D collaborations between firms. A firm's productivity is measured by its one-year lagged log R&D capital stock (Model A) or together with the one-year lagged log value-added per employee (Model B). We estimate two models with 50,000 MCMC draws where the first 10,000 draws are dropped during a burn-in phase, and every 10th of the remaining draws are kept to calculate the posterior mean (as point estimates) and posterior standard deviation (shown in parenthesis). The asterisks ***(**,*) indicate that a parameter's 99% (95%, 90%) highest posterior density range does not cover zero (Kruschke 2015).

Table I.2. Estimation results with a firm's productivity measured by log R&D capital stock defined at different depreciation rates (10%, 15% and 20%).

		10%		15%		20%	
Profits							
R&D Spillover	(ρ)	0.019***	(0.001)	0.017***	(0.001)	0.016***	(0.001)
Substitutability	(λ)	3.06e-5***	(1.14e-5)	3.78e-5***	(1.35e-5)	4.22e-5***	(1.17e-5)
Productivity	(δ)	0.829***	(0.002)	0.848***	(0.002)	0.861***	(0.002)
Sector Dummies		Yes		Yes		Yes	
Linking Cost							
Constant	(γ_0)	7.482***	(0.204)	6.843***	(0.180)	6.179***	(0.158)
Same Sector	(γ_1)	-1.315***	(0.064)	-1.194***	(0.055)	-1.066***	(0.050)
Same Country	(γ_2)	-0.424***	(0.051)	-0.379***	(0.048)	-0.339***	(0.042)
Diff-in-Productivity	(γ_3)	-0.099***	(0.012)	-0.090***	(0.011)	-0.082***	(0.009)
Noise							
Noise in Decisions	(ϑ)	1.574***	(0.047)	1.736***	(0.048)	1.946***	(0.055)
Sample Size	(n)	1,738					

Notes: The dependent variables are log-R&D expenditures and (presence or absence of) R&D collaborations between firms. A firm's productivity is measured by its one-year lagged log R&D capital stock defined at different depreciation rates (10%, 15%, and 20%). We estimate the model with 50,000 MCMC draws where the first 10,000 draws are dropped during a burn-in phase, and every 10th of the remaining draws are kept to calculate the posterior mean (as point estimates) and posterior standard deviation (shown in parenthesis). The asterisks ***(**,*) indicate that a parameter's 99% (95%, 90%) highest posterior density range does not cover zero (Kruschke 2015).

TABLE I.3. Estimation results for the case of heterogeneous technology spillovers *à la* Jaffe.

		Model (A) W/o Unobs. Heterogeneity		Model (B) With Unobs. Heterogeneity	
Profits					
R&D Spillover	(ρ)	0.040***	(0.002)	0.025***	(0.002)
Substitutability	(λ)	5.77e-5**	(1.82e-5)	3.68e-5***	(1.38e-5)
Productivity	(δ)	0.861***	(0.002)	0.860***	(0.002)
Unobs. Heterogeneity	(κ)	–		0.075***	(0.006)
Sector Dummies		Yes		Yes	
Linking Cost					
Constant	(γ_0)	6.610***	(0.229)	8.396***	(0.270)
Same Sector	(γ_1)	-0.979***	(0.078)	-1.299***	(0.091)
Same Country	(γ_2)	-0.507***	(0.060)	-0.693***	(0.084)
Diff-in-Productivity	(γ_3)	-0.125***	(0.015)	0.015	(0.014)
Noise/Uncertainty					
Noise in Decisions	(ϑ)	1.375***	(0.045)	1.322***	(0.039)
Unobs. Heterogeneity	(σ_z^2)	–		1.4864***	(0.1515)
Sample Size	(n)	1,738			

Notes: The dependent variables are log-R&D expenditures and (presence or absence of) R&D collaborations between firms. A firm's productivity is measured by its one-year-lagged log-R&D capital stock. To compute the R&D capital stock we use a perpetual inventory method based on the firms' R&D expenditures with a 15% depreciation rate (cf. Hall et al. 2000; Bloom et al. 2013). The parameters are given in the empirical profit function of Equation (12). We make 50,000 MCMC draws where the first 10,000 draws are dropped during a burn-in phase and every 10th of the remaining draws are kept to calculate the posterior mean (as point estimates) and posterior standard deviation (shown in parenthesis). All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery et al. (1992). The asterisks *** (**,*) indicate that a parameter's 99% (95%, 90%) highest posterior density range does not cover zero (Kruschke 2015).

TABLE I.4. Estimation results for the case of heterogeneous technology spillovers *à la* Mahalanobis.

