

# Network Formation with Local Complements and Global Substitutes: The Case of R&D Networks<sup>☆</sup>

Chih-Sheng Hsieh<sup>a</sup>, Michael D. König<sup>b</sup>, Xiaodong Liu<sup>c</sup>

<sup>a</sup>*Department of Economics, Chinese University of Hong Kong, CUHK Shatin, Hong Kong, China.*

<sup>b</sup>*Department of Economics, University of Zurich, Schönberggasse 1, CH-8001 Zurich, Switzerland.*

<sup>c</sup>*Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309-0256, United States.*

---

## Abstract

We introduce a stochastic network formation model where agents choose both actions and links. Neighbors in the network benefit from each other's action levels through local complementarities and there exists a global interaction effect reflecting a strategic substitutability in actions. We provide a complete equilibrium characterization in the form of a Gibbs measure, and show that the model is consistent with empirically observed networks. We then use our equilibrium characterization to show that the model can be conveniently estimated even for large networks. The policy relevance is demonstrated with examples of firm exit, mergers and subsidies in R&D collaboration networks.

*Key words:* network formation, peer effects, technology spillovers, key player, mergers and acquisitions, subsidies

*JEL:* C11, C63, C73, D83, L22

---

---

<sup>☆</sup>We would like to thank Anton Badev, Filomena Garcia, David Hemous, Matt Jackson, Terence Johnson, Alexey Kushnir, Lung-Fei Lee, Elena Manresa, Andrea Montanari, Nick Netzer, Onur Özgür, Agnieszka Rusinowska, María Sáez-Martí, Armin Schmutzler, Martin Summer, Adam Szeidl, Joachim Voth, Timothy Van Zandt, Yves Zenou, Fabrizio Zilibotti, and seminar participants at the Workshop on Networks: Dynamics, Information, Centrality, and Games at Centre d'Economie de la Sorbonne in Paris, the Second Annual NSF Conference on Network Science in Economics at Stanford University, the Econometric Society World Congress in Montreal, the congress of the European Economic Association in Mannheim, the 25th Stony Brook International Conference on Game Theory, the Public Economic Theory Conference in Lisbon, the University of Zurich, Bern, the National Taiwan University, the Austrian National Bank and Stanford University for the helpful comments and advice. We further thank Christian Helmers for data sharing and Sebastian Ottinger for the excellent research assistance. A previous version of this paper has been circulated under the title "Dynamic R&D Networks" as Working Paper No. 109 in the Department of Economics working paper series of the University of Zurich. 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 during 2010-2012.

*Email addresses:* [cshsieh@cuhk.edu.hk](mailto:cshsieh@cuhk.edu.hk) (Chih-Sheng Hsieh), [michael.koenig@econ.uzh.ch](mailto:michael.koenig@econ.uzh.ch) (Michael D. König), [xiaodong.liu@colorado.edu](mailto:xiaodong.liu@colorado.edu) (Xiaodong Liu)

# 1. Introduction

Networks are important in shaping individual behavior and aggregate outcomes in many social and economic applications [Jackson, 2008; Jackson et al., 2017]. A crucial aspect of such environments is the coevolution of networks and behaviors: an agent in a network adjust his behavior based on those of his connections and he choose his connections based on their behaviors. In this paper we introduce a tractable framework to study the joint evolution of networks and behaviors which can be applied to real world networks and used for policy analysis.

We consider a general linear-quadratic interdependent utility function [Ballester et al., 2006], where agents choose action levels and create links at a cost. Neighbors in the network benefit from each other’s action levels through local complementarities, such as those that arise between R&D collaborating firms sharing knowledge about a cost-reducing technology. The global interaction effect reflects a strategic substitutability in actions, for example, through business stealing effects that arise when a firm expands its production in the market. The tractability of the model allows us to provide a complete equilibrium characterization, and an efficiency analysis. Moreover, it is shown that the equilibrium networks generated by the model consistently reproduce features of real world networks. Further, using the equilibrium characterization in the form of a Gibbs measure, we show that the model can be conveniently estimated even for large networks. Finally, the model is amenable to policy analysis, and we illustrate this with examples of firm exits, mergers and acquisitions (M&A) and subsidies in the context of R&D collaboration networks.

**Overview of the results and contributions** Our paper attempts to make three interrelated contributions: a theoretical, an econometric and a policy contribution. Our model has a broad range of applications in various fields [Jackson et al., 2015]. To give a concrete example, in the following we will illustrate our contributions with the example of firms forming R&D collaborations to benefit from technology spillovers while, at the same time, being competitors in the product market [D’Aspremont and Jacquemin, 1988].

First, this paper provides the first fully tractable and estimable model of strategic R&D network formation with endogenous production and R&D collaboration choices, which takes into account the two-way flow of influence from the market structure to the incentives to form R&D collaborations and, in turn, from the formation of collaborations to the market structure.

We study the incentives of firms to form R&D collaborations with other firms and the implications of these alliance decisions for the overall market structure. We introduce a dynamic process in which firms can adjust both, quantities produced (as well as R&D efforts), and the R&D collaborations between them, based on a noisy profit maximization rationale [Blume, 2003; Brock and Durlauf, 2001], taking into account that the establishment of an R&D collaboration is fraught with ambiguity and uncertainty [see e.g., Czarnitzki et al., 2015; Kelly et al., 2002]. Using a potential function we show that the stationary states of this process are completely characterized by a “Gibbs measure” [Bisin et al., 2006; Grimmett, 2010]. Moreover, we show that when firms have heterogeneous marginal costs stemming from differences in their productivity, then the stochastically stable networks (in the limit of vanishing noise) are “nested split graphs” [König et al., 2014a],<sup>1</sup> providing an explanation for why nestedness has been observed in empir-

---

<sup>1</sup>A network is a nested split graph if the neighborhood of every node is contained in the neighborhoods of the nodes with higher degrees [Mahadev and Peled, 1995]. See supplementary Appendix B for further network

ical R&D networks [Tomasello et al., 2016]. Nested split graphs further 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]. In particular, Kitsak et al. [2010] find that firms in the core have a higher market value, consistent with the predictions of our model. Next, we show that when firms’ productivities are Pareto distributed, then the firms’ output levels and degrees also follow a Pareto distribution, which is consistent with the empirical data [Powell et al., 2005]. Further, we find that there exists a sharp transition between sparse and dense networks with decreasing linking costs. We also compute the stationary output levels and show that there exists an intermediate range of the linking cost for which multiple equilibria arise. The equilibrium selection is a path dependent process characterized by “hysteresis” [David, 1992]. Moreover, as in the case of the network density, there exists a sharp transition from a low output to a high output equilibrium.

It is also possible to generalize our model by introducing heterogeneous spillovers from collaborations between firms with differences in their technological characteristics, and/or heterogeneous costs of collaboration. In particular, assuming that firms can only benefit from collaborations if they have at least one technology in common, we show that our model is a generalization of a “random intersection graph” [Deijfen and Kets, 2009],<sup>2</sup> for which positive degree correlations (assortativity) can be obtained. We then investigate the efficient network and output structure that maximize social welfare, and find that equilibrium networks tend to be under-connected, compared to the social optimum.<sup>3</sup>

Second, we bring the model to the data by analyzing a unique dataset of firm R&D collaborations matched to firms’ balance sheets. The theoretical characterization of the stationary states via a Gibbs measure allows us to estimate the model’s parameters using a Markov Chain Monte Carlo (MCMC) method called double Metropolis-Hastings (DMH) algorithm [Hsieh and Lee, 2013; Mele, 2017; Liang, 2010], which can handle the problem of the “intractable normalization constant” in the probability likelihood function by introducing auxiliary network and outcome data.<sup>4</sup> We further use a novel adaptive exchange (AEX) algorithm to overcome the slow mixing problem faced by the DMH algorithm [Jin et al., 2013; Liang et al., 2015]. It applies importance sampling to prevent the “local trap problem” when the likelihood represented by the Gibbs measure is multi-modal. We also propose a likelihood partition approach in which we “integrate out” the intractable normalizing constant by direct analytic computations. The resulting likelihood allows us to use a standard MH algorithm for estimation, and this approach can be efficiently applied to large networks.

From our estimation results we observe that the estimated technology spillover effect is significantly positive and the estimated competition effect is significantly negative, which confirms our theoretical predictions that firms face a positive complementary effect from R&D collaborations and a negative substitution effect from competing firms in the same market. We also investigate

---

definitions and characterizations.

<sup>2</sup>A random intersection graph is constructed by assigning to each node a subset of a given set and two nodes are connected when their respective subsets intersect.

<sup>3</sup>In Section 4.3 we analyze the effectiveness of a subsidy on firms’ R&D collaboration costs, that gives firms incentives to form collaborations and thus increases the network connectivity.

<sup>4</sup>The intractable normalization constant refers to the denominator of the Gibbs likelihood function. The denominator involves a summation over all possible network configurations. If there are  $n$  nodes, then there are  $2^{\binom{n}{s}}$  possible networks to consider.

heterogeneous technology spillovers by weighting the R&D collaboration links with Jaffe and Mahalanobis technology proximity measures [Bloom et al., 2013]. The results show variations on the estimated spillover effects in correspondence with the average magnitudes of different proximity measures.

Third, we use our estimated model to investigate the impact of exogenous shocks on the network. In particular, we perform a “key player” analysis [Ballester et al., 2006; Zenou, 2015], to gauge the impact of firm exit on the economy, when firms are connected through R&D collaborations in a network.<sup>5</sup> The exit of a firm could be due to either financial reasons, such as the recession experienced by the American automobile manufacturing industry during the global financial downturn, or legal reasons, such as the recent emission-fraud scandal of Volkswagen. In the latter case, policy makers want to know the overall cost they impose on the economy by inflicting high pollution penalties and criminal fines that might threaten the continued existence of a firm. Focusing on the chemicals and pharmaceutical sector, our results indicate that the exit of *Amgen*, an American global pharmaceutical corporation, would lead to a reduction in welfare of 3.7%. We then provide a ranking of the firms in our sample according to their impact on welfare upon exit. The ranking shows that the most important firms are not necessarily the ones with the highest market share, but that we need to take into account the positions of these firms in a network of R&D collaborations, and how this network dynamically responds to shocks of firm exits.

Our framework also allows us to study mergers and acquisitions, and their impact on welfare [Farrell and Shapiro, 1990]. Traditional market concentration indices are not adequate to correctly account for the network effect of a merger on welfare [Encaoua and Hollander, 2002]. This is because the effect of a merger on industry profits, consumer surplus and overall welfare depends not only on the market structure, but also on the architecture of the R&D collaboration network between firms through which R&D spillovers are channelled. In such networked markets benefits from concentration of R&D activities can arise through economies of scale and faster diffusion of technologies in more centralized networks [Daughety, 1990].

By taking into account the network effect our results show that a merger between *Daiichi Sankyo Co. Ltd.*, the second largest pharmaceutical company in Japan and *Schering-Plough Corp.*, a U.S.-based multinational pharmaceutical company, would result in a welfare loss of 0.6%. In contrast, a welfare gain of 0.86% from a merger is obtained between *Isis Pharm. Inc.* and *Takeda Pharm. Inc.* Comparing the firms involved in mergers that lead to welfare gains as opposed to the ones that lead to welfare losses, we see that mergers between firms with a larger number of patents and a larger number of R&D collaborations typically lead to welfare gains, while mergers between firms with few collaborations, fewer patents and a larger market share typically lead to a welfare loss. This indicates that the R&D spillover effect is larger when two well connected, R&D intensive firms merge, while welfare losses from increased market concentration dominate when less connected and more market dominant firms are involved in the merger. Our counterfactual policy analysis is therefore potentially important for antitrust policy makers.

Finally, we investigate the impact of subsidizing R&D collaboration costs of selected pairs of

---

<sup>5</sup>We note that our model is formulated in a fairly flexible way, and because we consider the general payoff structure introduced in Ballester et al. [2006], one could use our framework also to investigate key players in criminal networks, or other related contexts [see also Zenou, 2015].

firms. Our study indicates that subsidizing an R&D collaboration between *Dynavax Technologies*, a U.S.-based biopharmaceutical company, and *Shionogi & CO. Ltd.*, a Japanese pharmaceutical company, would increase welfare by 0.76%. Not taking into account the endogeneity of the network would yield much lower welfare gains. As subsidies are increasingly being used by governmental organizations to stimulate collaborative R&D activities, our framework could assist governmental funding agencies that typically do not take into account the aggregate spillovers generated within a dynamic R&D network structure.

**Related literature** There is a growing literature on the stochastic evolution of networks going back to [Jackson and Watts \[2002\]](#),<sup>6</sup> using tools from evolutionary game theory [[Blume, 1993](#); [Kandori et al., 1993](#); [Sandholm, 2010](#)] to analyze the formation of social and economic networks. In this literature agents form links over time based on myopic improvements that the resulting network offers them relative to the current network. While there is a small probability of mistakes, the stochastically stable states are identified when this probability vanishes. Our paper uses similar techniques to analyze the stationary states in a stochastic network formation model, but different from the above mentioned works, we investigate the coevolution of links and actions, and develop an estimable framework from our theory that can be applied to real world networks.

There also exist related studies on the formation of R&D networks in the economics literature. Similar to our framework, [Goyal and Moraga-Gonzalez \[2001\]](#), [Dawid and Hellmann \[2014\]](#), 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 perturbed 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 a collaboration in these models is independent of the identity and the characteristics of the firms involved.<sup>7</sup>

Similarly, [Ehrhardt et al. \[2008\]](#) analyze the formation of a network in which agents play a coordination game with their neighbors, while [König et al. \[2014a\]](#), [Marsili et al. \[2004\]](#), and [Fosco et al. \[2010\]](#) study the coevolution of networks and behavior. As in the present paper, these authors show that the interplay between action choice and link creation may feed on each other to generate sharp transitions from sparse to dense networks. The underlying payoff structure, however, is different from ours. Further, while these authors assume that links decay at random, here link removal depends on whether the agents find this profitable.

Our analysis also bears similarities with a number of other recent contributions in the literature which analyze a similar payoff structure. In the paper by [Ballester et al. \[2006\]](#), the authors derive equilibrium outcomes in a linear quadratic game where agents' efforts are local complements in an exogenously given network. Different from [Ballester et al. \[2006\]](#), we make the network as well as effort choices endogenous.<sup>8</sup> [Cabrales et al. \[2011\]](#) allow the network to be formed endogenously,

---

<sup>6</sup>For more recent contributions to this literature see, for example, [Hojman and Szeidl \[2006\]](#), [Feri \[2007\]](#) and [Dawid and Hellmann \[2014\]](#).

<sup>7</sup>[Goyal and Moraga-Gonzalez \[2001\]](#) present a more general setup which relaxes this assumption but their analysis is restricted to regular graphs and networks comprising of four firms. In this paper we take into account general equilibrium structures with an arbitrary number of firms and make no ex ante restriction on the potential collaboration pattern between them.

<sup>8</sup>It is straightforward to see that the results obtained in this paper can be generalized to the payoff structure introduced in [Ballester et al. \[2006\]](#). See in particular the general payoff structure considered in Equation (1). We provide a complete equilibrium characterization for the model introduced in [Ballester et al. \[2006\]](#), but allow both

but assume that link strengths are proportional to effort levels, while we make the linking decision depending on marginal payoffs. [Hiller \[2013\]](#) studies the joint formation of links and actions using a similar payoff structure as we do here, however, abstracting from any global substitutability effects, and shows that equilibrium networks are nested split graphs [see also [König et al., 2014a](#)]. Similarly, [Belhaj et al. \[2016\]](#) analyze the design of optimal networks with the same payoff function but without global substitutabilities, and show that when the planner chooses links, but not the level of output (second best), the optimal network is a nested split graph. We find that when the planner chooses both actions and links (first best), both equilibrium and efficient structures are nested spit graphs, even when allowing for global substitutes and incorporating heterogeneous firms, and we provide a more precise equilibrium characterization beyond the general class of nested split graphs. In particular, we identify conditions under which both the output and the degree distributions follow a power law, consistent with the empirical data [[Gabaix, 2016](#); [Powell et al., 2005](#)].

Our approach is a further generalization of the endogenous network formation mechanisms proposed in [Snijders \[2001\]](#), [Chandrasekhar and Jackson \[2012\]](#), and [Mele \[2017\]](#). As in these papers, we use a potential function to characterize the stationary states [[Monderer and Shapley, 1996](#)], but here both, the action choices as well as the linking decisions are fully endogenized. Moreover, different from these papers we provide a microfoundation (from a Cournot competition model with externalities) for the potential function. Further, in a recent paper by [Badev \[2013\]](#) a potential function is used to analyze the formation of networks in which agents not only form links but also make a binary choice of adopting a certain behavior depending on the choices of their neighbors. Different from [Badev \[2013\]](#), we consider a continuum of choices, and provide a microfoundation derived from the payoff function introduced in [Ballester et al. \[2006\]](#). Moreover, different from the previous authors we provide an explicit equilibrium characterization, use an alternative estimation method (which can also be applied to large networks and addresses the local trap problem),<sup>9</sup> apply our model to a different context, and study a range of novel counterfactual policy scenarios. Relatedly, in a recent paper, [Hsieh and Lee \[2013\]](#) 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 between payoff maximizing agents.

Furthermore, [Bimpikis et al. \[2014\]](#) analyze the effect of mergers and acquisitions within and across different industries using a similar model as we do here. However, neither does their analysis incorporate the spillover effects from R&D collaborations, nor do they allow for these collaborations to respond to a merger. In contrast, our empirical analysis reveals that the network structure of R&D collaborations matters, and that when ignoring the network structure the impact of mergers is significantly underestimated. Moreover, we find that mergers between highly connected firms can be welfare improving, a feature that would not arise in the absence of the R&D network. We thus contribute to the ongoing debate about the validity of antitrust policies in innovative industries [[Encaoua and Hollander, 2002](#)], by adding another dimension

---

agents' actions and links to be endogenously determined.

<sup>9</sup>Note also that classical Maximum Likelihood Estimation (MLE) methods such as the one considered in [Badev \[2013\]](#) are greatly influenced by the choice of initial parameter values, and if these are not close enough to the true values, the method may converge to a sub-optimal solution [[Airolidi et al., 2009](#)].

that explicitly takes into account efficiency gains that can be realized from R&D spillovers across firms.

Finally, in König et al. [2014b] a similar market structure is considered. In particular, the authors investigate the impact of a subsidy per unit of R&D spent, while here we analyze subsidies to R&D collaborations directly. Moreover, the authors characterize key firms whose exit would have the largest impact on the output of the economy in the short run, taking the network as given. Here we develop a long run analysis, where the network is allowed to dynamically adjust upon the exit of a firm. To the best of your knowledge, this is the first paper to perform such a dynamic key player analysis in a fully strategic environment.

**Organization of the paper** The paper is organized as follows. The theoretical model is outlined in Section 2. In particular, Section 2.1 introduces the linear-quadratic payoff function considered in this paper. Section 2.2 defines the stochastic network formation and output adjustment process and provides a complete characterization of the stationary state. In Section 2.3 the welfare maximizing networks are derived. Section 2.4 discusses several extensions of the model that allow for firm heterogeneity. Next, Section 3 provides information about the data that we use and explains the estimation methods and results. Section 4 then uses the estimated model to analyze several counterfactual policy experiments. Finally, Section 5 concludes. All proofs are relegated to Appendix A.

Additional relevant material can be found in the supplementary appendices. In particular, supplementary Appendix B provides basic definitions and characterizations of networks. Supplementary Appendix C provides a motivation for the linear quadratic payoff function from a model of R&D collaborating firms that are competing on the product market à la Cournot. Supplementary Appendix D explains the distinction between continuous and discrete quantity choices. 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 in Section 3, while supplementary Appendix G provides additional details of the estimation algorithms. Supplementary Appendix H provides a simulation study to examine the performance and consistency of our various estimation algorithms, as well as the impact of missing observations on estimation. Finally, supplementary Appendix I shows the robustness of our results when analyzing alternative data.

## 2. Theoretical Framework

### 2.1. Payoffs

Each firm (agent)  $i \in \mathcal{N} = \{1, \dots, n\}$  in the network  $G \in \mathcal{G}^n$  selects an output (action) level  $q_i \in \mathcal{Q}$  and obtains a linear-quadratic profit (payoff)  $\pi_i : \mathcal{Q}^n \times \mathcal{G}^n \rightarrow \mathbb{R}$  given by<sup>10</sup>

$$\pi_i(\mathbf{q}, G) = \eta_i q_i - \nu q_i^2 - b q_i \sum_{j \neq i}^n q_j + \rho \sum_{j=1}^n a_{ij} q_i q_j - \zeta d_i, \quad (1)$$

---

<sup>10</sup>See also Ballester et al. [2006] and Jackson et al. [2015] for a more general discussion of the payoff function introduced in Equation (1).

where  $\mathcal{Q}$  is the (bounded) output choice set of a firm,  $\mathcal{G}^n$  denotes the set of all graphs with  $n \geq 2$  nodes,  $a_{ij} = 1$  if firms  $i$  and  $j$  set up a collaboration (0 otherwise) and  $a_{ii} = 0$ .<sup>11</sup> Equation (1) is concave in own output,  $q_i$ , with parameters  $\eta_i \geq 0$  and  $\nu \geq 0$ . Moreover,  $b > 0$  is a global substitutability parameter,  $\rho \geq 0$  a local complementarity parameter,  $\zeta \geq 0$  denotes a fixed linking cost and  $d_i$  is the number of collaborations of firm  $i$ . A derivation of the profit function in Equation (1) in the context of R&D collaborating firms competing à la Cournot can be found in supplementary Appendix C.

The profit function introduced in Equation (1) admits a (cardinal) potential function [Monderer and Shapley, 1996].

**Proposition 1.** *The profit function of Equation (1) admits a potential game where firms choose both output and links with a potential function  $\Phi: \mathcal{Q}^n \times \mathcal{G}^n \rightarrow \mathbb{R}$  given by*

$$\Phi(\mathbf{q}, G) = \sum_{i=1}^n (\eta_i q_i - \nu q_i^2) - \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j + \frac{\rho}{2} \sum_{i=1}^n \sum_{j=1}^n a_{ij} q_i q_j - \zeta m, \quad (2)$$

for any  $\mathbf{q} \in \mathcal{Q}^n$  and  $G \in \mathcal{G}^n$  where  $m$  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 exactly equivalent to the difference in the potential function from adding or removing a link. Similarly, the marginal profit of a firm from changing its output level is exactly equivalent to the change of the potential function.<sup>12</sup> The potential function thus allows to aggregate the incentives of the firms to either change their links or adjust their production levels in a single global function. The existence of a potential function will be crucial for the equilibrium characterization of the network formation process that will be introduced in the following section.

## 2.2. Network Dynamics and Equilibrium Characterization

In this section we introduce a dynamic model where firms choose both output and links, based on the profit function of Equation (1). In this model the network is formed endogenously, based on the decisions of firms with whom to collaborate,<sup>13</sup> and share knowledge about a cost reducing technology. The opportunities for change arrive as a Poisson process [Blume, 1993; Sandholm, 2010], similar to Calvo models of pricing [Calvo, 1983]. As in the best reply dynamics analyzed in Cournot [1838],<sup>14</sup> when a firm receives such an opportunity to change it adjusts its output or collaborations so as to maximize its profit while taking the output levels and collaborations of the other firms as given, assuming that firms are myopic (also see e.g., Jackson and Watts [2002]; Watts [2001]).<sup>15</sup> To capture the fact that R&D projects and the establishment of an R&D

<sup>11</sup>See supplementary Appendix B for further network definitions and characterizations.

<sup>12</sup>More formally, the potential  $\Phi$  has the property that for any  $\mathbf{q} \in \mathcal{Q}^n$  and  $G, G' \in \mathcal{G}^n$  with  $G' = G \oplus (i, j)$  or  $G' = G \ominus (i, j)$  we have that  $\Phi(\mathbf{q}, G') - \Phi(\mathbf{q}, G) = \pi_i(\mathbf{q}, G') - \pi_i(\mathbf{q}, G)$ , where  $G \oplus (i, j)$  ( $G \ominus (i, j)$ ) denotes the network obtained from  $G$  by adding (removing) the link  $(i, j)$ . Moreover, for  $q_i, q'_i \in \mathcal{Q}$ ,  $\mathbf{q}_{-i} \in \mathcal{Q}^{n-1}$  and  $G \in \mathcal{G}^n$  we have that  $\Phi(q'_i, \mathbf{q}_{-i}, G) - \Phi(q_i, \mathbf{q}_{-i}, G) = \pi_i(q'_i, \mathbf{q}_{-i}, G) - \pi_i(q_i, \mathbf{q}_{-i}, G)$ .

<sup>13</sup>Note that the formation of a collaboration requires the *mutual agreement* of both firms involved in a collaboration, while for the termination of a collaboration it is sufficient that one of the firms finds this profitable [Jackson and Watts, 2002].

<sup>14</sup>Cournot [1838] analyzed a dynamic process in which firms myopically best respond in the current period to the existing output levels of all rivals. See also Daughety [2005].

<sup>15</sup>The assumption of myopic firms is also common in boundedly rational dynamic decision-making, as considered in Gabaix [2014].



collaboration are fraught with ambiguity and uncertainty [Czarnitzki et al., 2015; Kelly et al., 2002], we will introduce noise in this decision process. The precise definition of the dynamics of output adjustment and network evolution is given in the following.

**Definition 1 (Cournot Best Reply Dynamics).** *The evolution of the population of firms and the collaborations between them is characterized by a sequence of states  $(\omega_t)_{t \in \mathbb{R}_+}$ ,  $\omega_t \in \Omega = \mathcal{Q}^n \times \mathcal{G}^n$ , where each state  $\omega_t = (\mathbf{q}_t, G_t)$  consists of a vector of firms' output levels,  $\mathbf{q}_t \in \mathcal{Q}^n$ , and a network of collaborations,  $G_t \in \mathcal{G}^n$ . We assume that firms choose quantities from a bounded set  $\mathcal{Q}$ . In a short time interval  $[t, t + \Delta t)$ ,  $t \in \mathbb{R}_+$ , one of the following events happens:*

**Output adjustment** *At rate  $\chi \geq 0$  a firm  $i \in \mathcal{N}$  is selected at random and given an adjustment opportunity of its current output level. When firm  $i$  receives such an adjustment opportunity, the probability to choose a certain output level is governed by a multinomial logistic function with choice set  $\mathcal{Q}$  and parameter  $\vartheta$ , so that the probability that we observe a switch by firm  $i$  to an output level  $q$  conditional on the output levels of all other firms,  $\mathbf{q}_{-it}$ , and the network,  $G_t$ , at time  $t$  is given by<sup>16</sup>*

$$\mathbb{P}(\omega_{t+\Delta t} = (q, \mathbf{q}_{-it}, G_t) | \omega_t = (\mathbf{q}_t, G_t)) = \chi \frac{e^{\vartheta \pi_i(q, \mathbf{q}_{-it}, G_t)}}{\int_{\mathcal{Q}} e^{\vartheta \pi_i(q', \mathbf{q}_{-it}, G_t)} dq'} \Delta t + o(\Delta t). \quad (3)$$

**Link formation** *With rate  $\tau \geq 0$  a pair of firms  $i, j$  which is not already connected receives an opportunity to form a link. The formation of a link depends on the marginal profits the firms receive from the link plus an additive pairwise i.i.d. error term  $\varepsilon_{ij,t}$ . The probability that link  $(i, j)$  is created is then given by  $\mathbb{P}(\omega_{t+\Delta t} = (\mathbf{q}_t, G_t \oplus (i, j)) | \omega_t = (\mathbf{q}_t, G_t)) = \tau \mathbb{P}(\{\pi_i(\mathbf{q}_t, G_t \oplus (i, j)) - \pi_i(\mathbf{q}_t, G_t) + \varepsilon_{ij,t} > 0\} \cap \{\pi_j(\mathbf{q}_t, G_t \oplus (i, j)) - \pi_j(\mathbf{q}_t, G_t) + \varepsilon_{ij,t} > 0\}) \Delta t + o(\Delta t)$ . Using the fact that  $\pi_i(\mathbf{q}_t, G_t \oplus (i, j)) - \pi_i(\mathbf{q}_t, G_t) = \pi_j(\mathbf{q}_t, G_t \oplus (i, j)) - \pi_j(\mathbf{q}_t, G_t) = \Phi(\mathbf{q}_t, G_t \oplus (i, j)) - \Phi(\mathbf{q}_t, G_t)$ , and assuming that the error term  $\varepsilon_{ij,t}$  is independently logistically distributed,<sup>17</sup> we obtain*

$$\mathbb{P}(\omega_{t+\Delta t} = (\mathbf{q}_t, G_t \oplus (i, j)) | \omega_t = (\mathbf{q}_t, G_t)) = \tau \frac{e^{\vartheta \Phi(\mathbf{q}_t, G_t \oplus (i, j))}}{e^{\vartheta \Phi(\mathbf{q}_t, G_t \oplus (i, j))} + e^{\vartheta \Phi(\mathbf{q}_t, G_t)}} \Delta t + o(\Delta t). \quad (4)$$

**Link removal** *With rate  $\xi \geq 0$  a pair of connected firms  $i, j$  receives an opportunity to terminate their collaboration. The link is removed if at least one firm finds this profitable. The marginal profits from removing the link  $(i, j)$  are perturbed by an additive pairwise i.i.d. error term  $\varepsilon_{ij,t}$ . The probability that the link  $(i, j)$  is removed is then given by  $\mathbb{P}(\omega_{t+\Delta t} = (\mathbf{q}_t, G_t \ominus (i, j)) | \omega_t = (\mathbf{q}_t, G_t)) = \xi \mathbb{P}(\{\pi_i(\mathbf{q}_t, G_t \ominus (i, j)) - \pi_i(\mathbf{q}_t, G_t) + \varepsilon_{ij,t} > 0\} \cup \{\pi_j(\mathbf{q}_t, G_t \ominus (i, j)) - \pi_j(\mathbf{q}_t, G_t) + \varepsilon_{ij,t} > 0\}) \Delta t + o(\Delta t)$ . Using the fact that  $\pi_i(\mathbf{q}_t, G_t \ominus (i, j)) - \pi_i(\mathbf{q}_t, G_t) = \pi_j(\mathbf{q}_t, G_t \ominus (i, j)) - \pi_j(\mathbf{q}_t, G_t) = \Phi(\mathbf{q}_t, G_t \ominus (i, j)) - \Phi(\mathbf{q}_t, G_t)$ , and assuming that the error term is independently logistically distributed we obtain*

$$\mathbb{P}(\omega_{t+\Delta t} = (\mathbf{q}_t, G_t \ominus (i, j)) | \omega_t = (\mathbf{q}_t, G_t)) = \xi \frac{e^{\vartheta \Phi(\mathbf{q}_t, G_t \ominus (i, j))}}{e^{\vartheta \Phi(\mathbf{q}_t, G_t \ominus (i, j))} + e^{\vartheta \Phi(\mathbf{q}_t, G_t)}} \Delta t + o(\Delta t). \quad (5)$$

We assume that the set  $\mathcal{Q}$  is a discretization of the bounded interval  $[0, \bar{q}]$ , with  $\bar{q} < \infty$ . The fact that  $\mathcal{Q}$  is a countable set allows us to use standard results for discrete state space,

<sup>16</sup>The multinomial choice probability can be derived from a random utility model where firms maximize profits subject to a random error term [Anderson et al., 2004; McFadden, 1976]. See supplementary Appendix D for more details.

<sup>17</sup>Let  $z$  be i.i. logistically distributed with mean 0 and scale parameter  $\vartheta$ , i.e.  $F_z(x) = \frac{e^{\vartheta x}}{1 + e^{\vartheta x}}$ . Consider the random variable  $\varepsilon = g(z) = -z$ . Since  $g$  is monotonic decreasing, and  $z$  is a continuous random variable, the distribution of  $\varepsilon$  is given by  $F_\varepsilon(y) = 1 - F_z(g^{-1}(y)) = \frac{e^{\vartheta y}}{1 + e^{\vartheta y}}$ .

continuous time Markov chains [Norris, 1998]. For an increasingly fine discretization we then can use Equation (3) as a continuous analogue of the standard multinomial logit probabilistic choice framework [see also Anderson et al., 2004, 2001; Ben-Akiva and Watanatada, 1981; McFadden, 1976], where the probability of choosing an output level is proportional to an exponential function of the firm’s profit. The standard derivation of the logit model is based on the assumption that profits are subject to noise from a type-I extreme value distribution [Anderson et al., 1992].<sup>18</sup> The parameter  $\vartheta$  is inversely related to the level of noise, so that in the limit of  $\vartheta \rightarrow \infty$  the noise vanishes and firms choose the output level that maximizes their profit, while in the limit of  $\vartheta \rightarrow 0$  the noise dominates and output adjustments in Equation (3) become random. The same holds for the link formation and removal decisions in Equations (4) and (5), respectively.<sup>19</sup>

Note that we can numerically implement the stochastic process introduced in Definition 1 using the “next reaction method” for simulating a continuous time Markov chain [Gibson and Bruck, 2000]. We will use this method throughout the paper to illustrate our theoretical predictions for various network statistics (see Figures 1 and 3). However, this method becomes computationally infeasible for large networks, where we have to rely on our theoretical equilibrium characterization and alternative simulation methods.

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

The one step transition probability matrix  $\mathbf{P}^\vartheta(t) : \Omega^2 \rightarrow [0, 1]^{|\Omega|^2}$  has elements which determine the probability of a transition from a state  $\omega \in \Omega$  to a state  $\omega' \in \Omega$  in a small time interval  $[t, t + \Delta t)$  of length  $\Delta t$  given by  $\mathbb{P}(\omega_{t+\Delta t} = \omega' | \mathcal{F}_t = \sigma(\omega_0, \omega_1, \dots, \omega_t = \omega)) = \mathbb{P}(\omega_{t+\Delta t} = \omega' | \omega_t = \omega) = q^\vartheta(\omega, \omega')\Delta t + o(\Delta t)$  if  $\omega' \neq \omega$  and  $\mathbb{P}(\omega_{t+\Delta t} = \omega | \mathcal{F}_t = \sigma(\omega_0, \omega_1, \dots, \omega_t = \omega)) = \mathbb{P}(\omega_{t+\Delta t} = \omega | \omega_t = \omega) = 1 + q^\vartheta(\omega, \omega)\Delta t + o(\Delta t)$ , where  $q^\vartheta(\omega, \omega')$  is the transition rate from state  $\omega$  to state  $\omega'$  and  $\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0$  (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)  $\mathbf{Q}^\vartheta = (q^\vartheta(\omega, \omega'))_{\omega, \omega' \in \Omega}$  of the Markov chain has the elements

$$q^\vartheta(\omega, \omega') = \begin{cases} \chi \frac{e^{\vartheta \pi_i(q', \mathbf{q}_{-i}, G)}}{\int_{\mathcal{Q}} e^{\vartheta \pi_i(q'', \mathbf{q}_{-i}, G)} dq''} & \text{if } \omega' = (q', \mathbf{q}_{-i}, G) \text{ and } \omega = (\mathbf{q}, G), \\ \tau \frac{e^{\vartheta \Phi(\mathbf{q}, G \oplus (i, j))}}{e^{\vartheta \Phi(\mathbf{q}, G \oplus (i, j))} + e^{\vartheta \Phi(\mathbf{q}, G)}} & \text{if } \omega' = (\mathbf{q}, G \oplus (i, j)) \text{ and } \omega = (\mathbf{q}, G), \\ \xi \frac{e^{\vartheta \Phi(\mathbf{q}, G \ominus (i, j))}}{e^{\vartheta \Phi(\mathbf{q}, G \ominus (i, j))} + e^{\vartheta \Phi(\mathbf{q}, G)}} & \text{if } \omega' = (\mathbf{q}, G \ominus (i, j)) \text{ and } \omega = (\mathbf{q}, G), \\ - \sum_{\omega' \neq \omega} q^\vartheta(\omega, \omega') & \text{if } \omega' = \omega, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

<sup>18</sup>See supplementary Appendix D for a derivation of the multinomial logit model with a continuous choice set from a multinomial model with a discrete choice set [see also McFadden, 1976]. This illustrates that our framework can also be applied to cases where agents choose from an arbitrary discrete set of alternatives.

<sup>19</sup>While in Definition 1 pairs of firms are selected in a global way to form collaborations, it is possible to consider a local approach where new links are more likely to be formed among firms which already have a common neighbor [Jackson and Rogers, 2007]. This can be captured by a linking cost that is decreasing in the number of common neighbors of the firms. We explicitly consider such a formulation in the empirical model in Section 3.2.

with  $\sum_{\omega' \in \Omega} q^\vartheta(\omega, \omega') = 0$ .<sup>20</sup> As the Markov chain is time homogeneous, the transition rates are independent of time. The stationary distribution  $\mu^\vartheta : \Omega \rightarrow [0, 1]$  is then the solution to  $\mu^\vartheta \mathbf{P}^\vartheta = \mu^\vartheta$ , or equivalently  $\mu^\vartheta \mathbf{Q}^\vartheta = \mathbf{0}$  [Norris, 1998].

With the potential function  $\Phi$  of Proposition 1 we can compute the stationary distribution in the form of a Gibbs measure [Grimmett, 2010].

**Theorem 1.** *The stochastic process  $(\omega_t)_{t \in \mathbb{R}_+}$  with states  $\omega_t \in \Omega = \mathcal{Q}^n \times \mathcal{G}^n$  is an ergodic Markov chain with a unique stationary distribution  $\mu^\vartheta : \Omega \rightarrow [0, 1]$  such that  $\lim_{t \rightarrow \infty} \mathbb{P}(\omega_t = (\mathbf{q}, G) | \omega_0 = (\mathbf{q}_0, G_0)) = \mu^\vartheta(\mathbf{q}, G)$ . The distribution  $\mu^\vartheta$  is given by*

$$\mu^\vartheta(\mathbf{q}, G) = \frac{e^{\vartheta(\Phi(\mathbf{q}, G) - m \ln(\frac{\xi}{\tau}))}}{\sum_{G' \in \mathcal{G}^n} \int_{\mathcal{Q}^n} e^{\vartheta(\Phi(\mathbf{q}', G') - m' \ln(\frac{\xi}{\tau}))} d\mathbf{q}'}, \quad (7)$$

for any  $\mathbf{q} \in \mathcal{Q}^n$  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

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{\omega_s \in \mathcal{A}\}} ds = \mu^\vartheta(\mathcal{A}), \quad \mathbb{P}\text{-a.s.}, \quad (8)$$

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

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

where  $\mathbb{E}_{\mu^\vartheta}(f)$  is the expected value of  $f$  under the invariant probability measure  $\mu^\vartheta$ . Note that the stationary distribution  $\mu^\vartheta$  in Equation (7) does not depend on the output adjustment rate  $\chi$ . It also does not depend on the link adjustment rates  $\tau$  and  $\xi$ , when the rates for link creation and removal are the same. In the following we will make the simplifying assumption that  $\tau = \xi$ .

In the limit of vanishing noise, i.e.,  $\vartheta \rightarrow \infty$ , we call the states in the support of  $\mu^\vartheta$  the *stochastically stable* states [Jackson and Watts, 2002; Kandori et al., 1993]. A state  $(\mathbf{q}, G) \in \Omega$  is stochastically stable if  $\lim_{\vartheta \rightarrow \infty} \mu^\vartheta(\mathbf{q}, G) > 0$ . In the following we will denote  $\mu^*(\mathbf{q}, G) \equiv \lim_{\vartheta \rightarrow \infty} \mu^\vartheta(\mathbf{q}, G)$ . The set of stochastically stable states is denoted by  $\Omega^* \subseteq \Omega$ . From the Gibbs distribution in Equation (7) it follows that  $(\mathbf{q}, G) \in \Omega^*$  if and only if  $\Phi(\mathbf{q}, G) \geq \Phi(\mathbf{q}', G') \forall \mathbf{q}' \in \mathcal{Q}^n$  and  $\forall G' \in \mathcal{G}^n$ .

As the potential function  $\Phi$  in Equation (2) is continuous on the compact set  $\Omega$  it has a global maximum, stochastically stable states always exist and  $\Omega^* \neq \emptyset$ . An explicit characterization of the stationary distribution  $\mu^\vartheta$  in Equation (7) requires the computation of the *partition function*

$$\mathcal{Z}_\vartheta = \sum_{G \in \mathcal{G}^n} \int_{\mathcal{Q}^n} e^{\vartheta \Phi(\mathbf{q}, G)} d\mathbf{q}, \quad (9)$$

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

<sup>20</sup>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  $\mathbf{Q}^\vartheta = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}^\vartheta(t + \Delta t) - \mathbf{I}_{|\Omega|}}{\Delta t}$ .

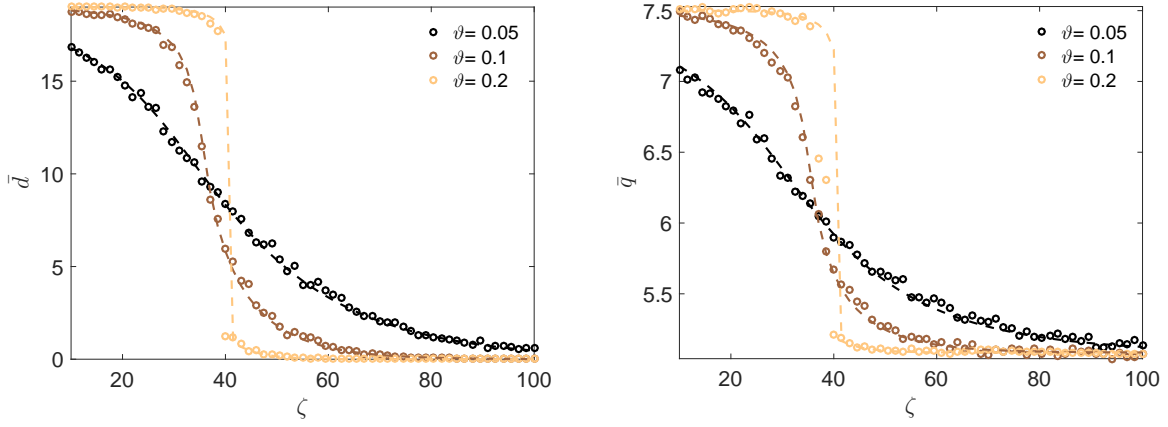


Figure 1: The average degree  $\bar{d}$  (left panel) and the average output  $\bar{q}$  (right panel) as a function of the linking cost  $\zeta$  for varying values of  $\vartheta \in \{0.05, 0.1, 0.2\}$  with  $n = 20$  firms and  $\tau = \xi = \chi = 1$ ,  $\eta = 300$ ,  $\rho = 1$ ,  $b = 1$  and  $\nu = 20$ . Dashed lines indicate the theoretical predictions of Equations (10) and Equation (12) in Proposition 2, respectively.

function more compactly as  $\mathcal{Z}_\vartheta = \int_{\mathcal{Q}^n} e^{\vartheta \mathcal{H}_\vartheta(\mathbf{q})} d\mathbf{q}$ .

In the following we provide an explicit characterization of the Gibbs distribution in Theorem 1 and derive the stochastically stable states. We first consider the special case of ex ante homogeneous firms with identical marginal costs.

**Proposition 2.** *Consider homogeneous firms such that  $\eta_i = \eta$  in the profit function of Equation (1) for all  $i = 1, \dots, n$ , and let the evolution of the firms output levels and collaborations be governed by the stochastic process in Definition 1. Denote  $\eta^* \equiv \eta/(n-1)$  and  $\nu^* \equiv \nu/(n-1)$ . Moreover, let the empirical average output be denoted by  $\bar{q} \equiv \frac{1}{n} \sum_{i=1}^n q_i$  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  $\bar{P}^\vartheta(k) \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{d_i=k\}}$ , and be denoted by  $P^\vartheta(k) \equiv \mathbb{E}_{\mu^\vartheta} (\bar{P}^\vartheta(k))$ .*

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

$$(b + 2\nu^*)q - \eta^* = \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right) q, \quad (10)$$

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

$$P^\vartheta(k) = \mathbb{E}_{\mu^\vartheta} \left( \frac{e^{-\bar{d}(q_1)} \bar{d}(q_1)^k}{k!} \right) (1 + o(1)), \quad (11)$$

where the expected degree for large  $\vartheta$  is given by

$$\mathbb{E}_{\mu^\vartheta} (\bar{d}) = \frac{n-1}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho(q^*)^2 - \zeta) \right) \right) + \frac{1}{2\vartheta} \mathcal{R}_\vartheta, \quad (12)$$

$q^*$  is given by Equation (10) and the expression of the remainder term  $\mathcal{R}_\vartheta$  can be found in the proof of the proposition.

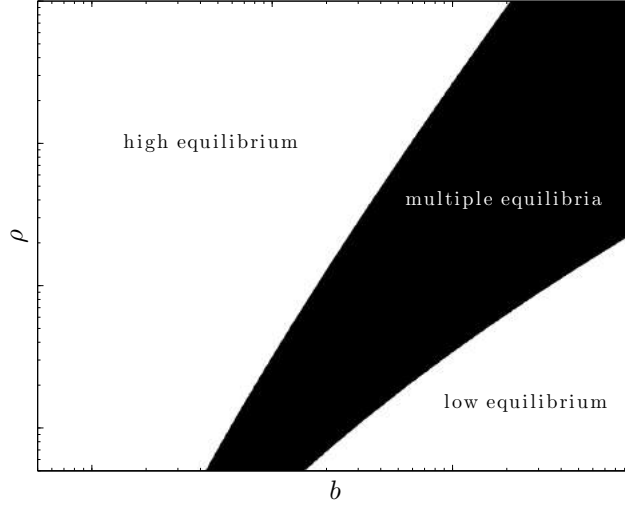


Figure 2: A phase diagram illustrating the regions with a unique and with multiple equilibria according to Equation (10) in Proposition 2 for varying values of  $b \in \{0, \dots, 0.01\}$  and  $\rho \in \{0, \dots, 0.01\}$  with  $n = 100$ ,  $\nu = 0.5$ ,  $\eta = 100$ ,  $\vartheta = 1$  and  $\zeta = 50$ .

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

$$q^* = \begin{cases} \frac{\eta^*}{b+2\nu^*-\rho}, & \text{if } \zeta < \frac{\rho(\eta^*)^2}{(b+2\nu^*)^2}, \\ \left\{ \frac{\eta^*}{b+2\nu^*-\rho}, \frac{\eta^*}{b+2\nu^*} \right\}, & \text{if } \frac{\rho(\eta^*)^2}{(b+2\nu^*)^2} < \zeta < \frac{\rho(\eta^*)^2}{(b+2\nu^*-\rho)^2}, \\ \frac{\eta^*}{b+2\nu^*}, & \text{if } \frac{\rho(\eta^*)^2}{(b+2\nu^*-\rho)^2} < \zeta, \end{cases} \quad (13)$$

and we refer to the two possible output levels in Equation (13) as the high equilibrium and the low equilibrium, respectively. The expected average degree in the high equilibrium is  $\mathbb{E}_{\mu^*}(\bar{d}) = \lim_{\vartheta \rightarrow \infty} \mathbb{E}_{\mu^\vartheta}(\bar{d}) = n - 1$ , which corresponds to a complete graph,  $K_n$ , and  $\mathbb{E}_{\mu^*}(\bar{d}) = \lim_{\vartheta \rightarrow \infty} \mathbb{E}_{\mu^\vartheta}(\bar{d}) = 0$  in the low equilibrium, which corresponds to an empty graph,  $\bar{K}_n$ .

Figure 1 shows the average output  $\bar{q}$  of Equation (10) and the average degree  $\bar{d}$  of Equation (12) in part (i) Proposition 2 as a function of the linking cost  $\zeta$ . With increasing cost, both the network connectivity and the output produced are decreasing. The transition from an economy with high output and collaboration intensity to an economy with low output and collaboration intensity is becoming sharper as  $\vartheta$  increases, consistent with the limit of part (ii) in Proposition 2. An illustration of  $\mu^\vartheta$  of the output distribution  $\mathcal{N}(q^*, \sigma^2)$  in part (i) of Proposition 2 together with the results of numerical simulations can be seen in the left panel in Figure 3. The figure shows that the analytic prediction reproduces the simulation results fairly well even for small values of  $\vartheta$ . A phase diagram illustrating the regions with a unique and with multiple equilibria according to Equation (10) can be seen in Figure 2. Note that the stationary output levels in Equation (10) are increasing in  $\rho$  and  $\eta$ , and decreasing in  $\zeta$  and  $b$  (see also Figure 1 in Appendix A). The latter implies that both higher collaboration cost (weaker spillovers) and more intense competition (larger market size/lower production costs) decrease overall production.

We next consider the more general case of firms with heterogeneous marginal production costs.