		Model (A) W/o Unobs. Heterogeneity		Model (B) With Unobs. Heterogeneity	
Profits					
R&D Spillover	(ρ)	0.019***	(0.001)	0.013***	(0.001)
Substitutability	(λ)	4.08e-5**	(1.43e-5)	7.05e-5***	(1.38e-5)
Productivity	(δ)	0.860***	(0.002)	0.863***	(0.002)
Unobs. Heterogeneity	(κ)	–		0.105***	(0.015)
Sector Dummies		Yes		Yes	
Linking Cost					
Constant	(γ_0)	6.688***	(0.229)	8.143***	(0.340)
Same Sector	(γ_1)	-1.003***	(0.081)	-1.281***	(0.104)
Same Country	(γ_2)	-0.531***	(0.062)	-0.686***	(0.080)
Diff-in-Productivity	(γ_3)	-0.127***	(0.014)	0.014	(0.014)
Noise/Uncertainty					
Noise in Decisions	(ϑ)	1.360***	(0.044)	1.357***	(0.047)
Unobs. Heterogeneity	(σ_z^2)	–		1.351***	(0.167)
Sample Size	(n)	1,738			

Notes: The dependent variables are log-R&D expenditures and (presence or absence of) R&D collaborations between firms. A firm's productivity is measured by its one-year-lagged log-R&D capital stock. To compute the R&D capital stock we use a perpetual inventory method based on the firms' R&D expenditures with a 15% depreciation rate (cf. Hall et al. 2000; Bloom et al. 2013). The parameters are given in the empirical profit function of Equation (12). We make 50,000 MCMC draws where the first 10,000 draws are dropped during a burn-in phase and every 10th of the remaining draws are kept to calculate the posterior mean (as point estimates) and posterior standard deviation (shown in parenthesis). All cases pass the convergence diagnostics provided by Geweke (1992) and Raftery et al. (1992). The asterisks *** (**,*) indicate that a parameter's 99% (95%, 90%) highest posterior density range does not cover zero (Kruschke 2015).

I.2. Goodness-of-fit

In the following we provide an additional illustration for the goodness-of-fit of our model. In Figure I.1 we compare observed links in the R&D network with the simulated links for the largest 150 firms in our sample. The left panel shows the observed network, and the middle panel shows the simulated links that show up in more than half of the 100 simulated networks. The right panel shows the difference between the simulated and observed networks, with yellow links indicating a match and violet links indicating a mismatch between the left and middle panels. We find that the simulated networks closely resemble the actual network. This illustrates that our model fits the data well also at the level of individual R&D collaborations.

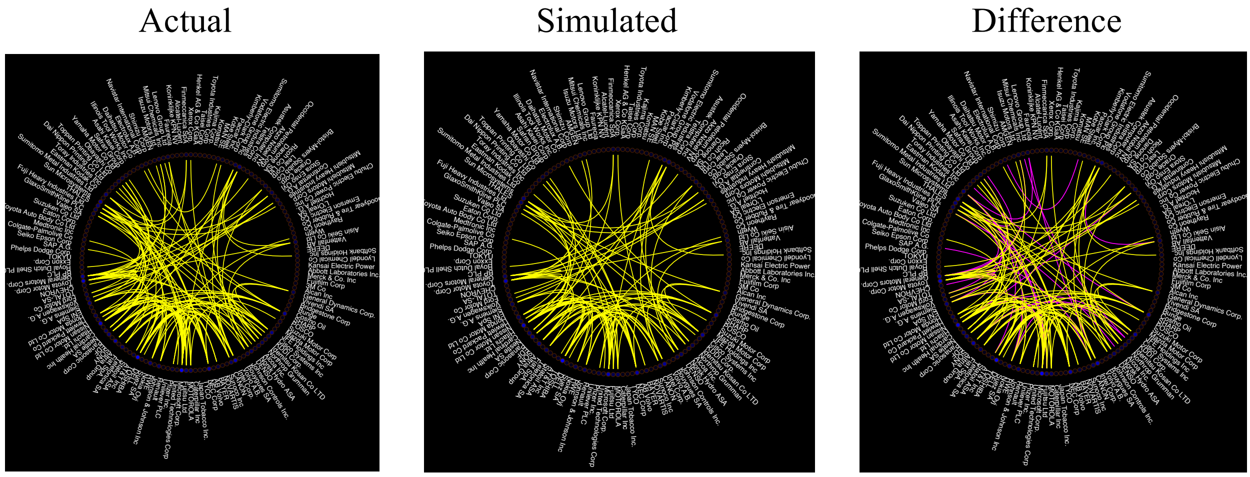


FIGURE I.1. The left panel shows the observed links of the largest 150 firms. Firm names are indicated along the circular plot. The middle panel shows the simulated links that show up in more than half of the 100 simulated networks. The right panel shows the difference between the simulated and observed networks, with yellow links indicating a match and violet links indicating a mismatch between the left and middle panels. Figures are created using the `Schemaball` library (https://github.com/okomarov/schemaball).