**Proposition 3.** *Let the firms' profits be given by Equation (1), and let the evolution of the firms' output levels and collaborations be governed by the stochastic process in Definition 1.*

(i) *For any  $\mathbf{q} \in \mathcal{Q}^n$  and  $G \in \mathcal{G}^n$  the stationary distribution of Equation (7) can be written as  $\mu^\vartheta(\mathbf{q}, G) = \mu^\vartheta(G|\mathbf{q})\mu^\vartheta(\mathbf{q})$ , where, for large  $\vartheta$ , the marginal distribution  $\mu^\vartheta(\mathbf{q})$  of the firms'*

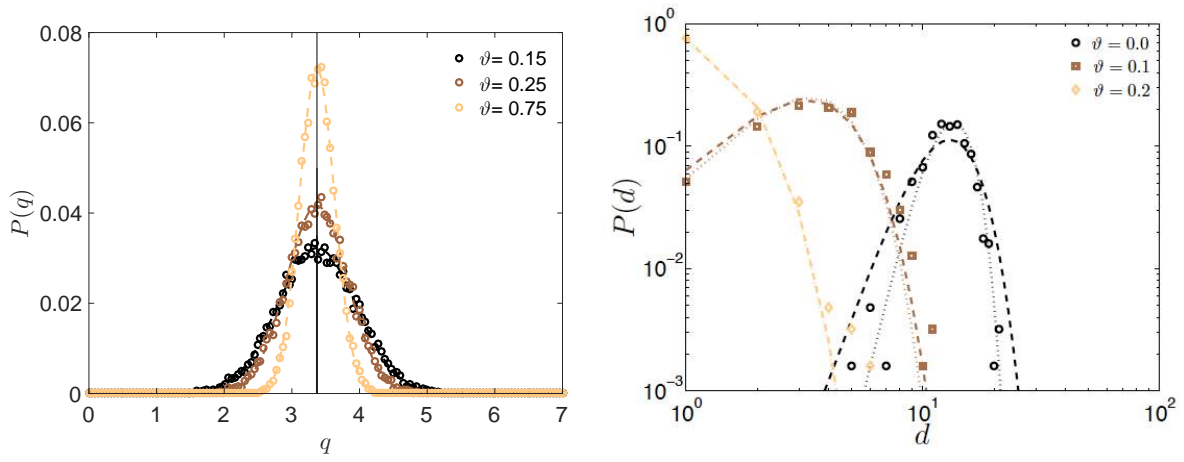


Figure 3: (Left panel) The stationary output distribution  $P(q)$  for  $n = 50$ ,  $\eta = 150$ ,  $b = 0.5$ ,  $\nu = 10$ ,  $\rho = 1$ ,  $\vartheta \in \{0.1, 0.25, 0.75\}$  and  $\zeta = 60$ . Dashed lines indicate the normal distribution  $\mathcal{N}(q^*, \sigma^2)$  of part(i) of Proposition 2). (Right panel) The stationary degree distribution  $P(k)$  for the same parameter values. The dashed lines indicate the solution in Equation (11) of Proposition 2.

output levels is asymptotically Gaussian and given by

$$\mu^\vartheta(\mathbf{q}) = \left(\frac{2\pi}{\vartheta}\right)^{-\frac{n}{2}} |\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top (-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)) (\mathbf{q} - \mathbf{q}^*) \right\} + o(\|\mathbf{q} - \mathbf{q}^*\|^2), \quad (14)$$

with mean  $\mathbf{q}^* \in \mathcal{Q}^n$  solving  $q_i^* = \eta_i + \sum_{j \neq i} \left( \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q_i^* q_j^* - \zeta) \right) \right) - b \right) q_j^*$ , and variance-covariance matrix given by the inverse of  $-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)$ ,<sup>21</sup> while the conditional distribution  $\mu^\vartheta(G|\mathbf{q})$  is given by

$$\mu^\vartheta(G|\mathbf{q}) = \prod_{i=1}^n \prod_{j=i+1}^n \frac{e^{\vartheta a_{ij}(\rho q_i q_j - \zeta)}}{1 + e^{\vartheta(\rho q_i q_j - \zeta)}}, \quad (15)$$

for any  $\mathbf{q} \in \mathcal{Q}^n$  and  $G \in \mathcal{G}^n$ , which corresponds to an inhomogeneous random graph with linking probability

$$p^\vartheta(q_i, q_j) = \frac{e^{\vartheta(\rho q_i q_j - \zeta)}}{1 + e^{\vartheta(\rho q_i q_j - \zeta)}}. \quad (16)$$

- (ii) In the limit of  $\vartheta \rightarrow \infty$ , the stochastically stable network  $G \in \mathcal{G}^n$  in the support of  $\mu^*$  is a nested split graph in which a link between the pair of firms  $i$  and  $j$  is present if and only if  $\rho q_i q_j > \zeta$ , and the output profile,  $\mathbf{q} \in \mathcal{Q}^n$ , is the fixed point to the following system of equations

$$q_i = \frac{\eta_i}{2\nu} + \frac{1}{2\nu} \sum_{j \neq i} q_j \left( \rho \mathbf{1}_{\{\rho q_i q_j > \zeta\}} - b \right), \quad \mu^* \text{-a.s.} \quad (17)$$

Moreover, if firms  $i$  and  $j$  are such that  $\eta_i > \eta_j$  then  $i$  has a higher output than  $j$ ,  $q_i > q_j$  and a larger number of collaborations,  $d_i > d_j$ ,  $\mu^*$ -a.s..

- (iii) Assume that  $(\eta_i)_{i=1}^n$  are identically and independently Pareto distributed with density function  $f(\eta) = (\gamma - 1)\eta^{-\gamma}$  for  $\eta \geq 1$ . Denote by  $\mathbf{M} \equiv \mathbf{I}_n + b\mathbf{B} - \rho\mathbf{A}$ , where  $\mathbf{B}$  is an  $n \times n$ -dim. matrix of ones with zero diagonal and  $\mathbf{A}$  has elements  $a_{ij} = \mathbf{1}_{\{\rho q_i q_j > \zeta\}}$ . Then the stochastically stable output distribution is given by  $\mu^*(\mathbf{q}) = (\gamma - 1)^n |\det(\mathbf{M})| \prod_{i=1}^n (\mathbf{M}\mathbf{q})_i^{-\gamma}$ . In particular, for  $\mathbf{q} = c\mathbf{u}$ , with  $c > 0$ , and  $\mathbf{u}$  being an  $n$ -dim. vector of ones, we have that  $\mu^*(c\mathbf{u}) \sim \prod_{i=1}^n O(c^{-\gamma})$  as  $c \rightarrow \infty$ , i.e., the output levels are asymptotically independently Pareto distributed.

<sup>21</sup>The variance-covariance matrix can be computed in closed form:  $(\Delta \mathcal{H}_\vartheta(\mathbf{q}))_{ii} = \frac{\vartheta \rho^2}{4} \sum_{j \neq i} q_j^2 \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right)^2 - 1$ , and  $(\Delta \mathcal{H}_\vartheta(\mathbf{q}))_{ij} = \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) \left( 1 + \frac{\vartheta \rho}{2} q_i q_j \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) \right) - b$ , for  $j \neq i$ .

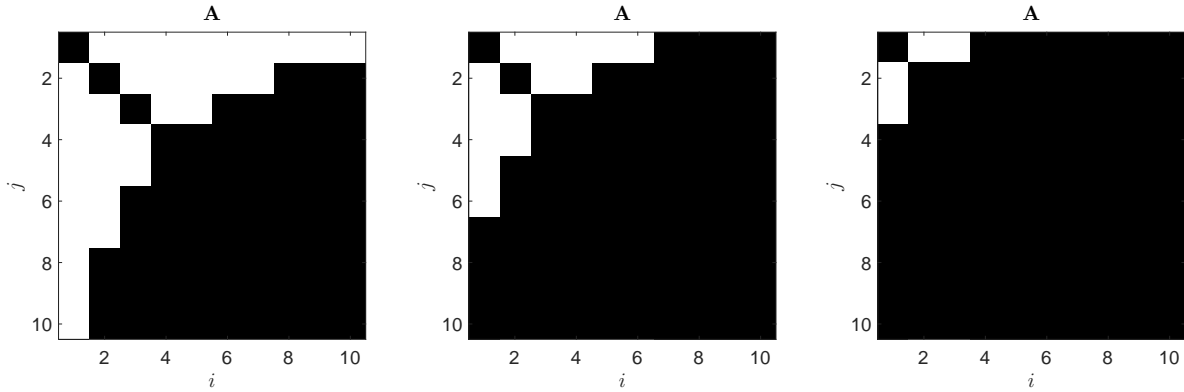


Figure 4: The (stepwise) adjacency matrix  $\mathbf{A} = (a_{ij})_{1 \leq i, j \leq n}$ , characteristic of a nested split graph, with elements given by  $a_{ij} = \mathbf{1}_{\{q_i q_j > \frac{\zeta}{\rho}\}}$ , where the vector  $\mathbf{q}$  is the solution to Equation (17) in Proposition 3. The panels from the left to the right correspond to increasing linking costs  $\zeta \in \{0.0075, 0.01, 0.02\}$ . The parameters used are  $n = 10$ ,  $\nu = 0.5$ ,  $b = 0.06$ ,  $\rho = 0.02$  and  $\boldsymbol{\eta} = (1.00, 0.71, 0.58, 0.50, 0.45, 0.41, 0.38, 0.35, 0.33, 0.32)^\top$ .

Note that the marginal probability derived in Equation (14) will be important for our estimation algorithm introduced in Section 3.3.1. Moreover, Figure 4 shows the adjacency matrix  $\mathbf{A} = (a_{ij})_{1 \leq i, j, n}$  whose elements are given by  $a_{ij} = \mathbf{1}_{\{q_i q_j > \frac{\zeta}{\rho}\}}$  and the vector  $\mathbf{q}$  is the solution to Equation (17) in part (ii) of Proposition 3. We observe that firms with higher  $\eta_i$  also have higher output levels,  $q_i$ , in the stationary state. Moreover, the corresponding adjacency matrix is stepwise, characterizing a nested split graph,<sup>22</sup> and becomes increasingly sparse with increasing linking cost  $\zeta$ . The fact that empirical R&D networks are characterized by nestedness has been documented in Tomasello et al. [2016]. Further, note that nested split graphs are paramount examples of core-periphery networks [Hojman and Szeidl, 2008]. The core-periphery structure of R&D alliance networks has also 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. Moreover, Kitsak et al. [2010] find that firms in the core have a higher market value, consistent with the predictions of our model. More productive firms have lower marginal costs of production, and thus higher output, R&D expenditures, and can form more links, which makes them more central in the network.

When the firms' marginal costs are not exogenously given, but when the  $(\eta_i)_{i=1}^n$  follow a power law distribution (which would correspond to Pareto distributed productivity levels that have been documented for example in König et al. [2016]), then part (iii) of Proposition 3 shows that the stationary output distribution is also a power law. The resulting degree distribution will then also be a power law.<sup>23</sup> An example based on a numerical simulation of the stochastic process of

<sup>22</sup>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 [König et al., 2014a; Mahadev and Peled, 1995]. See supplementary Appendix B for the definition of nested split graphs.

<sup>23</sup>Note that in the limit of  $\vartheta \rightarrow \infty$  the linking probability between two firms with output levels  $q$  and  $q'$ , respectively, is given by Equation (16):

$$\lim_{\vartheta \rightarrow \infty} p^\vartheta(q, q') = \lim_{\vartheta \rightarrow \infty} \frac{e^{\vartheta(\rho q q' - \zeta)}}{1 + e^{\vartheta(\rho q q' - \zeta)}} = \mathbf{1}_{\{\rho q q' > \zeta\}} = \mathbf{1}_{\{\log q + \log q' > \log(\frac{\zeta}{\rho})\}}. \quad (18)$$

When the output levels are power law distributed, with density  $f(x) = \frac{\gamma}{c} (\frac{c}{x})^{\gamma+1}$  for  $x > c$ , where  $c > 0$  is a lower-cut-off and  $\gamma > 0$  is a positive parameter, then the log-output levels,  $\ln q$ , follow an exponential distribution with density  $f(y) = \gamma c^\gamma e^{-\gamma y}$ . The linking probability in Equation (18) then induces an inhomogenous random graph identical to the one analyzed in Boguná and Pastor-Satorras [2003] (see also Appendix B). In particular, the authors show that this random graph is characterized by a power law degree distribution, a negative clustering degree correlation and a decaying average nearest neighbor degree distribution indicating a dissortative network. In Section 2.4.1 and Appendix E.1 we discuss how such network characteristics can also be obtained when firms

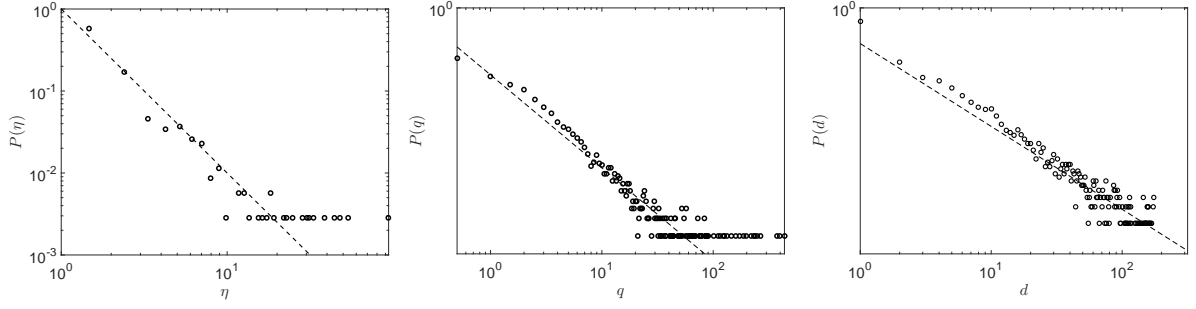


Figure 5: The distribution  $P(\eta)$  of  $\eta$  following a Pareto distribution with exponent 2 (left panel), the resulting stationary output distribution  $P(q)$  (middle panel) and the degree distribution  $P(d)$  (right panel) from a numerical simulation of the stochastic process of Definition 1. Dashed lines indicate a power-law fit. Observe that  $P(\eta)$  and  $P(q)$  exhibit a power law tail with the same exponent, consistent with part (iii) of Proposition 3. The parameters used are  $n = 350$ ,  $\nu = 0.95$ ,  $b = 0.75$ ,  $\rho = 2$  and  $\zeta = 75$ .

Definition 1 can be seen in Figure 5. Our model can thus provide an explanation for the joint occurrence of heavy tailed distributions not only in the firms' sizes [Gabaix, 2016], but also in their degrees [Powell et al., 2005].

### 2.3. Efficiency

Social welfare,  $W$ , is given by the sum of consumer surplus,  $U$ , and firms' profits,  $\Pi$ . Consumer surplus is given by  $U(\mathbf{q}) = \frac{1}{2} \sum_{i=1}^n q_i^2 + \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j$  (see supplementary Appendix C). In the special case of non-substitutable goods, when  $b \rightarrow 0$ , we obtain  $U(\mathbf{q}) = \frac{1}{2} \sum_{i=1}^n q_i^2$ , while in the case of perfectly substitutable goods, when  $b \rightarrow 1$ , we get  $U(\mathbf{q}) = \frac{1}{2} (\sum_{i=1}^n q_i)^2$ . Producer surplus is given by aggregate profits  $\Pi(\mathbf{q}, G) = \sum_{i=1}^n \pi_i(\mathbf{q}, G)$ . As a result, assuming homogeneous firms, total welfare is equal to

$$\begin{aligned} W(\mathbf{q}, G) &= U(\mathbf{q}) + \Pi(\mathbf{q}, G) = \frac{1}{2} \sum_{i=1}^n q_i^2 + \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j + \sum_{i=1}^n \pi_i(\mathbf{q}, G) \\ &= \frac{1}{2} \sum_{i=1}^n q_i^2 + \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j + \sum_{i=1}^n \left( \eta q_i - \nu q_i^2 - b \sum_{j \neq i}^n q_i q_j + \rho \sum_{j=1}^n a_{ij} q_i q_j \right) - 2\zeta m, \end{aligned} \quad (19)$$

where  $m$  denotes the number of links in  $G$ . The efficient state is then defined by the network  $G^* \in \mathcal{G}^n$  and output profile  $\mathbf{q}^* \in \mathcal{Q}^n$  that maximize welfare  $W(\mathbf{q}, G)$  in Equation (19), that is,  $W(\mathbf{q}^*, G^*) \geq W(\mathbf{q}, G)$  for all  $G \in \mathcal{G}^n$  and  $\mathbf{q} \in \mathcal{Q}^n$ .<sup>24</sup> The following proposition shows that the decentralized equilibrium is efficient only when the linking costs are sufficiently high. Otherwise, equilibrium networks are under-connected, and production is too low compared to what would be socially optimal.

**Proposition 4.** *Let the firms' profits be given by Equation (1), define welfare as in Equation (19), and let the evolution of the firms' output levels and collaborations be governed by the stochastic process in Definition 1. Further, denote by  $\eta^* \equiv \eta/(n-1)$  and  $\nu^* \equiv \nu/(n-1)$ .*

- (i) *In the case of homogeneous firms such that  $\eta_i = \eta$  for all  $i = 1, \dots, n$ , the efficient network  $G^* \in \mathcal{G}^n$  and output profile  $\mathbf{q}^* \in \mathcal{Q}^n$  are given by  $q^* \mathbf{u}$ , with  $\mathbf{u}$  denoting an  $n$ -dimensional*

are heterogeneous in terms of their marginal collaboration costs.

<sup>24</sup>Observe that this is different from the efficiency analysis in Ballester et al. [2006], where the planner chooses links, but not the effort levels, and there are not linking costs. It is also different from the efficiency analysis in Hiller [2013] and Belhaj et al. [2016], where global substitutability effects are not taken into account.



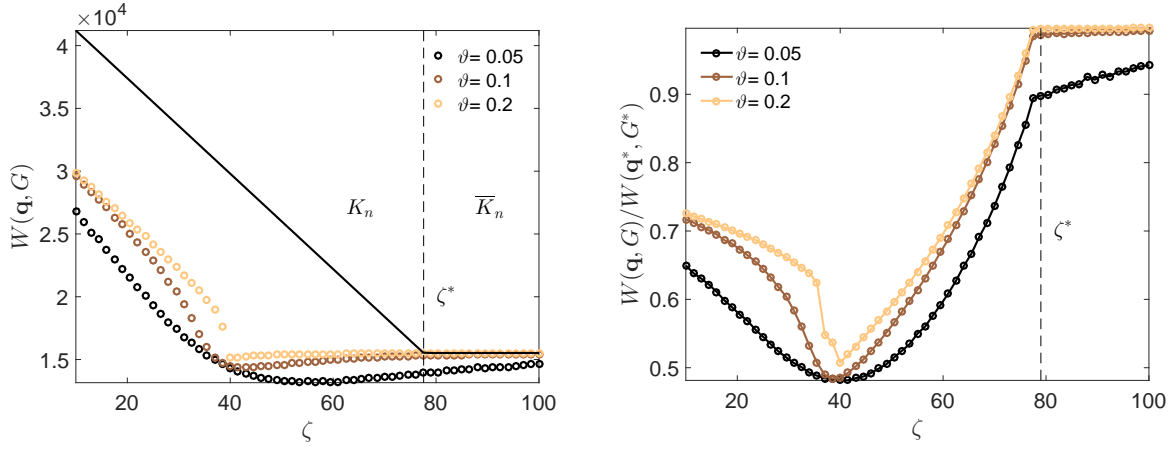


Figure 6: (Left panel) Welfare  $W(\mathbf{q}, G)$  as a function of the linking cost  $\zeta$  for varying values of  $\vartheta \in \{0.05, 0.1, 0.2\}$  with  $n = 20$  firms and  $\tau = \xi = \chi = 1$ ,  $\eta = 300$ ,  $\rho = 1$ ,  $b = 1$  and  $\nu = 20$ . The solid line indicates welfare in the efficient graph of Proposition 4 (which is either complete or empty). (Right panel) The ratio of welfare relative to welfare in the efficient graph.

vector of ones, and

$$(\mathbf{q}^*, G^*) = \begin{cases} \left( \frac{\eta^*}{b+2(\nu^*-\rho)-\frac{1}{n-1}}, K_n \right), & \text{if } \zeta \leq \zeta^*, \\ \left( \frac{\eta^*}{b+2\nu^*-\frac{1}{n-1}}, \bar{K}_n \right), & \text{if } \zeta^* < \zeta, \end{cases} \quad (20)$$

where  $K_n$  denotes the complete graph,  $\bar{K}_n$  denotes the empty graph and

$$\zeta^* = \frac{\rho(\eta^*)^2}{\left(b+2\nu^*-\frac{1}{n-1}\right)\left(b+2(\nu^*-\rho)-\frac{1}{n-1}\right)}. \quad (21)$$

Moreover, in the limit of  $\vartheta \rightarrow \infty$  the stochastically stable equilibrium network is efficient if  $\zeta > \zeta^*$ ,  $\mu^*$ -a.s..

(ii) In the case of heterogeneous firms, the efficient network  $G^* \in \mathcal{G}^n$  is a nested split graph, where the output profile  $\mathbf{q}^* \in \mathcal{Q}^n$  is the solution to the following system of equations

$$q_i = \frac{\eta_i}{2\nu-1} + \frac{1}{2\nu-1} \sum_{j \neq i}^n q_j \left( \rho \mathbf{1}_{\{\rho q_i q_j > \zeta\}} - b \right). \quad (22)$$

Further, when Equation (22) admits an interior solution, then the stochastically stable equilibrium output (and R&D expenditures) and the collaboration intensity are too low compared to the social optimum ( $\mu^*$ -a.s.).

The left panel in Figure 6 shows welfare as a function of the linking cost  $\zeta$  for varying values of  $\vartheta$ , while the right panel shows the ratio of welfare relative to welfare in the efficient graph in the case of homogeneous firms considered in part (i) of Proposition 4. It illustrates the region of inefficiency of the equilibrium network for linking costs  $\zeta < \zeta^*$ , where equilibrium networks typically tend to be under-connected.<sup>25</sup> As similar observation can also be made for heterogeneous firms considered in part (ii) of Proposition 4. In Section 4.3 we analyze the welfare improving impact

<sup>25</sup>In contrast, when the linking costs are very high, then the R&D externalities are not high enough to compensate for the costs of maintaining the network, and so the social planner prefers not to form any links. In this high cost region also the individual firms do not want to form links, so that the social planners solution and the decentralized equilibrium coincide.

of a subsidy on firms' R&D collaboration costs, that incentivises firms to form collaborations and thus increases the network connectivity.

## 2.4. 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.

### 2.4.1. Heterogeneous Collaboration Costs

We can extend the model by assuming that firms with higher productivity incur lower collaboration costs (see also supplementary Appendix E.1). 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 5),<sup>26</sup> consistent with previous empirical studies of R&D networks [see e.g., Powell et al., 2005], together with other empirically relevant correlations (see Propositions 6 and 7).<sup>27</sup>

### 2.4.2. Heterogeneous Technology Spillovers

We can further extend the model by assuming that there are heterogeneous spillovers between collaborating firms depending on their technology portfolios [Griffith et al., 2003] (see also supplementary Appendix E.2). For example, assume that firms can only benefit from collaborations if they have at least one technology in common. Then one can show that our model is a generalization of a “random intersection graph” [Deijfen and Kets, 2009] (see supplementary Appendix B) for which positive degree correlations can be obtained (i.e., “assortativity”, see Proposition 8).

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

---

<sup>26</sup>In particular, assume that the productivity  $s$  are distributed as a power law  $s^{-\gamma}$  with exponent  $\gamma$ . Then one can show that the asymptotic degree distribution is also power law distributed,  $P(k) \sim k^{-\frac{\gamma}{\gamma-1}}$ , with exponent  $\frac{\gamma}{\gamma-1}$ .

<sup>27</sup>We note that also other statistics such as the clustering degree distribution can be computed. See supplementary Appendix E.2 for further details. In particular, under the assumptions of a power law productivity distribution, we can generate two-vertex and three-vertex degree correlations, such as a decreasing average nearest neighbor connectivity,  $k_{nn}(d)$ , indicating a dissortative network, as well as a decreasing clustering degree distribution,  $C(d)$ , with the degree  $d$ .

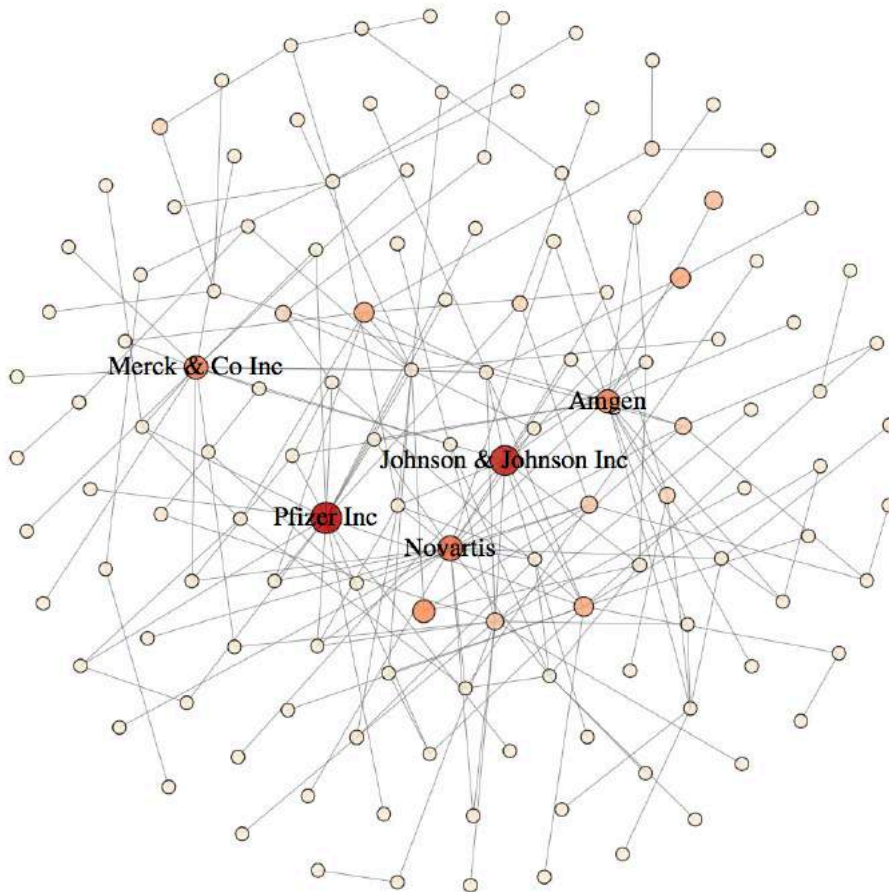


Figure 7: The largest connected component in the observed network of R&D collaborations for firms in the sector SIC-28 in the year 2006. The shade and size of a node indicates its R&D expenditures. The five largest firms in terms of their R&D expenditures are mentioned in the graph.

### 3. Empirical Study

#### 3.1. Data

To get a comprehensive picture of R&D alliances we use data of interfirm R&D collaborations stemming from two sources which have been widely used in the literature [Schilling, 2009].<sup>28</sup> 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.<sup>29</sup> 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].<sup>30</sup> The systematic collection

<sup>28</sup>Note however, that these databases have not been merged and jointly used in any previous empirical study except König et al. [2014b].

<sup>29</sup>For a comparison and summary of different alliance databases, including CATI and SDC, see Schilling [2009].

<sup>30</sup>See <https://sites.google.com/site/patentdatapoint>. We would like to thank Enghin Atalay and Ali

of inter-firm alliances in CATI started in 1987 and ended in 2006. As the CATI database only includes collaborations up to the year 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 3.1](#).

The combined CATI-SDC database only provides the names for each firm in an alliance. We therefore 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, to obtain information about their balance sheets and income statements [see e.g., [Bloom et al., 2013](#)]. For the purpose of matching firms across databases, we employ the above mentioned name matching algorithm. We could match roughly 25% of the firms in the alliance data for which balance sheet information was available.<sup>31</sup> From our match between the firms' names in the alliance database and the firms' names in the Compustat and Orbis databases, we obtained a firm's R&D expenditures, sales, employment, primary industry code, and location. We computed firm's labor productivity as sales relative to the number of employees.<sup>32,33</sup>

To be consistent with our theoretical model (see [Section 2](#) and supplementary [Appendix C](#)) where a firm's output is proportional to its R&D effort, we use the logarithm of firm's R&D expenditure to measure its R&D effort and thus the output. We further identify the patent portfolios of the firms in our dataset using the EPO Worldwide Patent Statistical Database (PATSTAT) [[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](#)]. We obtained matches for roughly 30% of the firms in the data. The technology classes were identified using the main international patent classification (IPC) numbers at the 4-digit level. After removing firm observations with missing values on sales, employment, and R&D expenditures, we regard the remaining 1,201 firms (with 428 R&D collaborations) as the full sample for our analysis.<sup>34</sup> Some descriptive statistics of this full sample are shown in the first row of [Table 1](#).

The size of the sample including all sectors is somewhat too large to make some of the estimation methods that we will introduce below become computationally infeasible (in particular the exchange algorithm and adaptive exchange algorithm in [Sections 3.3.2](#) and [3.3.3](#)),<sup>35</sup> we fur-

---

Hortacsu for sharing their name matching algorithm with us.

<sup>31</sup>Note that for many small private firms no balance sheet information is available, and hence these firms could not be matched by our algorithm. We therefore typically exclude smaller private firms from our analysis, but this is inevitable if one is going to use market value data. Nevertheless, R&D is mostly concentrated in publicly listed firms, which cover most of the R&D activities in the economy [see e.g., [Bloom et al., 2013](#)], and these firms are typically included in our sample.

<sup>32</sup>Labor productivity can vary across firms in different industries, in particular when industries differ greatly in their labor intensity of production. However, these difference should have a limited effect as we include industry fixed effects.

<sup>33</sup>We find that our results remain robust when using output-labor productivity instead of revenue-labor productivity. The results can be obtained upon request from the authors.

<sup>34</sup>To understand the impact of missing observations due to incomplete matching between databases, we conduct a Monte Carlo simulation study in supplementary [Appendix H](#).

<sup>35</sup>See [Section 3.3](#) and supplementary [Appendix H](#) for a more detailed discussion of the various estimation procedures that we introduce in this paper. Further, note that the restriction to small sample sizes does not apply to the likelihood partition approach introduced in [Section 3.3.1](#), for which we can use the full sample to estimate the parameters of the model. We find, however, that the estimation results are similar, irrespective of whether we use only a subsample of the data or the whole sample. Compared to several existing studies, [e.g. [Badev, 2013](#); [Hsieh](#)

Table 1: Descriptive statistics.

Sample	# of firms	Log R&D Expenditure			Productivity			Log # of Patents		
		mean	min	max	mean	min	max	mean	min	max
Full	1201	9.6496	2.5210	15.2470	1.6171	0.0002	20.2452	4.9320	0.0000	11.8726
SIC-28	351	9.6416	3.2109	15.2470	1.3385	0.0002	10.1108	4.7711	0.0000	11.8014
SIC-281	27	9.5288	7.5464	11.2266	2.0951	0.8124	4.5133	6.9610	2.3026	9.9499
SIC-282	22	10.1250	7.5123	12.1022	2.4637	0.1667	5.7551	6.7015	2.9957	10.3031
SIC-283	259	9.4797	3.2109	15.2470	1.0326	0.0002	6.5232	4.1962	0.0000	10.8752
SIC-284	12	11.0216	8.7933	13.2439	1.4869	0.6021	2.6405	7.7903	3.9890	10.9748
SIC-285	5	11.0548	9.8144	13.2205	1.5160	1.2591	1.7099	8.4910	7.1325	10.3017
SIC-286	8	9.3278	6.0924	11.3144	3.9443	1.1249	10.1108	3.6924	0.6931	6.6174
SIC-287	8	8.8004	6.1510	12.8862	1.8069	0.0672	2.7076	3.9510	0.6931	10.6792
SIC-289	10	9.0683	6.2913	10.5094	1.5494	0.0760	2.9324	5.3012	0.6931	9.8807

Note: The logarithm of a firm’s R&D expenditures (by thousand dollars) measures its R&D effort. A Firm’s productivity is measured by the ratio of sales to employment. The logarithm of the number of patents is used as a control variable in the linking cost function [Hanaki et al., 2010].

ther consider a subsample of the data where we can apply all of our estimation strategies. The subsample we consider in the following is restricted to firms in the SIC-28 sector, “chemicals and allied products”. It has the largest number of within sectoral R&D collaborations, as well as the smallest percentage of R&D collaborations to other 2-digit SIC sectors. The SIC-28 sector contains eight sub-sectors, ranging from “industrial inorganic chemicals” (SIC-281) to “miscellaneous chemical products” (SIC-289). The summary statistics of these sectors are provided in Table 1 and the number of R&D collaborations across major 2-digit SIC sectors (with more than twenty firms) and within the SIC-28 sectors are shown in Figure 8. Among the eight sub-sectors within SIC-28, the drugs development sector (SIC-283) is the largest one. It consists of 259 firms and 121 within sector R&D collaborations. In Appendix I we check the robustness of our results from the second largest 2-digit sector, the SIC-36 sector, “electronic and other electric equipment”. SIC-36 contains 198 firms and 29 R&D collaborations. We find the results from SIC-28 and SIC-36 are generally comparable.

### 3.2. Firm Heterogeneity

To account for the firm level heterogeneity that we observe in the data, we extend the profit function of Equation (1) in Section 2 to accommodate heterogeneous marginal costs of production, substitution effects, and heterogeneous technology spillovers (see also Section 2.4), so that the profit of firm  $i$  can be written as follows:

$$\pi_i(\mathbf{q}, G) = \eta_i q_i - \frac{1}{2} q_i^2 - b \sum_{j \neq i}^n b_{ij} q_j q_i + \rho \sum_{j=1}^n f_{ij} a_{ij} q_j q_i - \zeta_i(G). \quad (23)$$

As compared to Equation (1), in Equation (23) we have normalized  $\nu$  to 1/2. The term  $\eta_i$  represents an individual fixed effect for each firm and we capture it by  $\mathbf{X}_i \boldsymbol{\delta}$ , where  $\mathbf{X}_i$  includes firm  $i$ ’s productivity and a sector dummy (at the 2-digit SIC level). To allow for additional heterogeneity among firms, the substitution effect is considered at the sectoral level. Each firm faces a substitution effect from all other firms within the same 3-digit SIC sector, i.e., we set  $b_{ij} = 1$ , if both firms  $i$  and  $j$  are in the same 3-digit SIC sector, and zero otherwise. For the

---

and Lee, 2013; Mele, 2017] which analyze networks with a few hundred nodes, the size of one single network that we handle in this paper is significantly larger.

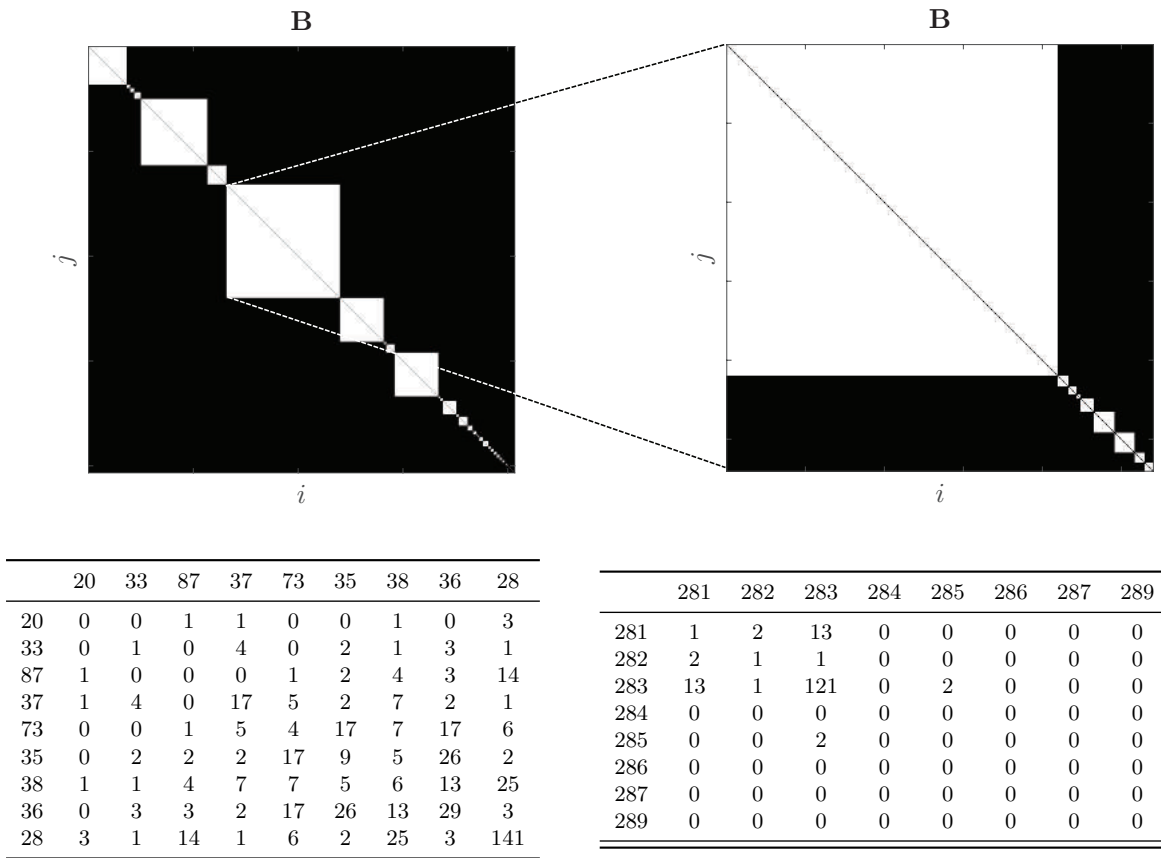


Figure 8: (Top left panel) The empirical competition matrix  $\mathbf{B}$  across all 2-digit SIC sectors. The largest sector is the SIC-28 sector with 351 firms, which comprises 29.22% of all firms in the sample. (Top right panel) The empirical competition matrix  $\mathbf{B}$  across all 3-digit SIC sectors within the SIC-28 sector. The largest sector is the SIC-283 “drugs” sector with 259 firms, which comprises 73.78% of all firms in the SIC-28 sector. (Bottom left panel) The number of R&D collaborations across all 2-digit SIC sectors. The sector SIC-28 has 141 within sector R&D collaborations. (Bottom right panel) The number of R&D collaborations within the sector SIC-28. The sector SIC-283 has 121 within sector R&D collaborations.

purpose of illustration, in Figure 8 we show the empirically observed sectoral competition matrix,  $\mathbf{B}$ , where  $\mathbf{B} = (b_{ij})_{1 \leq i, j \leq n}$  with  $b_{ii} = 0$ , among all 2-digit SIC sectors and among all 3-digit sectors within the SIC-28 sector. In Equation (23) we have further introduced the symmetric weights  $(f_{ij})_{1 \leq i, j \leq n}$ , with  $f_{ij} = f_{ji}$ , to capture heterogeneous technology spillovers across firms, based on the technological proximities of firms  $i$  and  $j$  measured either by Jaffe’s or the Mahalanobis patent similarity indices [Bloom et al., 2013; Jaffe, 1989] criteria (see the supplementary Appendices E.2 and F.4 for further details).

The total cost of R&D collaborations for firm  $i$  is captured by the term  $\zeta_i(G) = \sum_{j=1}^n a_{ij}(\psi_{ij} - \varphi_{ij})$ , with the pairwise symmetric functions  $\psi_{ij} = \gamma^\top \mathbf{c}_{ij}$  and  $\varphi_{ij} = \varkappa t_{ij}$ . The  $r$ -dimensional vector of dyadic-specific variables,  $\mathbf{c}_{ij}$ , represents measures of similarity between firms  $i$  and  $j$  regarding sector, location, technology, research quality, etc., that might have an effect on the collaboration costs (see e.g., Lychagin et al. [2016] and supplementary Appendix E.1 for a simple example). The term  $t_{ij} = \sum_{k \neq i, j}^n a_{ik} a_{jk}$  counts the number of common collaborators shared by firms  $i$  and  $j$  (“cyclic triangles” effect). It allows for R&D collaborations to be less costly between firms that have mutual third-party collaborators [Hanaki et al., 2010].<sup>36</sup>

<sup>36</sup>Hanaki et al. [2010] argue that the existence of mutual collaborators may enhance the effectiveness of penalties and improve the appropriability of the outcomes of joint R&D projects, and that firms can use pre-existing collaborations as conduits of information about the reliability of potential collaboration partners. The authors further find empirical evidence that R&D collaborations are more likely to be undertaken between firms that have mutual third-party collaborators.

The potential function  $\Phi: \mathbb{R}_+^n \times \mathcal{G}^n \rightarrow \mathbb{R}$  corresponding to Equation (23) is given by<sup>12</sup>

$$\Phi(\mathbf{q}, G) = \sum_{i=1}^n \left( \eta_i q_i - \frac{1}{2} q_i^2 \right) - \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n b_{ij} q_i q_j + \frac{\rho}{2} \sum_{i=1}^n \sum_{j \neq i}^n f_{ij} a_{ij} q_i q_j - \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n a_{ij} \psi_{ij} + \frac{1}{3} \sum_{i=1}^n \sum_{j \neq i}^n a_{ij} \varphi_{ij}. \quad (24)$$

In the vector-matrix form this is

$$\Phi(\mathbf{q}, G) = \boldsymbol{\eta}^\top \mathbf{q} - \frac{1}{2} \mathbf{q}^\top \mathbf{M}(G) \mathbf{q} - \frac{1}{2} \text{tr}(\mathbf{A} \boldsymbol{\Psi}^\top) + \frac{1}{3} \text{tr}(\mathbf{A} \boldsymbol{\varphi}^\top), \quad (25)$$

where  $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^\top$ ,  $\boldsymbol{\Psi} = (\psi_{ij})_{1 \leq i, j \leq n}$  and  $\boldsymbol{\varphi} = (\varphi_{ij})_{1 \leq i, j \leq n}$ . In the following, we denote by  $\mathbf{M}(G) \equiv \mathbf{I}_n + b\mathbf{B} - \rho(\mathbf{A} \circ \mathbf{F})$ , where  $\mathbf{F}$  is the matrix with elements  $f_{ij}$  for  $1 \leq i, j \leq n$ , and  $\circ$  denotes the Hadamard element wise matrix product.<sup>37</sup> The stationary distribution of the Markov process of Definition 1 is then given by the Gibbs measure  $\mu^\vartheta(\mathbf{q}, G)$  of Equation (7) in Theorem 1 with the potential function  $\Phi(\mathbf{q}, G)$  of Equation (25).

### 3.3. Exponential Random Graph Models

When both the quantities produced,  $\mathbf{q}$ , and the network,  $G$ , are endogenous, the stationary distribution  $\mu^\vartheta(\mathbf{q}, G)$  is determined by Equation (7). The parameters of the model can be summarized by  $\boldsymbol{\theta} = (\rho, b, \boldsymbol{\delta}^\top, \boldsymbol{\gamma}^\top, \varkappa) \in \Theta$  with parameter space  $\Theta$ .<sup>38</sup> This empirical model belongs to the family of exponential random graph models (ERGMs) or  $p^*$  models [see Frank and Strauss, 1986]. The closed form expression of the likelihood function given in Equation (7) establishes a straightforward channel to check identification of the parameter vector  $\boldsymbol{\theta}$ . Following the theory of exponential family distributions, identification of the parameters  $\boldsymbol{\theta}$  is guaranteed as long as the regressors in  $\Phi(\mathbf{q}, G)$  of Equation (25) are not linearly dependent [Badev, 2013; Mele, 2017; Lehmann and Casella, 2006]. ERGMs are notorious for the difficulty of estimation due to existence of an ‘‘intractable normalizing constant’’ in the probability likelihood function.<sup>39,40</sup> Using classical estimation methods such as a Maximum likelihood (MLE) approach, one needs to simulate a set of auxiliary networks in order to approximate the intractable normalizing constant (MCMC-MLE) [Badev, 2013]. However, the performance of the MCMC-MLE method is greatly influenced by the choice of initial parameter values,  $\boldsymbol{\theta}^{(0)}$  [Airoldi et al., 2009]. If  $\boldsymbol{\theta}^{(0)}$  is not close enough to the true value, without resorting to a global searching algorithm such as simulated annealing, the method may converge to a sub-optimal solution.

Alternatively, the Bayesian MCMC approach has recently gained more attention on ERGM estimation [see e.g., Hsieh and Lee, 2013; Mele, 2017; Liang, 2010; Snijders, 2002]. The intractable normalizing constant in the likelihood function also makes the standard MCMC algorithm in-

<sup>37</sup>Let  $\mathbf{A}$  and  $\mathbf{B}$  be  $m \times n$  matrices. The Hadamard product of  $\mathbf{A}$  and  $\mathbf{B}$  is defined by  $[\mathbf{A} \circ \mathbf{B}]_{ij} = [\mathbf{A}]_{ij} [\mathbf{B}]_{ij}$  for all  $1 \leq i \leq m, 1 \leq j \leq n$ , i.e. the Hadamard product is simply an element-wise multiplication.

<sup>38</sup>Similar to standard logistic regression frameworks the parameter  $\vartheta$  cannot be separately identified, and we therefore omit it for simplicity.

<sup>39</sup>This corresponds to the denominator of Equation (7) which involves a summation over all networks  $G \in \mathcal{G}^n$ , that is, a sum with  $2^{\binom{n}{2}}$  terms.

<sup>40</sup>Other than the difficulty of estimation, the most basic exponential random graphs are statistically equivalent to an Erdős-Rény random graph in the limit of large  $n$  unless the model contains at least one non-trivial negative network externality effect [Bhamidi et al., 2011; Mele, 2017]. We show in Appendix 2.4 how the introduction of various forms or firm heterogeneity leads to correlated networks with structural properties that differ significantly from an Erdős-Rény random graph.

feasible. The standard MH algorithm to update the parameters from  $\boldsymbol{\theta}$  to  $\boldsymbol{\theta}'$  depends on the acceptance probability,

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}) = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}'^\vartheta(\mathbf{q}, G|\boldsymbol{\theta}'))T_1(\boldsymbol{\theta}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})\mu^\vartheta(\mathbf{q}, G|\boldsymbol{\theta})T_1(\boldsymbol{\theta}'|\boldsymbol{\theta})} \right\}, \quad (26)$$

where  $\pi$  denotes the prior density and  $T_1(\boldsymbol{\theta}'|\boldsymbol{\theta})$  denotes the symmetric proposal density for the independent MH draw, i.e.,  $T_1(\boldsymbol{\theta}' - \boldsymbol{\theta}) = T_1(\boldsymbol{\theta} - \boldsymbol{\theta}')$ . In the above acceptance probability, two normalizing terms in  $\mu^\vartheta(\mathbf{q}, G|\boldsymbol{\theta}')$  and  $\mu^\vartheta(\mathbf{q}, G|\boldsymbol{\theta})$  do not cancel each other and therefore,  $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta})$  cannot be calculated.

In the following subsections, we will focus on the MCMC approach and discuss three strategies to bypass the evaluation of the intractable normalizing term. Among these three strategies, when the cyclic triangles effect is absent from the model, i.e., setting  $\varkappa = 0$ , link independence holds (conditional on output), and we can use a likelihood partition approach (Section 3.3.1), which is generally applicable to large network samples. In a link dependent case, where  $\varkappa \neq 0$ , we will use an exchange algorithm (Section 3.3.2) and an adaptive exchange algorithm (Section 3.3.3). The adaptive exchange algorithm is an extension of the exchange algorithm in order to avoid the local trap problem. In supplementary Appendix H, we conduct a simulation study to demonstrate the consistency of each method and compare their computational costs.

### 3.3.1. Likelihood Partition Approach

In the absence of cyclic triangles effects, i.e., when we set  $\varkappa = 0$ , conditional link independence holds (given the output levels of any pair of firms). The probability of observing a network  $G \in \mathcal{G}^n$ , given an output distribution  $\mathbf{q} \in \mathcal{Q}^n$ , is then determined by the conditional distribution (see Equation (15) and supplementary Appendix E):

$$\mu^\vartheta(G|\mathbf{q}) = \frac{\mu^\vartheta(\mathbf{q}, G)}{\mu^\vartheta(\mathbf{q})} = \prod_{i < j}^n \frac{e^{\vartheta a_{ij}(\rho f_{ij} q_i q_j - \boldsymbol{\gamma}^\top \mathbf{c}_{ij})}}{1 + e^{\vartheta(\rho f_{ij} q_i q_j - \boldsymbol{\gamma}^\top \mathbf{c}_{ij})}}. \quad (27)$$

Such a factorization would not be possible if the linking cost would allow for higher order network effects captured by  $\varkappa \neq 0$ . The marginal distribution of the firms' output levels  $\mathbf{q} \in \mathcal{Q}^n$  for large  $\vartheta$  is given by Equation (14) and supplementary Appendix E:

$$\mu^\vartheta(\mathbf{q}) \approx \left( \frac{2\pi}{\vartheta} \right)^{-\frac{n}{2}} |\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top (-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)) (\mathbf{q} - \mathbf{q}^*) \right\}, \quad (28)$$

where

$$(\Delta \mathcal{H}_\vartheta(\mathbf{q}))_{ii} = -1 + \frac{\vartheta \rho^2}{4} \sum_{j \neq i}^n f_{ij}^2 q_j^2 \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho f_{ij} q_i q_j - \boldsymbol{\gamma}^\top \mathbf{c}_{ij}) \right) \right)^2,$$

and

$$\begin{aligned} (\Delta \mathcal{H}_\vartheta(\mathbf{q}))_{ij} &= -bb_{ij} + \frac{\rho f_{ij}}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho f_{ij} q_i q_j - \boldsymbol{\gamma}^\top \mathbf{c}_{ij}) \right) \right) \\ &\quad \times \left( 1 + \frac{\vartheta \rho f_{ij}}{2} q_i q_j \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho f_{ij} q_i q_j - \boldsymbol{\gamma}^\top \mathbf{c}_{ij}) \right) \right) \right), \end{aligned}$$



for  $j \neq i$ . Further,  $\mathbf{q}^*$  in Equation (28) solves the following system of equations (see Equation (10) and supplementary Appendix E):

$$q_i = \eta_i + \sum_{j \neq i}^n \left( \frac{\rho f_{ij}}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} \left( \rho f_{ij} q_i q_j - \gamma^\top \mathbf{c}_{ij} \right) \right) \right) - b b_{ij} \right) q_j. \quad (29)$$

Equation (28) shows that in the limit of large  $\vartheta$ ,  $\mathbf{q}$  is asymptotically normally distributed with mean  $\mathbf{q}^*$  and variance-covariance matrix  $-\frac{1}{\vartheta} \Delta \mathcal{H}_\vartheta(\mathbf{q}^*)^{-1}$ . It then follows that the likelihood of the network  $G$  and the quantity profile  $\mathbf{q}$  in the large  $\vartheta$  limit can be partitioned as follows [Smyth, 1996]

$$\begin{aligned} \mu^\vartheta(\mathbf{q}, G) &= \mu^\vartheta(G|\mathbf{q}) \cdot \mu^\vartheta(\mathbf{q}) \approx \prod_{i < j}^n \frac{e^{\vartheta a_{ij}(\rho f_{ij} q_i q_j - \gamma^\top \mathbf{c}_{ij})}}{1 + e^{\vartheta(\rho f_{ij} q_i q_j - \gamma^\top \mathbf{c}_{ij})}} \\ &\times \left( \frac{2\pi}{\vartheta} \right)^{-\frac{n}{2}} |\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top (-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)) (\mathbf{q} - \mathbf{q}^*) \right\}, \end{aligned} \quad (30)$$

where we have inserted Equation (27) for  $\mu^\vartheta(G|\mathbf{q})$  and Equation (28) for  $\mu^\vartheta(\mathbf{q})$ . Based on the partition of  $\mu^\vartheta(\mathbf{q}, G)$  in Equation (30), we do not need to evaluate the intractable normalizing constant in the likelihood function of Equation (7), and can estimate the parameters by a standard Bayesian MCMC algorithm. The key step of solving  $\mathbf{q}^*$  from Equation (29) can be implemented efficiently by recognizing it as a fixed point system, and using a fast fixed point algorithm. As supplementary Appendix H shows, this likelihood partition (LP) approach is the least computationally costly compared to the other two methods outlined below. Therefore, we apply the LP approach to estimate the model for both, the full sample including all sectors and the SIC-28 sector subsample.