References

Almeida-Neto, M., Guimaraes, P., Guimaraes Jr., P. R., Loyola, R. D. and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos*, 117(8):1227–1239.

Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete choice theory of product differentiation*. MIT Press.

Baum, J., Cowan, R., and Jonard, N. (2009). Network-independent partner selection and the evolution of innovation networks. *UNU-MERIT Working Paper Series; 2009-022*.

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

Berliant, M. and Fujita, M. (2008). Knowledge creation as a square dance on the hilbert cube. *International Economic Review*, 49(4):1251–1295.

Berliant, M. and Fujita, M. (2009). Dynamics of knowledge creation and transfer: The two person case. *International Journal of Economic Theory*, 5(2):155–179.

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

Blundell, R., Griffith, R., and Van Reenen, J. (1995). Dynamic count data models of technological innovation. *The Economic Journal*, 105(429):333–344.

Bollaert, H. and Delanghe, M. (2015). Securities Data Company and Zephyr, data sources for M&A research. *Journal of Corporate Finance*.

Bollobás, B., Janson, S., and Riordan, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122.

Brualdi, R. A. and Solheid, Ernie, S. (1986). On the spectral radius of connected graphs. *Publications de l' Institute MathÃ©matique*, 53:45–54.

Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435.

Cohen, W. and Klepper, S. (1996a). Firm size and the nature of innovation within industries: The case of process and product R&D. *Review of Economics and Statistics*, 78(2):232–243.

Cohen, W. and Klepper, S. (1996b). A reprise of size and R&D. *The Economic Journal*, 106(437):925–951.

Cvetkovic, D. and Rowlinson, P. (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.*

Deijfen, M. and Kets, W. (2009). Random intersection graphs with tunable degree distribution and clustering. *Probability in the Engineering and Informational Sciences*, 23(04):661–674.

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

Dorogovtsev, S. N. and Mendes, J. F. (2013). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press.

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1):465–485.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Griffith, R., Redding, S., and Van Reenen, J. (2003). R&D and absorptive capacity: Theory and empirical evidence. *Scandinavian Journal of Economics*, 105(1):99–118.

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

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

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

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

Kamien, M. I., Muller, E., and Zang, I. (1992). Research joint ventures and R&D cartels. *The American Economic Review*, 82(5):1293–1306.

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

König, M. D. (2016). The formation of networks with local spillovers and limited observability. *Theoretical Economics*, 1:813–863.

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

König, M. D., Liu, X., and Zenou, Y. (2019). R&D networks: Theory, empirics and policy implications. *Review of Economics and Statistics*, 101(3):476–491.

König, M. D., Lorenz, J., and Zilibotti, F. (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, 11:1053–1102.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian Econometric Methods*. Cambridge University Press.

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis*, Vol. 476.

Liben-Nowell, D. and Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American society for information science and technology*, 58(7):1019–1031.

Lychagin, S., Pinkse, J., Slade, M. E., and Van Reenen, J. (2016). Spillovers in space: Does geography matter? *The Journal of Industrial Economics*, 64(2):295–335.

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

Mele, A. and Zhu, L. (2022). Approximate variational estimation for a model of network formation. *Review of Economics and Statistics, Forthcoming.*

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Melitz, M., Ottaviano, G., and Maggiore, S. (2008). Market Size, Trade, and Productivity. *Review of Economic Studies*, 75(1):295–316.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics.

Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., and Van Den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research Policy*, 36(7):1016–1034.

Papadopoulos, A. (2012). Sources of data for international business research: Availabilities and implications for researchers. In *Academy of Management Proceedings*, 2012(1):14092.

Peled, U. N., Petreschi, R., and Sterbini, A. (1999). (n, e)-graphs with maximum sum of squares of degrees. *Journal of Graph Theory*, 31(4):283–295.

Powell, W. W., White, D. R., Koput, K. W., and Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110(4):1132–1205.

Raftery, A. E., Lewis, S., Banfield, J. D., and Rafterv, A. E. (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, 4:115–130.

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

Singh, N. and Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics*, 15(4):546–554.

Söderberg, B. (2002). General formalism for inhomogeneous random graphs. *Physical Review E*, 66(6):066121.

Thoma, G., Torrisi, S., Gambardella, A., Guellec, D., Hall, B. H., and Harhoff, D. (2010). Harmonizing and combining large datasets-an application to firm level patent and accounting data. *National Bureau of Economic Research Working Paper No. w15851.*

Trajtenberg, M., Shiff, G., and Melamed, R. (2009). The “names game”: Harnessing inventors, patent data for economic research. *Annals of Economics and Statistics*, (93/94):79–108.

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, 23(176):88–93.

Westbrock, B. (2010). Natural concentration in industrial research collaboration. *The RAND Journal of Economics*, 41(2):351–371.

Wong, R. (2001). *Asymptotic Approximations of Integrals*, Society for Industrial Mathematics.