### 3.3.2. Exchange Algorithm

When we allow for cyclic triangles effects, i.e.,  $\varkappa \neq 0$ , then we can no longer use the LP estimation algorithm from the previous section. An alternative is the exchange algorithm, which provides a way to bypass evaluation of the intractable normalizing constant in the Metropolis-Hastings (MH) acceptance probability [see e.g. Liang et al., 2011]. The name “exchange” is given due to its similarity with the swapping operation of exchange Monte Carlo [Geyer, 1991]. It is different from the conventional MH algorithm by having a proposal density  $T_1(\boldsymbol{\theta}'|\boldsymbol{\theta})\mu(\mathbf{q}', G'|\boldsymbol{\theta}')$ , which involves simulation of auxiliary data  $(\mathbf{q}', G')$  from the distribution  $\mu(\mathbf{q}', G'|\boldsymbol{\theta}')$ . The acceptance probability of the exchange algorithm can be written as

$$\begin{aligned} \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, \mathbf{q}', G') &= \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}')\mu(\mathbf{q}, G|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})\mu(\mathbf{q}, G|\boldsymbol{\theta})} \cdot \frac{T_1(\boldsymbol{\theta}|\boldsymbol{\theta}')\mu(\mathbf{q}', G'|\boldsymbol{\theta})}{T_1(\boldsymbol{\theta}'|\boldsymbol{\theta})\mu(\mathbf{q}', G'|\boldsymbol{\theta}')} \right\} \\ &= \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}'^\Phi(\mathbf{q}, G, \boldsymbol{\theta}'))}{\pi(\boldsymbol{\theta})e^{\Phi(\mathbf{q}', G', \boldsymbol{\theta}')}} \cdot \frac{T_1(\boldsymbol{\theta}|\boldsymbol{\theta}'^\Phi(\mathbf{q}', G', \boldsymbol{\theta}'))}{T_1(\boldsymbol{\theta}'|\boldsymbol{\theta})e^{\Phi(\mathbf{q}, G, \boldsymbol{\theta})}} \right\}. \end{aligned} \quad (31)$$

Intractable normalizing terms in Equation (31) are cancelled out and thus the acceptance probability  $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, \mathbf{q}', G')$  can be computed.

A problem of the exchange algorithm is that it requires a perfect sampler of  $G'$  and  $\mathbf{q}'$  from  $\mu(\cdot|\boldsymbol{\theta}')$  [Propp and Wilson, 1996], which is computationally intense for most of ERGM applications. To overcome this issue, Liang [2010] proposed the double MH (DMH) algorithm to

replace the perfect sampler with a finite MH chain initialized at the observed network. In this paper, we use the DMH algorithm for parameter estimation, and more specific details about its implementation can be found in supplementary Appendix G.1.

To improve the computational efficiency of the DMH algorithm, we assume that during the dynamic network formation process, whenever a firm changes its R&D collaborations, all firms adjust output levels immediately and thus the new output levels will be the profit-maximizing output given by the best response function  $\mathbf{q}^* = \mathbf{M}(G)^{-1}\boldsymbol{\eta}$  plus a stochastic error term.<sup>41</sup> The size of error, according to the approximation derived in supplementary Appendix G.4, should be determined by  $\mathbf{M}(G)^{-1}$ . This assumption means that we impose two different time scales: a *fast time scale* of output adjustments, and a *slow time scale* of link adjustments [Khalil, 2002].<sup>42</sup> When simulating auxiliary data for the output levels and the network, this assumption saves firm’s infinitesimal adjustments on output and turns it into an implicit function of the network. However, we will have to evaluate  $\mathbf{M}(G)^{-1}$  whenever a link has been added or removed from the auxiliary network and this may still be computationally costly. To do this in an efficient way, we use a matrix perturbation result that is derived in supplementary Appendix G.5.

### 3.3.3. Adaptive Exchange Algorithm

Even though the DMH algorithm alleviates the computational burden by replacing perfect sampling, convergence of the finite MH chain in the DMH algorithm is not theoretically guaranteed. Therefore, the DMH estimates are only approximately correct no matter how long the algorithm has been run. Especially, if the network distribution represented by an ERGM is multi-modal, the finite MH run may be trapped at one of the local maxima of the likelihood function (“local trap problem”). Consequentially, the Markov chain of the DMH algorithm may mix very slowly and require unaffordable computation time for achieving convergence [Bhamidi et al., 2011].<sup>43</sup>

In this paper, we therefore use an adaptive exchange (AEX) algorithm, which has recently been developed by Jin et al. [2013] and Liang et al. [2015], to overcome the uncertainty of slow mixing faced by the DMH algorithm. The foundation of the AEX algorithm is an MCMC sampling scheme called stochastic approximation Monte Carlo (SAMC) algorithm [Liang et al., 2007]. The main feature of SAMC is that it applies importance sampling to prevent the local trap problem. In SAMC, the sample space is partitioned into non-overlapping subregions. Different importance weights are assigned to each subregion so that SAMC draws samples from a kind of “mixture distribution” that avoids being trapped by a local extremum. Additionally, SAMC contains a self-adjusting mechanism to the weights of each subregion so that it can escape from local extrema of the likelihood function very quickly.

---

<sup>41</sup>Note that  $\mathbf{q}^*$  maximizes the potential function of (25) for an exogenously given network  $G$ .

<sup>42</sup>Observe that the stationary distribution  $\mu(\mathbf{q}, G)$  in Equation (7) does not depend on the parameter  $\chi$  that governs the speed of the output adjustment. This means that the stationary distribution  $\mu(\mathbf{q}, G)$  is independent of how quickly the output levels adjust. As a consequence, we can use it as a degree of freedom in our estimation algorithms, that is, the stationary distribution is not affected whether we make the time scale separation assumption or not.

<sup>43</sup>To overcome the local trap problem and thus speed up convergence during network simulation, Snijders [2002] and Mele [2017] proposed MH samplers which consist of both local and non-local steps. In the local step, only one random edge is updated in an iteration. With a certain probability, the sampler will implement the non-local step where multiple (or even all) edges are updated in an iteration. Although Mele [2017] used a simulation study to demonstrate convergence of the MH sampler with a non-local step, the simulation result is based on a simple model specification with one indirect link effect and may not be extended to a more general case. Therefore, it is still questionable whether the finite MH runs with non-local steps can always achieve convergence for general ERGMs.

In the AEX algorithm, two Markov chains are running in parallel. In the first chain, we draw auxiliary data  $(\tilde{\mathbf{q}}, \tilde{G})$  from a family of distributions  $\mu(\tilde{\mathbf{q}}, \tilde{G}|\boldsymbol{\theta}_1), \dots, \mu(\tilde{\mathbf{q}}, \tilde{G}|\boldsymbol{\theta}_m)$  with frequencies determined by the SAMC algorithm, where  $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m)$  are pre-specified parameter points. In practice, we set  $m = 50$  and  $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{50})$  are chosen by the Max-Min procedure from the DMH draws [Liang et al., 2015]. We have also tried  $m = 30$  and  $m = 100$  in the simulation and empirical studies and the results are largely similar. In the second chain, we implement the exchange algorithm for updating the target parameters, where auxiliary data,  $\mathbf{q}'$  and  $G'$ , are re-sampled from the first chain by an importance sampling procedure. Convergence of the AEX algorithm, i.e.  $\mathbf{q}'$  and  $G'$  from the importance sampling procedure of the AEX converge to  $\mu(\cdot|\boldsymbol{\theta}')$  and the draws of  $\boldsymbol{\theta}$  from the target chain will converge weakly to the posterior  $\mu(\boldsymbol{\theta}|\mathbf{q}, G)$ , can be shown when the number of iterations of both the auxiliary and target chains go to infinity. Details of implementing the AEX algorithm and the proof of convergence are provided in supplementary Appendices G.2 and G.3.

### 3.4. Empirical Results

Table 2 presents our estimation results for the full sample of all sectors and the subsample restricted to the SIC-28 sector.<sup>44</sup> Due to the aforementioned computational constraints, we only apply the LP algorithm to estimate the full sample, while applying all three estimation algorithms to the subsample of the SIC-28 sector. In the full sample, the result shows that both estimates of the technology spillover parameter  $\rho$  (0.0355) and the competition parameter  $b$  (0.0002) are significant and have the expected signs. The two effects match our theoretical predictions from Section 2, showing that firms face a positive complementary effect from R&D collaborations and a negative substitution effect from competing firms in the same market. Furthermore, we find that the technology spillover effect  $\rho$  is much larger than the product market rivalry effect  $b$  [Bloom et al., 2013]. This suggests that the marginal returns from R&D collaborations are positive even between competing firms. We also find that a higher labor productivity is associated with higher R&D expenditures [see e.g., Cohen et al., 1987]. The estimation results further show significant effects from the control variables in the linking cost function, including the same sector (at 3-digit SIC level) and the same country dummies, and the sum of log patent counts from each of the two firms involved in a collaboration. Results on the two dummies reveal that the R&D collaboration cost would be lower among two firms in the same sector or in the same country. The sum of the log patent numbers acts as a proxy for the research capacities of collaborating firms [Hanaki et al., 2010] and a higher value indicates lower R&D collaboration cost between them.

In the SIC-28 subsample, the estimates are similar across different estimation methods, which shows the robustness of the estimates. Compared to the result of the full sample, there are additional significant effects from the differences in productivities between collaborating firms and its square in the linking cost. This non-monotonic effect from the differences in productivities indicates that R&D collaborations between similar but not too similar firms involve lower collaboration and coordination costs [see e.g. Hanaki et al., 2010; Nooteboom et al., 2007].

In Table 3 we provide additional estimation results by taking into account heterogeneous technology spillover effects among collaborating firms. We also allow for a cyclic triangles effects in the linking cost function, where firms with common collaborators may experience lower col-

---

<sup>44</sup>The estimation result of the SIC-36 sector is available in Appendix I.

Table 2: Estimation results of the full sample and the SIC-28 sector

		Full sample		SIC-28 subsample	
		LP	LP	DMH	AEX
R&D Spillover	$(\rho)$	0.0355*** (0.0008)	0.0386*** (0.0015)	0.0408*** (0.0021)	0.0458*** (0.0010)
Substitutability	$(b)$	0.0002*** (0.0000)	0.0001** (0.0001)	0.0002*** (0.0001)	0.0002*** (0.0000)
Prod.	$(\delta_1)$	0.2099*** (0.0127)	0.4475*** (0.0457)	0.3769*** (0.0509)	0.3787*** (0.0424)
Sector FE	$(\delta_2)$	Yes	Yes	Yes	Yes
.....					
<b>Linking Cost</b>					
Constant	$(\gamma_0)$	13.1415*** (0.1336)	13.2627*** (0.3507)	14.4023*** (1.1547)	14.3366*** (0.1180)
Same Sector	$(\gamma_1)$	-2.1458*** (0.1053)	-1.9317*** (0.2551)	-1.9648*** (0.5749)	-1.8579*** (0.3972)
Same Country	$(\gamma_2)$	-0.8841*** (0.1030)	-0.4186*** (0.1591)	-0.6359* (0.3903)	-0.6555*** (0.1907)
Diff-in-Prod.	$(\gamma_3)$	0.0231 (0.0554)	-1.2698*** (0.2937)	-1.4300** (0.6450)	-1.3255*** (0.1436)
Diff-in-Prod. Sq.	$(\gamma_4)$	-0.0014 (0.0044)	0.3276*** (0.0876)	0.4023** (0.1910)	0.4505*** (0.0563)
Patents	$(\gamma_5)$	-0.0943*** (0.0053)	-0.0783*** (0.0150)	-0.1176** (0.0562)	-0.0410** (0.0210)
Sample size		1,201		351	

Note: The dependent variable is log R&D expenditures. The parameters  $\theta = (\rho, b, \delta^\top, \gamma^\top, \varkappa)$  correspond to Equation (24), where  $\psi_{ij} = \gamma^\top \mathbf{c}_{ij}$  and  $\eta_i = \mathbf{X}_i \delta$  (see Section 3.2). We make 50,000 MCMC draws where we drop the first 2,000 draws during a burn-in phase and keep every 20th of the remaining draws 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 MCMC draws for  $\rho$  and  $b$  are shown in Figure 1 in supplementary Appendix G. The asterisks \*\*\*(\*\*, \*) indicate that its 99% (95%, 90%) highest posterior density range does not cover zero.

Table 3: Homogeneous versus heterogeneous spillovers

		Homogeneous		Jaffe		Mahalanobis	
		DMH	Logit	DMH	Logit	DMH	Logit
R&D Spillover	$(\rho)$	0.0396*** (0.0019)	0.0356*** (0.0030)	0.0524*** (0.0090)	0.0070 (0.0042)	0.0275*** (0.0042)	0.0038** (0.0019)
Substitutability	$(b)$	0.0002*** (0.0001)	-	0.0001*** (0.0001)	-	0.0001*** (0.0001)	-
Prod.	$(\delta_1)$	0.3696*** (0.0526)	-	0.4367*** (0.0556)	-	0.4372*** (0.0612)	-
Sector FE	$(\delta_2)$	Yes	-	Yes	-	Yes	-
<hr/>							
<b>Linking Cost</b>							
Constant	$(\gamma_0)$	13.5645*** (0.6067)	12.8064*** (0.5075)	13.5182*** (0.2966)	11.4667*** (0.4764)	14.3226*** (0.5195)	11.4501*** (0.4859)
Same Sector	$(\gamma_1)$	-2.0559*** (0.4247)	-1.7129*** (0.2681)	-1.8892*** (0.3261)	-2.0271*** (0.2547)	-2.8818*** (0.7106)	-2.0253*** (0.2609)
Same Country	$(\gamma_2)$	-0.3782 (0.3267)	-0.3677** (0.1781)	-0.6871*** (0.3082)	-0.4679*** (0.1740)	-0.9134*** (0.3905)	-0.4674*** (0.1669)
Diff-in-Prod.	$(\gamma_3)$	-0.8575* (0.3881)	-1.2679*** (0.3116)	-3.3302*** (0.4379)	-1.3288*** (0.2981)	-3.1080*** (0.6717)	-1.3145*** (0.3106)
Diff-in-Prod. Sq.	$(\gamma_4)$	0.2655** (0.1270)	0.3046** (0.0936)	0.9665*** (0.1916)	0.3187*** (0.0889)	0.9984*** (0.2880)	0.3167*** (0.0929)
Patents	$(\gamma_5)$	-0.0909** (0.0449)	-0.0384 (0.0295)	-0.2128*** (0.0336)	-0.2340*** (0.0269)	-0.1957*** (0.0534)	-0.2310*** (0.0270)
Cyclic Triangles	$(\varkappa)$	-1.6277*** (0.4095)	-1.5486*** (0.1753)	-3.5815*** (0.3898)	-2.2637*** (0.1587)	-3.0555*** (0.4338)	-2.2509*** (0.1537)

Note: The dependent variable is log R&D expenditures. The parameters  $\theta = (\rho, b, \delta^\top, \gamma^\top, \varkappa)$  correspond to Equation (24), where  $\psi_{ij} = \gamma^\top \mathbf{c}_{ij}$ ,  $\varphi_{ij} = \varkappa t_{ij}$  and  $\eta_i = \mathbf{X}_i \delta$  (see Section 3.2). The estimation results are based on 351 firms from the SIC-28 sector. We make 50,000 MCMC draws where we drop the first 2,000 draws during a burn-in phase and keep every 20th of the remaining draws 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 its 99% (95%, 90%) highest posterior density range does not cover zero. Heterogeneous spillovers are captured by the technological proximity matrix with elements  $f_{ij}$  using either the Jaffe or the Mahalanobis patent proximity metrics [Bloom et al., 2013; Jaffe, 1989].

laboration costs [Hanaki et al., 2010]. In specifying heterogeneous technology spillovers, R&D collaborations are weighted by different technology proximity measures. More precisely, in Table 3 from left to right we consider as a benchmark a homogenous R&D collaboration matrix (i.e., assuming all weights are set to one), a matrix weighted by the technological proximity measure introduced by Jaffe [1989], and a matrix weighted by the Mahalanobis technological proximity (see Bloom et al. [2013] and supplementary Appendix F.4 for more details), respectively.<sup>45</sup> Each model specification is estimated using the DMH algorithm outlined in Section 3.3.2 based on the subsample of firms in the SIC-28 sector. We are then able to determine potential variations of the estimated spillover effects due to alternative weights for the different technology proximity measures. Although the Jaffe and Mahalanobis measures are highly correlated (see the supplementary Appendix F.4), the estimated spillover coefficient based on the Jaffe measure (0.0524) is larger than the one based on the Mahalanobis measure (0.0275), and the homogeneous spillover coefficient (0.0396) lies between them.

Further note that the cyclic triangles effect in the linking cost function for each model specification breaks the link independence condition (even conditional on output). Thus, in Table 3 we also compare the legitimate estimate from the DMH algorithm with the naive estimate of a

<sup>45</sup>We do not impose any row-normalization on these weighted R&D collaboration matrices.

logit model conditional on output. The DMH results show that the cyclic triangles effect is large and significant, indicating that having mutual third-party collaborations effectively reduces the R&D collaboration cost. The results from the logit model show clear biases on the estimate of technology spillover effect, particularly in the heterogeneous spillovers cases, and other estimates in the linking cost function.

### 3.5. Model Fit

In order to investigate whether the network formation model that we propose fits the observed network data, we adopt a model goodness-of-fit examination following [Hunter et al. \[2008\]](#). We take the observed network data from the full sample of all sectors. Then we simulate one hundred artificial networks from our network formation model with parameters reported under the LP algorithm in [Table 2](#). Model fitness is examined by the similarity between simulated networks and observed networks in the distribution of four network statistics – degree, edge-wise shared partner, minimum geodesic distance, and average nearest neighbor connectivity.<sup>46</sup> The examination results are shown in [Figure 9](#). We present the distribution of statistics for the observed network by solid curves, distributions for simulated networks by box plots and the 5<sup>th</sup> and 95<sup>th</sup> percentiles by dotted lines. From the figure we find that the simulated networks and the observed network display similar distributions over these four statistics. This suggests that our estimated model is able to simulate the unobserved network generating process.

## 4. Counterfactual Analyses

With our estimates from the previous section ([Table 2](#)) we are now able to perform various counterfactual studies. The first, discussed in [Section 4.1](#), studies the impact on welfare of a firm exiting from the network. The second, discussed in [Section 4.2](#), analyzes the welfare impact of a merger between firms in the same sector. The third policy intervention, discussed in [Section 4.3](#), studies the welfare impact of a subsidy on the collaboration costs between pairs of firms, and aims at identifying the pair for which the subsidy yields the highest welfare gains. In all these counterfactual scenarios the output levels of the firms and the links between them are fully endogenous and respond to changes in the network or the parameters. Further, to evaluate the impact of any intervention under consideration we take a long run perspective and analyze welfare in the new stationary state, after the policy has been implemented.

### 4.1. Firm Exit and Key Players

In this section we evaluate the expected welfare loss from the exit of a firm from the network. The exit of a firm could either be due to financial reasons, such as the recession experienced by the American automobile manufacturing industry during the global financial downturn, or legal reasons, such as the recent emission-fraud scandal of Volkswagen. In the former case, policy

---

<sup>46</sup>See supplementary [Appendix B](#) for the definition of various network statistics. Moreover, the edge-wise shared partner statistic contains important information of a network, for example, the count of triangles in a network  $G$  is a function of  $G$ . Its distribution consists of values  $EP_G(0)/E_G, \dots, EP_G(m-2)/E_G$ , where  $EP_G(k)$  denotes the number of edges whose endpoints both share edges with exactly  $k$  other individuals and  $E_G$  is the total number of edges in network  $G$ . The minimum geodesic distance is a higher-order network statistics. Its distribution consists of the proportions of the possible values of geodesic distances between two nodes [[Hunter et al., 2008](#)].

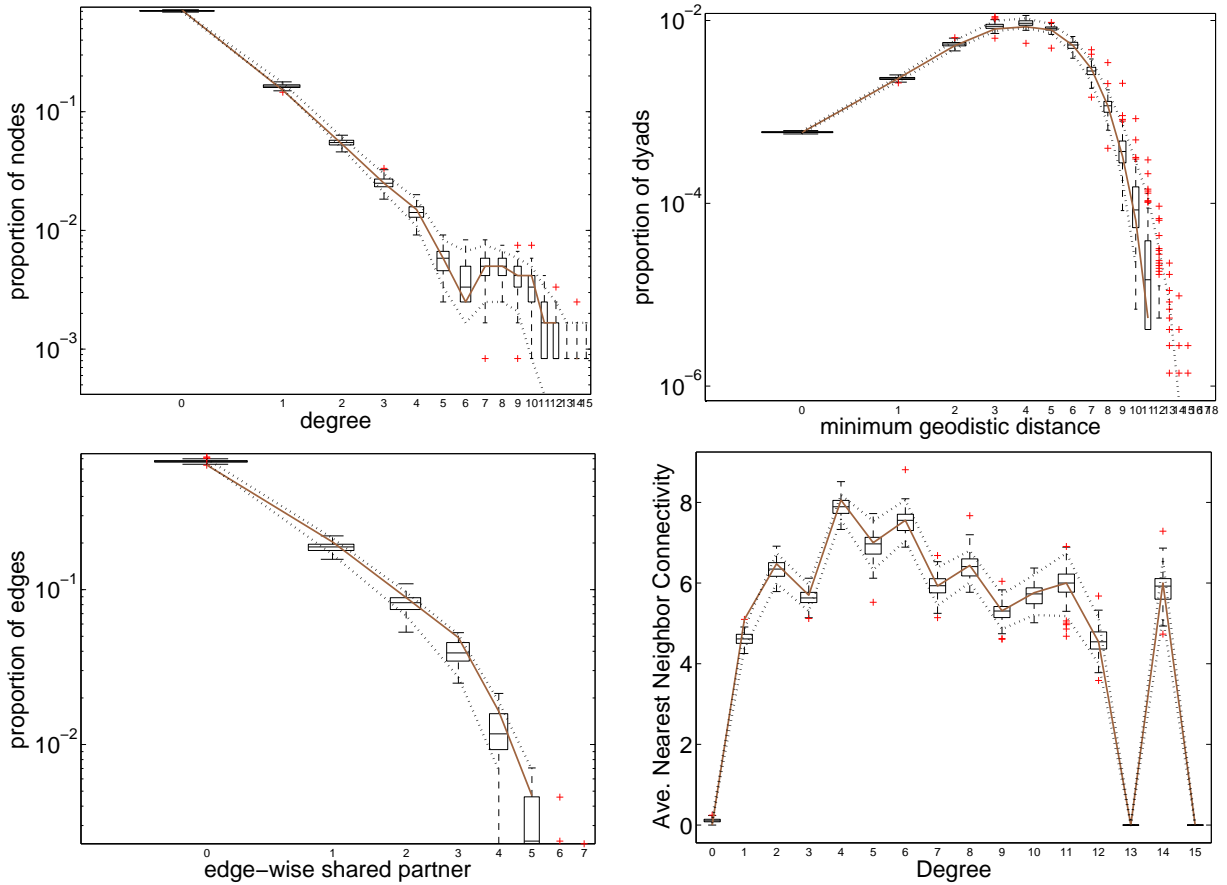


Figure 9: Goodness-of-fit statistics.

makers want to know the overall welfare gains of “bailing out” a bankrupting firm, while, in the latter case, policy makers want to know the overall cost they impose on the economy by inflicting high penalties that might threaten the continued existence of a firm.

The firm whose exit results in the highest expected welfare loss is termed the “key player” [Zenou, 2015]. This counterfactual analysis is related to Ballester et al. [2006], who perform a key player analysis where agents are ranked according to the reduction in aggregate output when they are removed from the network, and König et al. [2014b] who do this for the reduction in welfare similar to our setup. However, while these authors have assumed that the network is exogenously given and does not adapt to the exit of a firm, here we can relax this assumption and allow the network to reconfigure endogenously after the removal of a firm. Formally, the *key firm* is defined as

$$i^* = \operatorname{argmin}_{i \in \mathcal{N}} \left\{ \sum_{G \in \mathcal{G}^{n-1}} \int_{\mathcal{Q}^{n-1}} W_{-i}(\mathbf{q}, G) \mu^\vartheta(\mathbf{q}, G) d\mathbf{q} \right\}, \quad (32)$$

where the probability measure  $\mu^\vartheta(\mathbf{q}, G)$  is given by Equation (7), the potential function from Equation (25), the welfare function  $W(\mathbf{q}, G)$  is defined in Equation (19) and  $W_{-i}(G, \mathbf{q})$  denotes the welfare function with firm  $i$  removed from the network.

We proceed by removing each firm from the network one at a time. Using the estimated model from Table 2, we then simulate the network evolution for the remaining  $n - 1$  firms. We use the same graph simulation algorithm that is part of the DMH estimation algorithm, which is explained in greater detail in supplementary Appendix G.1. We run the simulation for  $10^4$

Table 4: Key player ranking for firms in the chemicals and allied products sector (SIC-28).

Firm	Mkt. Sh. [%] <sup>a</sup>	Patents	Degree	$\Delta W$ [%] <sup>b</sup>	$\Delta W_F$ [%] <sup>c</sup>	$\Delta W_N$ [%] <sup>d</sup>	SIC	Rank
Pfizer Inc.	2.7679	78061	15	-1.8764	-1.7943	-0.3843	283	1
Novartis	2.0691	18924	15	-1.7369	-1.8271	-0.3273	283	2
Amgen	0.8193	6960	13	-1.6272	-1.4240	-0.4753	283	3
Bayer	3.8340	133433	10	-1.3781	-1.2910	-0.3445	280	4
Merck & Co. Inc.	1.2999	52847	10	-1.0182	-1.1747	-0.2892	283	5
Dyax Corp.	0.0007	227	6	-0.7709	-0.6660	-0.3289	283	6
Medarex Inc.	0.0028	168	9	-0.7452	-0.8749	-0.3847	283	7
Exelixis	0.0057	58	7	-0.7293	-0.8603	-0.3686	283	8
Xoma	0.0017	648	7	-0.6039	-0.6863	-0.2254	283	9
Genzyme Corp.	0.1830	1116	3	-0.5904	-0.2510	-0.2987	283	10
Johnson & Johnson Inc.	3.0547	1212	7	-0.5368	-0.8556	-0.3520	283	11
Abbott Lab. Inc.	1.2907	11160	3	-0.5162	-0.1867	-0.3543	283	12
Infinity Pharm. Inc.	0.0011	44	4	-0.4623	-0.5155	-0.2724	283	13
Curagen	0.0023	174	3	-0.4335	-0.4388	-0.3742	283	14
Cell Genesys Inc.	0.0001	236	5	-0.4133	-0.4629	-0.2450	283	15
Solvay SA	1.2445	22689	3	-0.4048	-0.3283	-0.2480	280	16
Takeda Pharm. Co. Ltd.	0.6445	19460	7	-0.3934	-0.7817	-0.3818	283	17
Daiichi Sankyo Co. Ltd.	0.4590	14	5	-0.3691	-0.5581	-0.3377	283	18
Maxygen	0.0014	252	3	-0.3455	-0.3013	-0.2268	283	19
Compugen Ltd.	0.0000	246	5	-0.3130	-0.5251	-0.3202	283	20

<sup>a</sup> Market share in the primary 3-digit SIC sector in which the firm is operating.

<sup>b</sup> The relative welfare loss due to exit of a firm  $i$  is computed as  $\Delta W = (\mathbb{E}_{\mu^\vartheta} [W_{-i}(\mathbf{q}, G)] - W(\mathbf{q}^{\text{obs}}, G^{\text{obs}})) / W(\mathbf{q}^{\text{obs}}, G^{\text{obs}})$ , where  $\mathbf{q}^{\text{obs}}$  and  $G^{\text{obs}}$  denote the observed R&D expenditures and network, respectively.

<sup>c</sup>  $\Delta W_F$  denotes the relative welfare loss due to exit of a firm assuming a fixed network of R&D collaborations.

<sup>d</sup>  $\Delta W_N$  denotes the relative welfare loss due to exit of a firm in the absence of a network of R&D collaborations.

iterations,<sup>47</sup> use the observation of the last iteration as the simulation outcome, and calculate the corresponding welfare value. We then repeat this procedure 100 times and report the average welfare value.

The results for the key player analysis focusing on the SIC-28 sector can be seen in Table 4. In column six ( $\Delta W_F$ ) we consider the case of an exogenously fixed network, while in column five ( $\Delta W$ ) we allow the network to dynamically adjust to the exit of a firm. For the key firm, *Pfizer Inc.*, a U.S. based global pharmaceutical corporation which is among the world’s largest pharmaceutical companies, the reduction in welfare due to its exit amounts to 1.79% when we assume that the network is fixed, and 1.88% when we allow for an adaptive network. Thus, network adaptivity and endogeneity amplify the impact of firm exit (at least for the most important firms), so that studies that assume an exogenous network typically underestimate the effect of the exit of a firm. Moreover, in column seven ( $\Delta W_N$ ) in Table 4 we observe that the welfare loss ignoring the network structure altogether (0.38%) is much lower than the welfare loss incurred in the presence of the R&D collaboration network.

Table 4 also illustrates that the most important firms are not necessarily the ones with the highest market share, number of patents or degree, nor can they be identified with standard centrality measures in the literature (such as the betweenness or eigenvector centrality; see [Wasserman and Faust \[1994\]](#)). Rather, our results illustrate that in order to identify the key firms that are systemically relevant, we need to consider not only the market structure but also the spillovers generated through a network of R&D collaborations, and this network must be allowed to dynamically adjust upon the exit of a firm.

<sup>47</sup>We also tried  $1.5 \times 10^4$  and  $2 \times 10^4$  iterations and get similar results.



## 4.2. Mergers and Acquisitions

Our framework can also be used to study the counterfactual scenario for mergers and acquisitions in innovative industries, and their impact on welfare [Farrell and Shapiro, 1990; Salant et al., 1983]. Market concentration indices are not adequate to correctly account for the network effects of a merger on welfare [see e.g. Encaoua and Hollander, 2002]. This is because the effect of a merger of two firms on industry profits, consumer surplus, and overall welfare depends not only on the market structure (as for example in Bimpikis et al. [2014]), but also on the R&D spillovers in the network which might lead to R&D-efficiency gains from concentration. Due to these spillovers, mergers and the increased concentration they generate (both, in terms of the product market and the collaboration network) can be good or bad for welfare [Daughety, 1990], depending on the identities of the involved firms and their positions in the market as well as the network structure. This counterfactual analysis is therefore potentially important for antitrust policy makers. For example, in 2014, more than half of the merger proposals that were investigated by the U.S. Department of Justice involved R&D-efficiency claims [Marshall and Parra, 2015].

Based on our model we can assess the impact of a merger between two firms  $i$  and  $j$  which are competing in the same market. The merger that results (in expectation) in the greatest reduction in welfare is defined as

$$(i, j)^* = \underset{(i, j) \in \mathcal{N} \times \mathcal{N}}{\operatorname{argmin}} \left\{ \sum_{G \in \mathcal{G}^{n-1}} \int_{\mathcal{Q}^{n-1}} W_{i \cup j}(\mathbf{q}, G) \mu^\vartheta(\mathbf{q}, G) d\mathbf{q} \right\}, \quad (33)$$

where the probability measure  $\mu^\vartheta(\mathbf{q}, G)$  is given by Equation (7), the potential function from Equation (25), the welfare function  $W(\mathbf{q}, G)$  is defined in Equation (19) and  $W_{i \cup j}(\mathbf{q}, G)$  denotes the welfare function with firms  $i$  and  $j$  being merged to a single firm  $k$  in the network  $G$ . That is, two incident nodes,  $i$  and  $j$  in  $G$ , are merged into a new node  $k$ , where each of the edges incident to  $k$  correspond to an edge incident to either  $i$  or  $j$ . In a similar way, the merger that results (in expectation) in the greatest increase in welfare can be defined.<sup>48</sup>

Using the same simulation procedure as in Section 4.1, the merger rankings (for both, welfare gains and losses) can be found in Table 5. To prevent ambiguity on the sector identity of the merged firms, we restrict mergers to firms within the same 3-digit SIC sectors. The ranking for a fixed network in column ten ( $\Delta W_F$ ) in Table 5 is based on the assumption that the network does not adjust to the merger and can be interpreted as a short run analysis. In contrast, the ranking in column nine ( $\Delta W$ ) is based on an endogenous network which adapts to a merger and can be interpreted as a long run analysis. The mergers resulting in the highest reduction or gain in welfare involve only firms in the drugs development sector (SIC 283), which is also the largest sector in our sample.

The highest loss in the endogenous network case is triggered by a merger between *Daiichi Sankyo Co. Ltd.*, a global pharmaceutical company and the second largest pharmaceutical company in Japan, and *Schering-Plough Corp.*, a U.S.-based pharmaceutical company, yielding a reduction in welfare of 0.6%. The relatively large market shares of these companies amplify the

---

<sup>48</sup>We note that we only consider mergers between firms in the same market. We also do not consider firms operating in multiple markets simultaneously, such as for example in Bimpikis et al. [2014]. However, we believe that our analysis of a very specific type of a merger can generate useful insights because it takes into account multiple sources of externalities that are typically ignored in other studies.

negative effect of the merger. Strikingly, assuming a fixed network ( $\Delta W_F$ ) would predict a weak welfare gain from the merger, while ignoring the network structure altogether ( $\Delta W_N$ ) would predict a smaller welfare loss. This highlights the importance of taking into account the endogeneity of the network. Further, the welfare loss due to a merger is typically lower than the welfare loss due to firm exit. In contrast, the highest welfare gain from a merger is obtained between *Isis Pharm. Inc.* and *Takeda Pharm. Co. Ltd.*, two large multinational pharmaceutical companies. In the endogenous case ( $\Delta W$ ) the welfare gain is 0.86%, while in the exogenous case ( $\Delta W_F$ ) it is only 0.34%. This indicates the importance of dynamic network effects in the industry.

Further, comparing the firms involved in mergers that lead to welfare gains in Table 5, as opposed to the ones that lead to a welfare loss, we see that mergers between firms with a larger number of patents and a larger number of collaborations (high degree) lead to welfare gains, while mergers between firms with few collaborations, fewer patents and a large market share lead to a welfare loss. This indicates that welfare gains are largest when two well connected and R&D intensive firms merge, while welfare losses dominate when less connected and more market dominant firms are involved in the merger. Moreover, from Table 5 we find that the highest welfare gains from a merger are larger than the highest welfare losses from a merger, indicating that the R&D spillover effects are larger than the market distortion effects. Finally, as expected, in the absence of the R&D collaboration network in column eleven in Table 5 ( $\Delta W_N$ ), the merger between two firms always leads to a reduction in welfare [Salant et al., 1983].

### 4.3. R&D Collaboration Subsidy

Many governments provide R&D subsidies to foster the R&D activities of private 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 industrial R&D. The main aim of this EU programme is to strengthen European competitiveness in the field of R&D by means of promoting market-driven collaborative research and technology development. Another example is Germany, where the federal government provides R&D subsidies to stimulate collaboration activities between private organizations [Broekel and Graf, 2012]. In this section we analyze the impact of changes in the collaboration cost due to R&D subsidies on aggregate welfare.<sup>49</sup>

We analyze a counterfactual policy that selects a specific firm-pair,  $(i, j)$ , and compensates their collaboration costs through a subsidy, i.e., setting  $\psi_{ij} = 0$  (see Equation (23) and thereafter).<sup>50</sup> We then can evaluate the changes in welfare due to such a subsidy using our estimated model. The pair of firms for which the subsidy results in the largest gain in welfare is defined as

$$(i, j)^* = \operatorname{argmax}_{(i, j) \in \mathcal{E}} \left\{ \sum_{G \in \mathcal{G}^n} \int_{\mathcal{Q}^n} W(\mathbf{q}, G | \psi_{ij} = 0) \mu^\vartheta(\mathbf{q}, G) d\mathbf{q} \right\}, \quad (34)$$

where the probability measure  $\mu^\vartheta(\mathbf{q}, G)$  is given by Equation (7), the welfare function  $W(\mathbf{q}, G)$

<sup>49</sup>For a theoretical and empirical analysis of R&D subsidies provided by the Finnish Funding Agency for Technology and Innovation see Takalo et al. [2013].

<sup>50</sup>Observe that in terms of the objective function of the planner the net effect of this policy intervention is zero. This is because the cost of the subsidy is exactly compensated for by the gain in firms' profits, which in turn, are part of the welfare function.

Table 5: Merger ranking for firms in the chemicals and allied products sector (SIC-28).

Firm $i$	Firm $j$	Mkt. Sh. $i$ [%] <sup>a</sup>	Mkt. Sh. $j$ [%]	Pat. $i$	Pat. $j$	$d_i$	$d_j$	$\Delta W$ [%] <sup>b</sup>	$\Delta W_F$ [%] <sup>c</sup>	$\Delta W_N$ [%] <sup>d</sup>	SIC	Rank
<b>WELFARE LOSS</b>												
Daiichi Sankyo Co. Ltd.	Schering-Plough Corp.	0.4590	0.6057	14	52847	5	1	-0.6036	0.0476	-0.2386	283	1
MorphoSys AG	Daiichi Sankyo Co. Ltd.	0.0038	0.4590	20	14	4	5	-0.5976	0.0132	-0.3948	283	2
Vical Inc.	Cephalon	0.0008	0.1005	170	810	1	1	-0.5639	0.3903	-0.3111	283	3
Galapagos NV	Medarex Inc.	0.0025	0.0028	30	168	2	9	-0.5581	0.1017	-0.3253	283	4
Galapagos NV	Coley Pharm. Group Inc.	0.0025	0.0012	30	125	2	1	-0.5409	0.2329	-0.3935	283	5
Infinity Pharm. Inc.	Abnylam Pharm. Inc.	0.0011	0.0015	44	114	4	3	-0.5339	0.0484	-0.3309	283	6
Icagen	Biosite Inc.	0.0005	0.0177	423	182	1	3	-0.5261	0.3587	-0.3244	283	7
Clinical Data Inc.	Renovis	0.0037	0.0006	9	58	4	1	-0.5179	0.3005	-0.3890	283	8
Clinical Data Inc.	Curagen	0.0037	0.0023	9	174	4	3	-0.5134	0.0108	-0.3450	283	9
EntreMed Inc.	AVI BioPharma Inc.	0.0004	0.0000	62	67	3	1	-0.5120	0.2734	-0.3213	283	10
<b>WELFARE GAIN</b>												
Isis Pharm. Inc.	Takeda Pharm. Co. Ltd.	0.0014	0.6445	4472	19460	4	7	0.8643	0.3406	-0.3517	283	1
Cell Genesys Inc.	Pfizer Inc.	0.0001	2.7679	236	78061	5	15	0.8636	0.6395	-0.3692	283	2
Exelixis	Pfizer Inc.	0.0057	2.7679	58	78061	7	15	0.8235	0.5397	-0.4127	283	3
Dyax Corp	Pfizer Inc.	0.0007	2.7679	227	78061	6	15	0.7717	0.5548	-0.4120	283	4
Bristol-Myers Squibb Co.	Novartis	1.0287	2.0691	22312	18924	6	15	0.7696	0.4889	-0.2978	283	5
Exelixis	Takeda Pharm. Co. Ltd.	0.0057	0.6445	58	19460	7	7	0.7661	0.5511	-0.3254	283	6
Exelixis	Novartis	0.0057	2.0691	58	18924	7	15	0.7637	0.5130	-0.3872	283	7
Genzyme Corp.	Pfizer Inc.	0.1830	2.7679	1116	78061	3	15	0.7441	0.4206	-0.3572	283	8
Medarex Inc.	Allergan Inc.	0.0028	0.1759	168	6154	9	3	0.7441	0.3586	-0.2983	283	9
Medarex Inc.	Amgen	0.0028	0.8193	168	6960	9	13	0.7411	0.7776	-0.2699	283	10

<sup>a</sup> Market share in the primary 3-digit sector in which the firm is operating.

<sup>b</sup> The relative welfare change due to a merger of firms  $i$  and  $j$  is computed as  $\Delta W = (\mathbb{E}_{\mu^{\theta}} [W_{i,j}(G, \mathbf{q})] - W(\mathbf{q}^{\text{obs}}, G^{\text{obs}})) / W(\mathbf{q}^{\text{obs}}, G^{\text{obs}})$ , where  $\mathbf{q}^{\text{obs}}$  and  $G^{\text{obs}}$  denote the observed R&D expenditures and network, respectively.

<sup>c</sup>  $\Delta W_F$  denotes the relative welfare change due to a merger of firms assuming a fixed network of R&D collaborations.

<sup>d</sup>  $\Delta W_N$  denotes the relative welfare change due to a merger of firms in the absence of a network of R&D collaborations.

is defined in Equation (19) and  $W(\mathbf{q}, G | \psi_{ij} = 0)$  denotes the welfare function with firms  $i$  and  $j$  receiving a subsidy such that they do not incur a pair-specific collaboration cost (by setting  $\psi_{ij} = 0$  permanently). The results can be seen in Table 6. In column nine ( $\Delta W$ ) we find that a subsidy between *Dynavax Technologies*, a U.S.-based clinical-stage biopharmaceutical company, and *Shionogi & Co. Ltd.*, a Japanese pharmaceutical company, would yield a welfare gain of 0.76 %. The welfare gain in column ten ( $\Delta W_F$ ) is much weaker when we impose a fixed network, highlighting the importance of dynamic network effects. The ranking illustrates that welfare gains from subsidizing R&D collaborations can be obtained for firms which otherwise would only be involved in few if any collaborations. Our framework could be used to guide governmental funding agencies that typically do not take into account the spillovers generated within a dynamic R&D network structure.

## 5. Conclusion

In this paper we have introduced a tractable model to analyze the coevolution of networks and behavior, and we have given an application to the formation of R&D collaboration networks in which firms are competitors on the product market. We provide a complete equilibrium characterization and show that our model can reproduce the observed patterns in real world networks. Moreover, the model can be conveniently estimated using state of the art Bayesian algorithms, and can be estimated even for large networks. Further, the model is amenable to policy analysis, and we illustrate this with examples for firm exit, M&As and subsidies in the context of R&D collaboration networks.

Due to the generality of the payoff function we consider (see Section 2.1), we believe that our model – both from a theoretical and empirical perspectives – can be applied to a variety of related contexts, where externalities can be modelled in the form of an adaptive network. Examples include peer effects in education, crime, risk sharing, scientific co-authorship, etc. [Jackson et al., 2015]. Our methodology can also be applied to study discrete choice models [Badev, 2013; König, 2016], and network games with local substitutes [Bramoullé and Kranton, 2007], when we assume a negative sign for the local externality parameter in our payoff function.

Table 6: Subsidy ranking for firms in the chemicals and allied products sector (SIC-28).

Firm $i$	Firm $j$	Mkt. Sh. $i$ [%] <sup>a</sup>	Mkt. Sh. $j$ [%]	Pat. $i$	Pat. $j$	$d_i$	$d_j$	$\Delta W$ [%] <sup>b</sup>	$\Delta W_F$ [%] <sup>c</sup>	SIC $i$	SIC $j$	Rank
Dynavax Technologies	Shionogi & Co. Ltd.	0.0003	0.0986	162	10156	0	0	0.7646	0.0509	283	283	1
Ar-Qule	Kemira Oy.	0.0004	0.3340	43	510	1	0	0.7622	0.0252	283	280	2
Indevus Pharm. Inc.	Solvay SA	0.0029	1.2445	37	22689	0	3	0.7603	0.0713	283	280	3
Nippon Kayaku Co. Ltd.	Koninklijke DSM NV	0.1342	1.1059	4398	4674	0	1	0.7543	0.0369	280	280	4
Encysive Pharm. Inc.	Johnson & Johnson Inc.	0.0011	3.0547	280	1212	0	7	0.7466	0.1111	283	283	5
Kaken Pharm. Co. Ltd.	Elancorp	0.0377	0.0322	821	462	0	3	0.7315	0.0986	283	283	6
Tsumura & Co.	Syngenta AG	0.0451	4.1430	23	5397	0	0	0.7215	-0.0188	283	287	7
NOF Corp.	Alkermes Inc.	0.1361	0.0138	431	31	0	0	0.7166	0.0132	280	283	8
Toagosei Co. Ltd.	Mitsubishi Tanabe Pharma Corp.	0.1412	0.0877	771	5296	0	1	0.7160	-0.0004	280	283	9
DOV Pharm. Inc.	Mochida Pharm. Co.	0.0015	0.0366	80	575	1	0	0.7158	0.0188	283	283	10
Geron	Elancorp	0.0002	0.0322	240	462	1	3	0.7146	0.0039	283	283	11
Tanox Inc.	PPG Industries Inc.	0.0032	7.5437	139	29784	0	0	0.7145	0.0283	283	285	12
Gedeon Richter	Dade Behring Inc.	0.0572	0.0999	11115	152	0	0	0.7103	0.0173	283	283	13
Nippon Kayaku Co. Ltd.	Valeant Pharm.	0.1342	0.0521	4398	312	0	0	0.7087	0.0695	280	283	14
Geron	Akzo Nobel NV	0.0002	11.7496	240	11366	1	2	0.7080	0.0114	283	285	15
Rigel Pharm. Inc.	Kyorin Holdings Inc.	0.0019	0.0381	259	2986	1	0	0.7074	0.0319	283	283	16
Indevus Pharm. Inc.	MannKind Corporation	0.0029	0.0000	37	32	0	0	0.7064	0.0144	283	283	17
Biosite Inc.	Toyama Chemical Co. Ltd.	0.0177	0.0083	182	2320	1	0	0.7062	-0.0179	283	283	18
Tsumura & Co	Alnylam Pharm. Inc.	0.0451	0.0015	23	114	0	3	0.7053	0.0222	283	283	19
Gen-Probe Inc.	Mitsubishi Tanabe Pharma Corp.	0.0201	0.0877	1179	5296	1	1	0.7046	0.0101	283	283	20

<sup>a</sup> Market share in the primary 3-digit sector in which the firm is operating.

<sup>b</sup> The relative welfare gain due to subsidizing the R&D collaboration costs between firms  $i$  and  $j$  is computed as  $\Delta W = (\mathbb{E}_{\mu^{\psi}} [W(\mathbf{q}, G|\psi_{ij} = 0)] - W(\mathbf{q}^{\text{obs}}, G^{\text{obs}})) / W(\mathbf{q}^{\text{obs}}, G^{\text{obs}})$ , where  $\mathbf{q}^{\text{obs}}$  and  $G^{\text{obs}}$  denote the observed R&D expenditures and network, respectively.

<sup>c</sup>  $\Delta W_F$  denotes the relative welfare gain due to a subsidy assuming a fixed network of R&D collaborations.

## References

- Airoldi, E., Goldenberg, A., Zheng, A., and Fienberg, S. (2009). A survey of statistical network models. *Machine Learning*, 2(2):129–233.
- Anderson, S., Goeree, J., and Holt, C. (2004). Noisy directional learning and the logit equilibrium. *Scandinavian Journal of Economics*, 106(3):581–602.
- Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete choice theory of product differentiation*. MIT Press.
- Anderson, S. P., Goeree, J. K., and Holt, C. A. (2001). Minimum-effort coordination games: Stochastic potential and logit equilibrium. *Games and Economic Behavior*, 34(2):177–199.
- Badev, A. (2013). Discrete games in endogenous networks: Theory and policy. *University of Pennsylvania Working Paper PSC 13–05*.
- Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who’s who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417.
- Belhaj, M., Bervoets, S., and Deroian, F. (2016). Optimal networks in games with local complementarities. *Theoretical Economics*, 11.
- Ben-Akiva, M. and Watanatada, T. (1981). Application of a continuous spatial choice logit model. *Structural analysis of discrete data with econometric applications*, pages 320–343.
- Bhamidi, S., Bresler, G., Sly, A., et al. (2011). Mixing time of exponential random graphs. *The Annals of Applied Probability*, 21(6):2146–2170.
- Bimpikis, K., Ehsani, S., and Ilklic, R. (2014). Cournot competition in networked markets. *Mimeo, Stanford University*.
- Bisin, A., Horst, U., and Özgür, O. (2006). Rational expectations equilibria of economies with local interactions. *Journal of Economic Theory*, 127(1):74–116.
- Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.
- Blume, L. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424.
- Blume, L. (2003). How noise matters. *Games and Economic Behavior*, 44(2):251–271.
- 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.
- Brock, W. and Durlauf, S. (2001). Discrete choice with social interactions. *Review of Economic Studies*, 68(2):235–260.
- 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.
- Cabrales, A., Calvó-Armengol, A., and Zenou, Y. (2011). Social interactions and spillovers. *Games and Economic Behavior*, 72(2):339–360.
- 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. *The American Economic Review*, 84(2):159–163.
- Cohen, W., Levin, R., and Mowery, D. (1987). Firm size and R&D intensity: a re-examination. *The 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. *The American Economic Review*, 78(5):1133–1137.
- Daughety, A. F. (1990). Beneficial concentration. *The American Economic Review*, 80(5):1231–1237.
- Daughety, A. F. (2005). *Cournot oligopoly: characterization and applications*. Cambridge University Press.
- David, P. A. (1992). Heroes, herds and hysteresis in technological history: Thomas Edison and “The battle of the systems” reconsidered. *Industrial and Corporate Change*, 1(1):129–180.
- Dawid, H. and Hellmann, T. (2014). The evolution of R&D networks. *Journal of Economic Behavior & Organization*, 105:158–172.
- 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.
- Ehrhardt, G., Marsili, M., and Vega-Redondo, F. (2008). Networks emerging in a volatile world. *Economics Working Papers*.
- Encaoua, D. and Hollander, A. (2002). Competition policy and innovation. *Oxford Review of Economic Policy*, 18(1):63–79.
- Farrell, J. and Shapiro, C. (1990). Horizontal mergers: an equilibrium analysis. *The American Economic Review*, 80(1):107–126.
- 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.
- Feri, F. (2007). Stochastic stability in networks with decay. *Journal of Economic Theory*, 135(1):442–457.
- Fosco, C., Vega-Redondo, F., and Marsili, M. (2010). Peer effects and peer avoidance: The diffusion of behavior in coevolving networks. *Journal of the European Economic Association*, 8(1):169–202.
- Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842.
- Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 166:1710.

- Gabaix, X. (2016). Power laws in economics: An introduction. *The Journal of Economic Perspectives*, 30(1):185–205.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, pages 169–193.
- Geyer, C. J. (1991). *Markov chain Monte Carlo maximum likelihood*. Interface Foundation of North America.
- 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.
- 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.
- 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.
- Hanaki, N., Nakajima, R., and Ogura, Y. (2010). The dynamics of R&D network in the IT industry. *Research Policy*, 39(3):386–399.
- Hiller, T. (2013). Peer effects in endogenous networks. *Available at SSRN 2331810*.
- Hojman, D. and Szeidl, A. (2008). Core and periphery in networks. *Journal of Economic Theory*, 139(1):295–309.
- Hojman, D. A. and Szeidl, A. (2006). Endogenous networks, social games, and evolution. *Games and Economic Behavior*, 55(1):112–130.
- Horn, R. A. and Johnson, C. R. (1990). *Matrix Analysis*. Cambridge University Press.
- Hsieh, C.-s. and Lee, L.-f. (2013). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *CUHK Working Paper*.
- 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. (2008). *Social and Economic Networks*. Princeton University Press.
- Jackson, M. O., Rogers, B., and Zenou, Y. (2017). The economic consequences of social network structure. *Journal of Economic Literature*, 55(1):49–95.
- Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.
- 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., Zenou, Y., et al. (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.
- Jin, I. H., Yuan, Y., and Liang, F. (2013). Bayesian analysis for exponential random graph models using the adaptive exchange sampler. *Statistics and its interface*, 6(4):559.
- Kandori, M., Mailath, G., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56.
- Kelly, M. J., Schaan, J.-L., and Joncas, H. (2002). Managing alliance relationships: key challenges in the early stages of collaboration. *R&D Management*, 32(1):11–22.
- Khalil, H. K. (2002). *Nonlinear Systems*. Prentice Hall.
- 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., Tessone, C., and Zenou, Y. (2014a). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics*, 9:695–752.
- König, M. D. (2016). Diffusion of behavior in dynamic networks. *University of Zurich, Department of Economics Working Paper No. 222*.
- König, M. D., Liu, X., and Zenou, Y. (2014b). R&D networks: Theory, empirics and policy implications. *CEPR Discussion Paper No. 9872*.
- König, M. D., Lorenz, J., and Zilibotti, F. (2016). Innovation vs. imitation and the evolution of productivity distributions. *Theoretical Economics*, 11:1053–1102.
- Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.
- Liang, F. (2010). A double Metropolis–Hastings sampler for spatial models with intractable normalizing constants. *Journal of Statistical Computation and Simulation*, 80(9):1007–1022.
- Liang, F., Jin, I. H., Song, Q., and Liu, J. S. (2015). An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *Journal of the American Statistical Association*, 111(513):377–393.
- Liang, F., Liu, C., and Carroll, R. (2011). *Advanced Markov chain Monte Carlo methods: learning from past samples*, volume 714. John Wiley & Sons.
- Liang, F., Liu, C., and Carroll, R. J. (2007). Stochastic approximation in Monte Carlo computation. *Journal of the American Statistical Association*, 102(477):305–320.
- 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.
- Marshall, G. and Parra, A. (2015). Mergers in innovative industries. *Mimeo, University of Illinois at Urbana-Champaign*.
- Marsili, M., Vega-Redondo, F., and Slanina, F. (2004). The rise and fall of a networked society: a formal model. *Proceedings of the National Academy of Sciences*, 101(6):1439–1442.
- McFadden, D. (1976). The mathematical theory of demand models. *Behavioral Travel Demand Models*, pages 305–314.
- Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.
- Monderer, D. and Shapley, L. (1996). Potential Games. *Games and Economic Behavior*, 14(1):124–143.

- 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.
- Norris, J. R. (1998). *Markov chains*. Cambridge University Press.
- Park, J. and Newman, M. (2004). Statistical mechanics of networks. *Physical Review E*, 70(6):66117.
- 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.
- Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures & Algorithms*, 9(1-2):223–252.
- Raftery, A. E., Lewis, S., Banfield, J. D., and Raftery, A. E. (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, page 4.
- 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.
- Rudin, W. (1987). *Real and complex analysis*. Tata McGraw-Hill Education.
- Salant, S. W., Switzer, S., and Reynolds, R. J. (1983). Losses from horizontal merger: the effects of an exogenous change in industry structure on Cournot-Nash equilibrium. *The Quarterly Journal of Economics*, 98(2):185–199.
- Sandholm, W. (2010). *Population games and evolutionary dynamics*. MIT Press.
- Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal*, 30(3):233–260.
- Shalizi, C. R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics*, 41(2):508–535.
- Smyth, G. K. (1996). Partitioned algorithms for maximum likelihood and other non-linear estimation. *Statistics and Computing*, 6(3):201–216.
- Snijders, T. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.
- 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.
- Tomasello, M. V., Napoletano, M., Garas, A., and Schweitzer, F. (2016). The rise and fall of R&D networks. Available at SSRN 2749230.
- 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.
- Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- 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. *The RAND Journal of Economics*, 41(2):351–371.
- Wong, R. (2001). *Asymptotic approximations of integrals*, volume 34. Society for Industrial Mathematics.
- Zenou, Y. (2015). Key players. *Oxford Handbook on the Economics of Networks*, Y. Bramoulle, B. Rogers and A. Galeotti (Eds.), Oxford University Press.



## 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 exactly equivalent to the difference in the potential function from adding or removing a link. Similarly, the marginal profit of a firm from changing its output level is exactly equivalent to the change of the potential function.

**PROOF OF PROPOSITION 1.** The potential  $\Phi(\mathbf{q}, G)$  has the property that  $\Phi(\mathbf{q}, G \oplus (i, j)) - \Phi(\mathbf{q}, G) = \rho q_i q_j - \zeta = \pi_i(\mathbf{q}, G \oplus (i, j)) - \pi_i(\mathbf{q}, G)$ , and similarly,  $\Phi(\mathbf{q}, G \ominus (i, j)) - \Phi(\mathbf{q}, G) = \zeta - \rho q_i q_j = \pi_i(\mathbf{q}, G \ominus (i, j)) - \pi_i(\mathbf{q}, G)$  for any  $\mathbf{q} \in \mathcal{Q}^n$  and  $G \in \mathcal{G}^n$ . From the properties of  $\pi_i(\mathbf{q}, G)$  it also follows that  $\Phi(q'_i, \mathbf{q}_{-i}, G) - \Phi(q_i, \mathbf{q}_{-i}, G) = \pi_i(q'_i, \mathbf{q}_{-i}, G) - \pi_i(q_i, \mathbf{q}_{-i}, G)$ .  $\square$

We next show that the stationary distribution can be characterized by a Gibbs measure.

**PROOF OF THEOREM 1.** First, note from Equation (6) that  $q^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') > 0$  for any  $\boldsymbol{\omega} \neq \boldsymbol{\omega}'$  and finite  $\vartheta$ , 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  $\mathbf{Q}^\vartheta = (q^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}'))_{\boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega}$  is therefore irreducible. Moreover, for an irreducible Markov chain on a finite state space  $\Omega$  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 (6). This equation is satisfied when the probability distribution  $\mu^\vartheta$  satisfies the following detailed balance condition [Norris, 1998]

$$\forall \boldsymbol{\omega}, \boldsymbol{\omega}' \in \Omega : \mu^\vartheta(\boldsymbol{\omega}) q^\vartheta(\boldsymbol{\omega}, \boldsymbol{\omega}') = \mu^\vartheta(\boldsymbol{\omega}') q^\vartheta(\boldsymbol{\omega}', \boldsymbol{\omega}). \quad (35)$$

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 quantity 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  $q'_i \neq q_i$  for some  $i \in \mathcal{N}$ . For the case of link creation with a transition from  $\boldsymbol{\omega} = (\mathbf{q}, G)$  to  $\boldsymbol{\omega}' = (\mathbf{q}, G \oplus (i, j))$  we can write the detailed balance condition as follows

$$\frac{1}{\mathcal{Z}_\theta} e^{\vartheta(\Phi(\mathbf{q}, G) - m \ln(\frac{\xi}{\tau}))} \frac{e^{\vartheta\Phi(\mathbf{q}, G \oplus (i, j))}}{e^{\vartheta\Phi(\mathbf{q}, G \oplus (i, j))} + e^{\vartheta\Phi(\mathbf{q}, G)}} \tau = \frac{1}{\mathcal{Z}_\theta} e^{\vartheta(\Phi(\mathbf{q}, G \oplus (i, j)) - (m+1) \ln(\frac{\xi}{\tau}))} \frac{e^{\vartheta\Phi(\mathbf{q}, G)}}{e^{\vartheta\Phi(\mathbf{q}, G)} + e^{\vartheta\Phi(\mathbf{q}, G \oplus (i, j))}} \xi.$$

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

$$\frac{1}{\mathcal{Z}_\theta} e^{\vartheta(\Phi(\mathbf{q}, G) - m \ln(\frac{\xi}{\tau}))} \frac{e^{\vartheta\Phi(\mathbf{q}, G \ominus (i, j))}}{e^{\vartheta\Phi(\mathbf{q}, G \ominus (i, j))} + e^{\vartheta\Phi(\mathbf{q}, G)}} \xi = \frac{1}{\mathcal{Z}_\theta} e^{\vartheta(\Phi(\mathbf{q}, G \ominus (i, j)) - (m-1) \ln(\frac{\xi}{\tau}))} \frac{e^{\vartheta\Phi(\mathbf{q}, G)}}{e^{\vartheta\Phi(\mathbf{q}, G)} + e^{\vartheta\Phi(\mathbf{q}, G \ominus (i, j))}} \tau.$$

For a change in the output level with a transition from  $\boldsymbol{\omega} = (q_i, \mathbf{q}_{-i}, G)$  to  $\boldsymbol{\omega}' = (q'_i, \mathbf{q}_{-i}, G)$  we get for the following detailed balance condition

$$\frac{1}{\mathcal{Z}_\theta} e^{\vartheta(\Phi(q_i, \mathbf{q}_{-i}, G) - m \ln(\frac{\xi}{\tau}))} \frac{e^{\vartheta\pi_i(q'_i, \mathbf{q}_{-i}, G)}}{\int_{\mathcal{Q}} e^{\vartheta\pi_i(q', \mathbf{q}_{-i}, G)} dq'} \chi = \frac{1}{\mathcal{Z}_\theta} e^{\vartheta(\Phi(q'_i, \mathbf{q}_{-i}, G) - m \ln(\frac{\xi}{\tau}))} \frac{e^{\vartheta\pi_i(q_i, \mathbf{q}_{-i}, G)}}{\int_{\mathcal{Q}} e^{\vartheta\pi_i(q', \mathbf{q}_{-i}, G)} dq'} \chi.$$

This can be written as  $e^{\vartheta(\Phi(q_i, \mathbf{q}_{-i}, G) - \Phi(q'_i, \mathbf{q}_{-i}, G))} = e^{\vartheta(\pi_i(q_i, \mathbf{q}_{-i}, G) - \pi_i(q'_i, \mathbf{q}_{-i}, G))}$ , which is satisfied since we have for the potential  $\Phi(q_i, \mathbf{q}_{-i}, G) - \Phi(q'_i, \mathbf{q}_{-i}, G) = \pi_i(q_i, \mathbf{q}_{-i}, G) - \pi_i(q'_i, \mathbf{q}_{-i}, G)$ . Hence, the probability measure  $\mu^\vartheta$  satisfies a detailed balance condition of Equation (35) and therefore is the stationary distribution of the Markov chain with transition rate matrix  $\mathbf{Q}^\vartheta$ .  $\square$

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

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

$$\sum_{\mathbf{h} \in \mathcal{H}^n} e^{\sum_{i=1}^n h_i c_i} = \prod_{i=1}^n \sum_{h_i \in \{0,1\}} e^{h_i c_i}. \quad (36)$$

PROOF OF LEMMA 1. We give a proof by induction. For the induction basis consider  $n = 2$  (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 (36) 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  $h_1, \dots, h_n$  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 (36) holds for  $n$ . This concludes the proof.  $\square$

PROOF OF PROPOSITION 2. We start with the proof of the first part of the proposition. Observe that the potential of Equation (2) can be written as

$$\Phi(\mathbf{q}, G) = \underbrace{\sum_{i=1}^n \left( \eta - \nu q_i - \frac{b}{2} \sum_{j \neq i}^n q_j \right) q_i}_{\psi(\mathbf{q})} + \sum_{i=1}^n \sum_{j=i+1}^n a_{ij} \underbrace{(\rho q_i q_j - \zeta)}_{\sigma_{ij}} = \psi(\mathbf{q}) + \sum_{i=1}^n \sum_{j=i+1}^n a_{ij} \sigma_{ij}. \quad (37)$$

We then have that  $e^{\vartheta \Phi(\mathbf{q}, G)} = e^{\vartheta \psi(\mathbf{q})} e^{\vartheta \sum_{i < j}^n a_{ij} \sigma_{ij}}$ , where only the second factor is network dependent. Observing that the sequence  $(a_{ij})_{1 \leq i < j \leq n} = (a_{12}, a_{13}, \dots, a_{n-1, n})$  is a binary sequence as in Lemma 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<sup>51</sup>

$$\sum_{G \in \mathcal{G}^n} e^{\vartheta \sum_{i < j} a_{ij} \sigma_{ij}} = \prod_{i=1}^n \prod_{j=i+1}^n \left(1 + e^{\vartheta \sigma_{ij}}\right). \quad (38)$$

From Equation (38) we then obtain

$$\sum_{G \in \mathcal{G}^n} e^{\vartheta \Phi(\mathbf{q}, G)} = e^{\vartheta \psi(\mathbf{q})} \prod_{i < j}^n \left(1 + e^{\vartheta \sigma_{ij}}\right) = \prod_{i=1}^n e^{\vartheta \left(\eta_i - \nu q_i - \frac{b}{2} \sum_{j \neq i} q_j\right) q_i} \prod_{i < j}^n \left(1 + e^{\vartheta (\rho q_i q_j - \zeta)}\right). \quad (39)$$

We can use Equation (39) to compute the marginal distribution

$$\begin{aligned} \mu^\vartheta(\mathbf{q}) &= \frac{1}{\mathcal{Z}_\vartheta} \sum_{G \in \mathcal{G}^n} e^{\vartheta \Phi(\mathbf{q}, G)} \\ &= \frac{1}{\mathcal{Z}_\vartheta} \prod_{i=1}^n e^{\vartheta \left(\eta_i - \nu q_i - \frac{b}{2} \sum_{j \neq i} q_j\right) q_i} \prod_{i < j}^n \left(1 + e^{\vartheta (\rho q_i q_j - \zeta)}\right) \\ &= \frac{1}{\mathcal{Z}_\vartheta} e^{\vartheta \sum_{i=1}^n \left(\eta_i - \nu q_i - \frac{b}{2} \sum_{j \neq i} q_j\right) q_i} e^{\sum_{i < j} \ln \left(1 + e^{\vartheta (\rho q_i q_j - \zeta)}\right)} \\ &= \frac{1}{\mathcal{Z}_\vartheta} e^{\vartheta \mathcal{H}_\vartheta(\mathbf{q})}, \end{aligned} \quad (40)$$

where we have introduced the Hamiltonian

$$\mathcal{H}_\vartheta(\mathbf{q}) \equiv \sum_{i=1}^n \left( \eta q_i - \nu q_i^2 + \sum_{j > i} \left( \frac{1}{\vartheta} \ln \left(1 + e^{\vartheta (\rho q_i q_j - \zeta)}\right) - b q_i q_j \right) \right). \quad (41)$$

Using the fact that  $\int_{\mathcal{Q}^n} \mu^\vartheta(\mathbf{q}) d\mathbf{q} = 1$ , it follows from Equation (40) that we can write the partition function as

$$\mathcal{Z}_\vartheta = \int_{\mathcal{Q}^n} e^{\vartheta \mathcal{H}_\vartheta(\mathbf{q})} d\mathbf{q}.$$

We next make the Laplace approximation [Wong, 2001, Theorem 3, p. 495]

$$\mathcal{Z}_\vartheta \sim \left(\frac{2\pi}{\vartheta}\right)^{\frac{n}{2}} \left| \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)_{q_i = q^*} \right|^{-\frac{1}{2}} e^{\vartheta \mathcal{H}_\vartheta(\mathbf{q}^*)}, \quad (42)$$

for large  $\vartheta$ , where  $\mathbf{q}^* = \operatorname{argmax}_{\mathbf{q} \in [0, \bar{q}]^n} \mathcal{H}_\vartheta(\mathbf{q})$ , and the Hessian is given by  $\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j}$  for  $1 \leq i, j \leq n$ . From Equation (41) we find that

$$\frac{\partial \mathcal{H}_\vartheta}{\partial q_i} = \eta - 2\nu q_i + \sum_{j \neq i} \left( \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) - b \right) q_j. \quad (43)$$

The first order conditions  $\frac{\partial \mathcal{H}_\vartheta}{\partial q_i} = 0$  in Equation (43) imply that

$$\eta - 2\nu q_i = \sum_{j \neq i} \left( b - \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) \right) q_j.$$

---

<sup>51</sup>Note that Equation (38) requires the  $\sigma_{ij}$  to be constant, and in particular, to be independent. In this case the summation over all networks only needs to count the number of possible networks in which the link  $ij$  is present. In contrast, when  $\sigma_{ij}$  depends on the other links in the network, then this simple summation formula would no longer hold.

This system of equations has a symmetric solution,  $q_i = q$  for all  $i = 1, \dots, n$ , where

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

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

$$(b + 2\nu^*)q - \eta^* = \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right) q. \quad (44)$$

Let the RHS of Equation (44) be denoted by  $F(q)$  so that we can write it as  $(b + 2\nu^*)q - \eta^* = F(q)$ . Then we have that  $F(0) = 0$ ,  $F'(q) \geq 0$  and  $F(q) \sim \rho q$  for  $q \rightarrow \infty$ . It follows that  $(b + 2\nu^*)q - \eta^* = F(q)$  has at least one solution when  $b + 2\nu^* > \rho$ .<sup>52</sup> Moreover, any iteration  $(b + 2\nu^*)q_{t+1} - \eta^* = F(q_t)$  starting at  $q_0 = 0$  converges to the smallest fixed point  $q^*$  such that  $(b + 2\nu^*)q^* - \eta^* = F(q^*)$ .

We next compute the average output level  $\bar{q} = \frac{1}{n} \sum_{i=1}^n q_i$ . We have that

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

where we have denoted by  $\mathcal{F}_\vartheta \equiv -\ln \mathcal{Z}_\vartheta$ . The average output is then given by

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

With Equation (42) we get

$$\mathcal{F}_\vartheta \sim -\frac{n}{2} \ln \left( \frac{2\pi}{\vartheta} \right) + \frac{1}{2} \ln \left| \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)_{q_i=q^*} \right| - \vartheta \mathcal{H}_\vartheta(\mathbf{q}^*).$$

We then find that

$$\begin{aligned} \frac{\partial \mathcal{F}_\vartheta}{\partial \eta} &= -\vartheta \frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \eta} + \frac{1}{2} \frac{\partial}{\partial \eta} \ln \left| \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)_{q_i=q^*} \right| \\ &= -\vartheta \frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \eta} + \frac{1}{2} \text{tr} \left( \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)^{-1} \frac{\partial}{\partial \eta} \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right) \right)_{q_i=q^*}, \end{aligned}$$

where we have used Jacobi's formula [Horn and Johnson, 1990].<sup>53</sup> From Equation (43) we further have that

$$\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i^2} = -2\nu + \frac{\vartheta \rho^2}{4} \sum_{j \neq i}^n q_j^2 \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right)^2, \quad (45)$$

<sup>52</sup>Since the RHS,  $F(q)$ , of Equation (44) is increasing (one can see this from taking the derivative), is zero at  $q = 0$ , i.e.  $F(0) = 0$ , and asymptotically grows linearly as  $\rho q$ , it follows that when  $b + 2\nu^* > \rho$  there must exist at least one fixed point. This is because the LHS,  $(b + 2\nu^*)q - \eta^*$ , of Equation (44) starts below zero at  $q = 0$  (where it is  $-\eta^*$ ), both LHS and RHS are increasing, and the RHS approaches asymptotically a line with a slope smaller than the slope  $b + 2\nu^*$  of the LHS. Hence they must intersect at some  $q \geq 0$ .

<sup>53</sup>For any invertible matrix  $\mathbf{M}(x)$  for all  $x$ , Jacobi's formula states that  $\frac{d}{dx} |\mathbf{M}(x)| = |\mathbf{M}(x)| \text{tr} \left( \mathbf{M}(x)^{-1} \frac{d}{dx} \mathbf{M}(x) \right)$ , which can be written more compactly as  $\frac{d}{dx} \ln |\mathbf{M}(x)| = \text{tr} \left( \mathbf{M}(x)^{-1} \frac{d}{dx} \mathbf{M}(x) \right)$ .

and for  $j \neq i$  we have that

$$\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} = -b + \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) \left( 1 + \frac{\vartheta \rho}{2} q_i q_j \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) \right). \quad (46)$$

This shows that  $\frac{\partial}{\partial \eta} \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right) = 0$ , so that  $\frac{\partial \mathcal{F}_\vartheta}{\partial \eta} = -\vartheta \frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \eta}$ , and the expected average output level is then given by

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

Using the fact that  $\frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \eta} = \sum_{i=1}^n q_i = nq^*$ , we then get in leading order terms for large  $\vartheta$  that

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

Next, we compute the output distribution. It can be written as follows

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

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

$$\mathcal{H}_\vartheta(\mathbf{q}) = \mathcal{H}_\vartheta(\mathbf{q}^*) + (\mathbf{q} - \mathbf{q}^*)^\top \nabla \mathcal{H}_\vartheta(\mathbf{q}^*) + \frac{1}{2} (\mathbf{q} - \mathbf{q}^*)^\top \Delta \mathcal{H}_\vartheta(\mathbf{q}^*) (\mathbf{q} - \mathbf{q}^*) + o(\|\mathbf{q} - \mathbf{q}^*\|^2),$$

as  $\vartheta \rightarrow \infty$ , where  $\mathbf{q}^* = \operatorname{argmax}_{\mathbf{q} \in [0, \bar{q}]^n} \mathcal{H}_\vartheta(\mathbf{q})$ , the gradient is  $\nabla \mathcal{H}_\vartheta(\mathbf{q}) = \left( \frac{\partial \mathcal{H}_\vartheta}{\partial q_i} \right)_{i=1, \dots, n}$ , and the Hessian is  $\Delta \mathcal{H}_\vartheta(\mathbf{q}) = \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)_{i, j=1, \dots, n}$ . As the gradient  $\nabla \mathcal{H}_\vartheta(\mathbf{q})$  vanishes at  $\mathbf{q}^*$ , we have that

$$\mathcal{H}_\vartheta(\mathbf{q}) = \mathcal{H}_\vartheta(\mathbf{q}^*) + \frac{1}{2} (\mathbf{q} - \mathbf{q}^*)^\top \Delta \mathcal{H}_\vartheta(\mathbf{q}^*) (\mathbf{q} - \mathbf{q}^*) + o(\|\mathbf{q} - \mathbf{q}^*\|^2).$$

We then can write

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

Normalization implies that

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

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

$$\mu^\vartheta(\mathbf{q}) = \left( \frac{2\pi}{\vartheta} \right)^{-\frac{n}{2}} |\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top (-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)) (\mathbf{q} - \mathbf{q}^*) \right\} + o(\|\mathbf{q} - \mathbf{q}^*\|^2). \quad (47)$$

That is, in the limit of large  $\vartheta$ ,  $\mathbf{q}$  is asymptotically normally distributed with mean  $\mathbf{q}^*$  and variance  $-\frac{1}{\vartheta} \nabla \mathcal{H}_\vartheta(\mathbf{q}^*)^{-1}$ .

Imposing symmetry,  $q_i = q$  for all  $i = 1, \dots, n$ , in Equation (45) we can write

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

and for  $j \neq i$  we have from Equation (45) that

$$\left. \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right|_{q_i=q_j=q} = -b + \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right) \left( 1 + \frac{\vartheta \rho}{2} q^2 \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right) \right).$$

Using Equation (44), from which we get

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

and

$$\frac{\rho}{2} \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) = \frac{((n-1)(\rho - b) - 2\nu)q + \eta}{(n-1)q},$$

we then can write

$$\left. \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i^2} \right|_{q_i=q} = -2\nu + \frac{\vartheta(((n-1)(\rho - b) - 2\nu)q + \eta)((n-1)b + 2\nu)q - \eta}{n-1},$$

and

$$\left. \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right|_{q_i=q_j=q} = -b + \frac{((n-1)b + 2\nu)q - \eta}{(n-1)q} \left( 1 + \frac{\vartheta q(((n-1)(\rho - b) - 2\nu)q + \eta)}{n-1} \right).$$

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

$$\left. \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i^2} \right|_{q_i=q} = (n-1) \left( -2\nu^* + \vartheta q^2 \left( \rho - b - 2\nu^* + \frac{\eta^*}{q} \right) \left( b + 2\nu^* - \frac{\eta^*}{q} \right) \right), \quad (48)$$

and

$$\left. \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right|_{q_i=q_j=q} = -b + \left( b + 2\nu^* - \frac{\eta^*}{q} \right) \left( 1 + \vartheta q^2 \left( \rho - b - 2\nu^* + \frac{\eta^*}{q} \right) \right). \quad (49)$$

Note that due to symmetry, the Hessian  $\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)$  with components in Equations (48) and (49) is a special case of a circulant matrix. Denoting by  $a$  the diagonal elements of  $\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)$  and by  $b$  the off-diagonal elements, the determinant in Equation (47) follows from the general formula [Horn and Johnson, 1990]:

$$|-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)| = \begin{vmatrix} a & b & b & \dots \\ b & a & b & \dots \\ b & b & a & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} = (a-b)^{n-1} (a + (n-1)b).$$

Similarly, for a circulant matrix (by applying the Sherman-Morrison formula (Horn and Johnson

[1990]) we get for the inverse in Equation (47) that

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

For large  $n$  we see from Equations (48) and (49) that the off-diagonal elements vanish relative to the diagonal elements. As  $\mathbf{q}$  is asymptotically normally distributed with mean  $\mathbf{q}^*$  and variance  $-\frac{1}{\vartheta} \Delta \mathcal{H}_\vartheta(\mathbf{q}^*)^{-1}$ , this implies that, in the limit of  $n \rightarrow \infty$ , the individual firms' output levels become independent. The diagonal entries are given by

$$-\frac{1}{\vartheta} (\Delta \mathcal{H}_\vartheta(\mathbf{q}^*))_{ii}^{-1} \sim \frac{1}{\vartheta} \frac{n}{2\nu^* + \vartheta(bq - \eta^* + 2\nu^*q)(q(b + 2\nu^* - \rho) - \eta^*)} \equiv \sigma^2.$$

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

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

where we have denoted by  $\mathcal{F}_\vartheta \equiv -\ln \mathcal{Z}_\vartheta$ . From the Laplace approximation in Equation (42) we find that

$$\begin{aligned} \frac{\partial \mathcal{F}_\vartheta}{\partial \zeta} &= -\vartheta \frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \zeta} + \frac{1}{2} \frac{\partial}{\partial \zeta} \ln \left| \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)_{q_i=q^*} \right| \\ &= -\vartheta \frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \zeta} + \frac{1}{2} \text{tr} \left( \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)^{-1} \frac{\partial}{\partial \zeta} \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right) \right)_{q_i=q^*}, \end{aligned}$$

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

$$\mathbb{E}_{\mu^\vartheta}(m) = -\frac{\partial \mathcal{H}_\vartheta(\mathbf{q}^*)}{\partial \zeta} + \frac{1}{2\vartheta} \text{tr} \left( \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)^{-1} \frac{\partial}{\partial \zeta} \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right) \right)_{q_i=q^*}.$$

Further, we have that

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

and in the symmetric equilibrium this is

$$\frac{\partial \mathcal{H}_\vartheta}{\partial \zeta} \Big|_{q_i=q} = -\frac{n(n-1)}{4} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right).$$

The expected number of links can then be written as

$$\mathbb{E}_{\mu^\vartheta}(m) = \frac{n(n-1)}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right) + \frac{1}{2\vartheta} \text{tr} \left( \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right)^{-1} \frac{\partial}{\partial \zeta} \left( \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right) \right)_{q_i=q^*}.$$

Using the fact that

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

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

$$\left. \frac{\partial \mathcal{H}_\vartheta}{\partial \zeta} \right|_{q_i=q} = -\frac{n(n-1)}{2\rho} \left( b + 2\nu^* - \frac{\eta^*}{q} \right).$$

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

$$\left. \frac{\partial \mathcal{H}_\vartheta}{\partial \zeta} \right|_{q_i=q} = 0.$$

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

$$\left. \frac{\partial \mathcal{H}_\vartheta}{\partial \zeta} \right|_{q_i=q} = -\frac{n(n-1)}{2}.$$

Further, the derivatives with respect to  $\zeta$  in Equation (45) are given by

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

and for  $j \neq i$  from Equation (46) we get that

$$\frac{\partial}{\partial \zeta} \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} = -\frac{\vartheta \rho}{4} \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right)^2 \left( 1 - \vartheta \rho q_i q_j \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right).$$

Imposing symmetry,  $q_i = q$  for all  $i = 1, \dots, n$ , we then can write

$$\left. \frac{\partial}{\partial \zeta} \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i^2} \right|_{q_i=q} = \frac{(n-1)\vartheta^2 \rho^2}{4} \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right)^2, \quad (50)$$

and

$$\left. \frac{\partial}{\partial \zeta} \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \right|_{q_i=q_j=q} = -\frac{\vartheta \rho}{4} \left( 1 - \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right)^2 \left( 1 - \vartheta \rho q^2 \tanh \left( \frac{\vartheta}{2} (\rho q^2 - \zeta) \right) \right). \quad (51)$$

For a circulant matrix (by applying the Sherman-Morrison formula [Horn and Johnson, 1990] we have that

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

and

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



so that

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

The expected number of links can then be written as follows

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

where

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

with

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

In the following we compute the degree distribution. From our previous discussion we know that each firm  $i$  has an output level  $q$  distributed identically and independently with density  $\mu^\vartheta(q)$  given by  $\mathcal{N}(q^*, \sigma^2)$  and converging to  $\delta(q - q^*)$  in the limit  $\vartheta \rightarrow \infty$ . With the marginal distribution from Equation (40) and the potential in Equation (37) we then can write the conditional distribution as

$$\begin{aligned} \mu^\vartheta(G|\mathbf{q}) &= \frac{\mu^\vartheta(\mathbf{q}, G)}{\mu^\vartheta(\mathbf{q})} = \frac{e^{\vartheta \Phi(\mathbf{q}, G)}}{\sum_{G' \in \mathcal{G}^n} e^{\vartheta \Phi(\mathbf{q}, G')}} = \frac{e^{\psi(\mathbf{q})} e^{\vartheta \sum_{i < j} a_{ij} (\rho q_i q_j - \zeta)}}{e^{\psi(\mathbf{q})} \prod_{i < j} (1 + e^{\vartheta (\rho q_i q_j - \zeta)})} \\ &= \frac{e^{\vartheta \sum_{i < j} a_{ij} (\rho q_i q_j - \zeta)}}{\prod_{i < j} (1 + e^{\vartheta (\rho q_i q_j - \zeta)})} \\ &= \prod_{i < j} \frac{e^{\vartheta a_{ij} (\rho q_i q_j - \zeta)}}{1 + e^{\vartheta (\rho q_i q_j - \zeta)}} \\ &= \prod_{i < j} \left( \frac{e^{\vartheta (\rho q_i q_j - \zeta)}}{1 + e^{\vartheta (\rho q_i q_j - \zeta)}} \right)^{a_{ij}} \left( 1 - \frac{e^{\vartheta (\rho q_i q_j - \zeta)}}{1 + e^{\vartheta (\rho q_i q_j - \zeta)}} \right)^{1 - a_{ij}} \\ &= \prod_{i < j} p^\vartheta(q_i, q_j)^{a_{ij}} \left( 1 - p^\vartheta(q_i, q_j) \right)^{1 - a_{ij}}. \end{aligned} \quad (52)$$

Hence, we obtain the likelihood of an inhomogeneous random graph with link probability<sup>54</sup>

$$p^\vartheta(q_i, q_j) = \frac{e^{\vartheta (\rho q_i q_j - \zeta)}}{1 + e^{\vartheta (\rho q_i q_j - \zeta)}} = \frac{g^\vartheta(q_i, q_j)}{1 + g^\vartheta(q_i, q_j)}, \quad (53)$$

where we have denoted by  $g^\vartheta(q, q') \equiv e^{\vartheta (\rho q q' - \zeta)}$ . The probability of observing the network  $G$ ,

<sup>54</sup>See also supplementary Appendix B, Boguná and Pastor-Satorras [2003] and Söderberg [2002].

given the output levels  $\mathbf{q}$  can then be written as follows

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

which can be written as

$$\mu^\vartheta(G|\mathbf{q}) = \mathcal{C}_\vartheta(\mathbf{q}) \prod_{i=1}^n \prod_{j=i+1}^n g^\vartheta(q_i, q_j)^{a_{ij}},$$

with the normalizing constant

$$\mathcal{C}_\vartheta(\mathbf{q}) \equiv \prod_{i=1}^n \prod_{j=i+1}^n \left( 1 + g^\vartheta(q_i, q_j) \right).$$

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

$$\mathcal{C}_\vartheta(\mathbf{q}) = \sum_{G \in \mathcal{G}^n} \prod_{i=1}^n \prod_{j=i+1}^n g^\vartheta(q_i, q_j)^{a_{ij}}.$$

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

$$\mathbb{E}_{\mu^\vartheta} \left( \prod_{i=1}^n x_i^{d_i(G)} \middle| \mathbf{q} \right) = \mathbb{E} \left( \prod_{i=1}^n \prod_{j=i+1}^n (x_i x_j)^{a_{ij}} \middle| \mathbf{q} \right) \quad (54)$$

$$\begin{aligned} &= \sum_{G \in \mathcal{G}^n} \mathbb{P}(G|\mathbf{q}) \prod_{i=1}^n \prod_{j=i+1}^n (x_i x_j)^{a_{ij}} \\ &= \frac{1}{\mathcal{C}_\vartheta(\mathbf{q})} \sum_{G \in \mathcal{G}^n} \prod_{i=1}^n \prod_{j=i+1}^n g^\vartheta(q_i, q_j)^{a_{ij}} \prod_{i=1}^n \prod_{j=i+1}^n (x_i x_j)^{a_{ij}} \\ &= \frac{1}{\mathcal{C}_\vartheta(\mathbf{q})} \sum_{G \in \mathcal{G}^n} \prod_{i=1}^n \prod_{j=i+1}^n \left( g^\vartheta(q_i, q_j) x_i x_j \right)^{a_{ij}} \\ &= \frac{\sum_{G \in \mathcal{G}^n} \prod_{i=1}^n \prod_{j=i+1}^n \left( g^\vartheta(q_i, q_j) x_i x_j \right)^{a_{ij}}}{\prod_{i=1}^n \prod_{j=i+1}^n (1 + g^\vartheta(q_i, q_j))} \\ &= \prod_{i=1}^n \prod_{j=i+1}^n \frac{1 + g^\vartheta(q_i, q_j) x_i x_j}{1 + g^\vartheta(q_i, q_j)}, \end{aligned} \quad (55)$$

where we have used the fact that  $\sum_{G \in \mathcal{G}^n} \prod_{i=1}^n \prod_{j=i+1}^n \left( g^\vartheta(q_i, q_j) x_i x_j \right)^{a_{ij}} = \prod_{i=1}^n \prod_{j=i+1}^n (1 + g^\vartheta(q_i, q_j) x_i x_j)$ . To compute the generating function of  $d_1(G)$ , we simply set  $x_i = 1$  for all  $i > 1$ . Then

$$\begin{aligned} \mathbb{E}_{\mu^\vartheta} \left( x_1^{d_1(G)} \right) &= \mathbb{E}_{\mu^\vartheta} \left( \mathbb{E}_{\mu^\vartheta} \left( x_1^{d_1(G)} \middle| q_1 \right) \right) \\ &= \mathbb{E}_{\mu^\vartheta} \left( \mathbb{E}_{\mu^\vartheta} \left( \prod_{j=2}^n \frac{1 + g^\vartheta(q_1, q_j) x_1}{1 + g^\vartheta(q_1, q_j)} \middle| q_1 \right) \right) \\ &= \mathbb{E}_{\mu^\vartheta} \left( \left( \mathbb{E}_{\mu^\vartheta} \left( \frac{1 + g^\vartheta(q_1, q_2) x_1}{1 + g^\vartheta(q_1, q_2)} \middle| q_1 \right) \right)^{n-1} \right), \end{aligned}$$

where we have used symmetry and the independence of  $q_1, \dots, q_n$ . Further, note that

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

Hence, for  $g^\vartheta(q_1, q_2)$  small in the sparse graph limit, we can write

$$\begin{aligned}\mathbb{E}_{\mu^\vartheta} \left( \frac{1 + g^\vartheta(q_1, q_2)x_1}{1 + g^\vartheta(q_1, q_2)} \middle| q_1 \right) &= \int_{\mathcal{Q}} \frac{1 + g^\vartheta(q_1, q_2)x_1}{1 + g^\vartheta(q_1, q_j)} \mu^\vartheta(dq_2) \\ &= 1 + (x_1 - 1) \int_{\mathcal{Q}} g^\vartheta(q_1, q_2) \mu^\vartheta(dq_2) + o(1) \\ &= 1 + (x_1 - 1) \nu^\vartheta(q_1) + o(1),\end{aligned}$$

where we have denoted by  $\nu^\vartheta(q) \equiv \int_{\mathcal{Q}} g^\vartheta(q, q') \mu^\vartheta(dq')$ . It then follows that

$$\mathbb{E}_{\mu^\vartheta} \left( x_1^{d_1(G)} \right) = \mathbb{E}_{\mu^\vartheta} \left( \left( 1 + (x_1 - 1) \nu^\vartheta(q_1) \right)^{n-1} \right) (1 + o(1)) = \mathbb{E}_{\mu^\vartheta} \left( e^{(x_1-1)(n-1)\nu^\vartheta(q_1)} \right) (1 + o(1)),$$

where we have used the fact that  $e^{(x_1-1)\nu^\vartheta(q)} = 1 + (x_1 - 1)\nu^\vartheta(q) + o(1)$ . This is the probability generating function of a mixed Poisson random variable with mixing parameter  $\nu^\vartheta(q)$ . In particular, since  $p^\vartheta(q, q') = g^\vartheta(q, q') + o(1)$ , we can write  $n\nu^\vartheta(q) = n \int_{\mathcal{Q}} p(q, q') \mu^\vartheta(dq') = \sum_{j=1}^n \int_{\mathcal{Q}} p^\vartheta(q, q_j) \mu^\vartheta(dq_j) = \sum_{j=1}^n \mathbb{P}(a_{1j} = 1 | q_1 = q) = \mathbb{E}_{\mu^\vartheta} (d_1(G) | q_1 = q)$ , which is the expected degree of a firm with output  $q$ , and we denote it by  $\bar{d}(q)$ . Further, it then follows that

$$\begin{aligned}\mathbb{E}_{\mu^\vartheta} \left( x_1^{d_1(G)} \right) &= \sum_{k=0}^n x_1^k \mathbb{P}(d_1(G) = k) \\ &= \mathbb{E}_{\mu^\vartheta} \left( e^{(x_1-1)\bar{d}(q_1)} \right) (1 + o(1)) \\ &= \mathbb{E}_{\mu^\vartheta} \left( e^{-\bar{d}(q_1)} \sum_{k=0}^n \frac{(x_1 \bar{d}(q_1))^k}{k!} \right) (1 + o(1)) \\ &= \sum_{k=0}^n x_1^k \mathbb{E}_{\mu^\vartheta} \left( \frac{e^{-\bar{d}(q_1)} \bar{d}(q_1)^k}{k!} \right) (1 + o(1)).\end{aligned}$$

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

$$P^\vartheta(k) = \mathbb{P}(d_1(G) = k) = \mathbb{E}_{\mu^\vartheta} \left( \frac{e^{-\bar{d}(q_1)} \bar{d}(q_1)^k}{k!} \right) (1 + o(1)).$$

We now give a proof of part (ii) of the proposition. In the limit of  $\vartheta \rightarrow \infty$  we obtain from the FOC in Equation (44) that

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

This shows that the right hand side of Equation (10) has a point of discontinuity at  $\sqrt{\frac{\zeta}{\rho}}$  (see Figure 1). It then follows that, in the limit of  $\vartheta \rightarrow \infty$  (for the stochastically stable equilibrium), we have

$$q^* = \begin{cases} \frac{\eta}{b+2\nu^*-\rho}, & \text{if } \zeta < \frac{\rho(\eta^*)^2}{(b+2\nu^*)^2}, \\ \left\{ \frac{\eta^*}{b+2\nu^*-\rho}, \frac{\eta^*}{b+2\nu^*} \right\}, & \text{if } \frac{\rho(\eta^*)^2}{(b+2\nu^*)^2} < \zeta < \frac{\rho\eta^2}{(b+2\nu^*-\rho)^2}, \\ \frac{\eta^*}{b+2\nu^*}, & \text{if } \frac{\rho(\eta^*)^2}{(b+2\nu^*-\rho)^2} < \zeta, \end{cases} \quad (56)$$

which is increasing in  $\rho$  and  $\eta^*$ , and decreasing in  $\zeta$  and  $b$  (see Figure 1). Next, note that

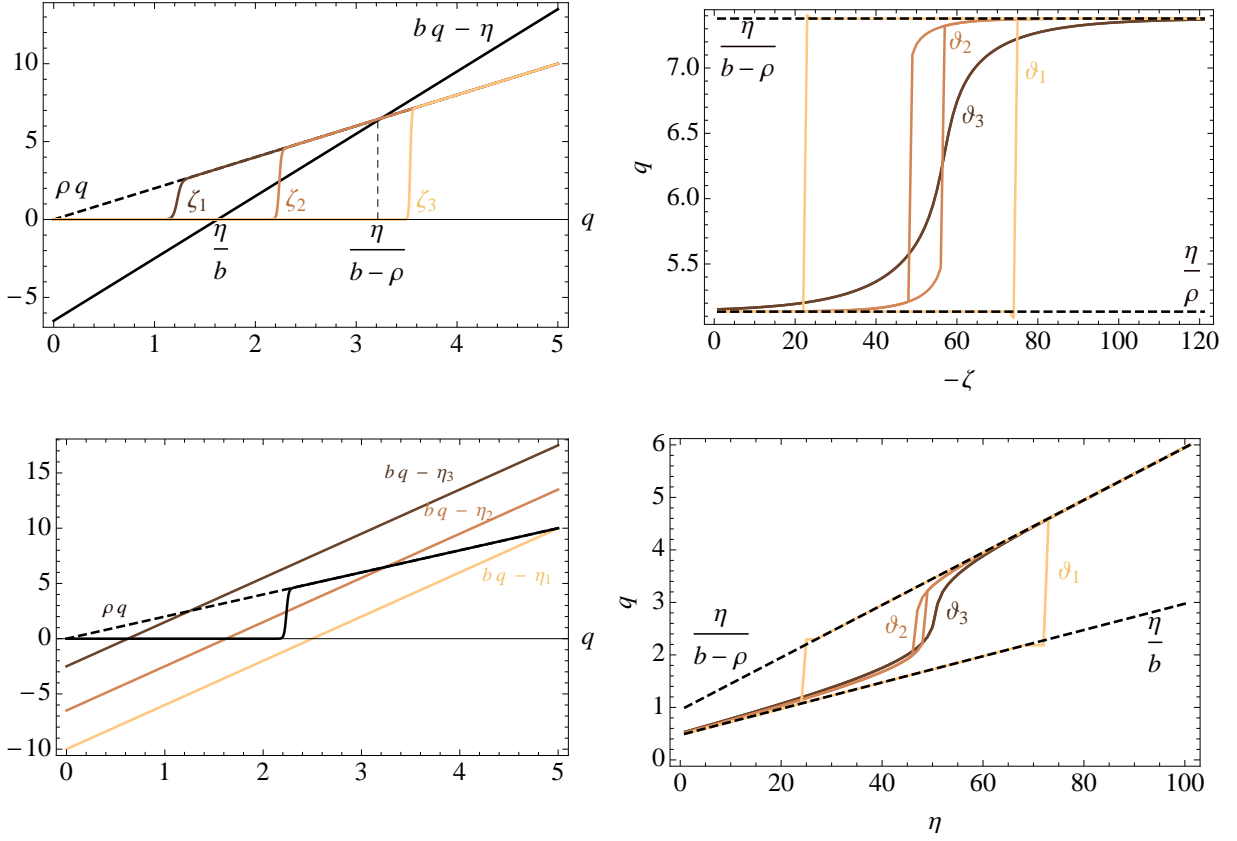


Figure 1: (Top left panel) The right hand side of Equation (10) for different values of  $\zeta_1 = 25$ ,  $\zeta_2 = 10$ ,  $\zeta_3 = 3$  and  $b = 4$ ,  $\rho = 2$ ,  $\eta = 6.5$ ,  $\nu = 0$  and  $\vartheta = 10$ . (Top right panel) The values of  $q$  solving Equation (10) for different values of  $\zeta$  with  $b = 1.48$ ,  $\rho = 0.45$  and  $\vartheta_1 = 49.5$ ,  $\vartheta_2 = 0.495$ ,  $\vartheta_3 = 0.2475$ . (Bottom left panel) The right hand side of Equation (10) 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  $q$  solving Equation (10) for different values of  $\eta$  with  $b = 4$ ,  $\rho = 2$  and  $\vartheta_1 = 10$ ,  $\vartheta_2 = 0.26$ ,  $\vartheta_3 = 0.2$ .

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

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

$$\frac{\partial^2 \ln \mathcal{Z}_\vartheta}{\partial \eta^2} = \frac{1}{\mathcal{Z}_\vartheta} \frac{\partial^2 \mathcal{Z}_\vartheta}{\partial \eta^2} - \frac{1}{\mathcal{Z}_\vartheta^2} \left( \frac{\partial \mathcal{Z}_\vartheta}{\partial \eta} \right)^2 = \frac{1}{\mathcal{Z}_\vartheta} \frac{\partial^2 \mathcal{Z}_\vartheta}{\partial \eta^2} - \left( \frac{\partial \ln \mathcal{Z}_\vartheta}{\partial \eta} \right)^2 = \vartheta^2 \mathbb{E}_{\mu^\vartheta} \left( \sum_{i=1}^n q_i^2 \right) - \vartheta^2 \mathbb{E}_{\mu^\vartheta} \left( \sum_{i=1}^n q_i \right)^2.$$

We then get

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

The variance of the mean is then given by

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

We have that

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

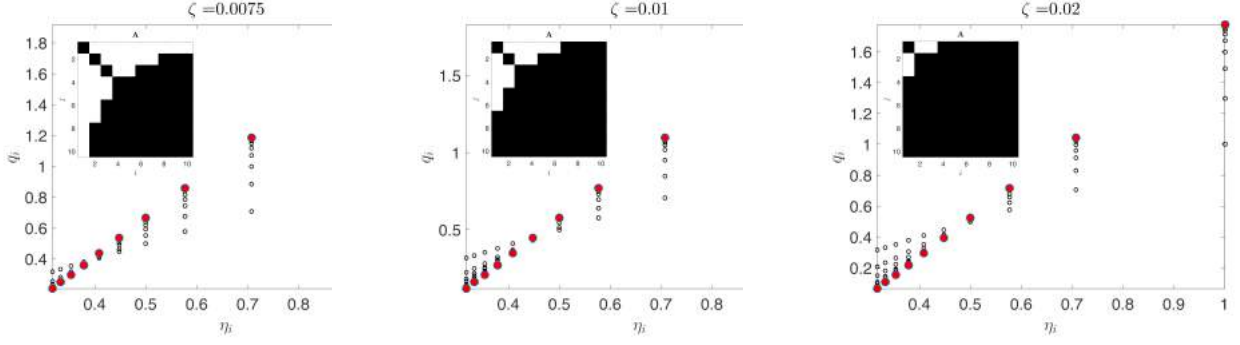


Figure 2: The output iteration of Equation (59) over the firm fixed effects  $\eta_i$ . Filled circles indicate the fixed points. The insets show the adjacency matrix  $\mathbf{A} = (a_{ij})_{1 \leq i, j, n}$  with elements are given by  $a_{ij} = \mathbb{1}_{\{q_i q_j > \frac{\zeta}{\rho}\}}$  (see also Figure 4), where the vector  $\mathbf{q}$  is the fixed point of Equation (59). The panels from the left to the right correspond to increasing linking costs  $\zeta \in \{0.0075, 0.01, 0.02\}$ . The parameters used are  $n = 10$ ,  $\nu = 0.5$ ,  $b = 0.06$ ,  $\rho = 0.02$  and  $\boldsymbol{\eta} = (1.00, 0.71, 0.58, 0.50, 0.45, 0.41, 0.38, 0.35, 0.33, 0.32)^\top$ .

and we get

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

Note that the variance of the average output 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 output is equal to  $q^*$  with probability one in the large  $\vartheta$  limit.

Further, in the limit of  $\vartheta \rightarrow \infty$ , for both, the low equilibrium, where  $q = \frac{\eta^*}{b+2\nu^*}$  and therefore  $\frac{\eta^*}{q} = b + 2\nu^*$ , as well as the high equilibrium, where  $q = \frac{\eta^*}{b+2\nu^*-\rho}$ , and  $\frac{\eta^*}{q} = b + 2\nu^* - \rho$  we find from Equation (50) that

$$\frac{\partial}{\partial \zeta} \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i^2} \Big|_{q_i=q} = 0,$$

and from Equation (51) we get

$$\frac{\partial}{\partial \zeta} \frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j} \Big|_{q_i=q_j=q} = 0.$$

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

We next give the proof of Proposition 3, which generalizes Proposition 2 by allowing for firm heterogeneity.

**PROOF OF PROPOSITION 3.** We first give a proof of part (i) of the proposition. We have that  $\mu^\vartheta(\mathbf{q}, G) = \mu^\vartheta(G|\mathbf{q})\mu^\vartheta(\mathbf{q})$ . Analogous to the proof of Proposition 2 one can show that  $\mu^\vartheta(G|\mathbf{q}) = \prod_{i < j} p^\vartheta(q_i, q_j)^{a_{ij}} (1 - p^\vartheta(q_i, q_j))^{1-a_{ij}}$  where  $p^\vartheta(q_i, q_j)$  is given by Equation (53), which corresponds to an inhomogeneous random graph with linking probability  $p^\vartheta : \mathcal{Q} \times \mathcal{Q} \rightarrow [0, 1]$ . The output distribution is given by

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

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

sion around  $\mathbf{q}^*$  (for large  $\vartheta$ ) we have that

$$\mathcal{H}_\vartheta(\mathbf{q}) = \mathcal{H}_\vartheta(\mathbf{q}^*) + (\mathbf{q} - \mathbf{q}^*)^\top \nabla \mathcal{H}_\vartheta(\mathbf{q}^*) + \frac{1}{2}(\mathbf{q} - \mathbf{q}^*)^\top \Delta \mathcal{H}_\vartheta(\mathbf{q}^*)(\mathbf{q} - \mathbf{q}^*) + o\left(\|\mathbf{q} - \mathbf{q}^*\|^2\right),$$

as  $\vartheta \rightarrow \infty$ , where  $\mathbf{q}^* = \operatorname{argmax}_{\mathbf{q} \in [0, \bar{q}]^n} \mathcal{H}_\vartheta(\mathbf{q})$ , the gradient is  $\nabla \mathcal{H}_\vartheta(\mathbf{q}) = \left(\frac{\partial \mathcal{H}_\vartheta}{\partial q_i}\right)_{i=1, \dots, n}$ , and the Hessian is  $\Delta \mathcal{H}_\vartheta(\mathbf{q}) = \left(\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j}\right)_{i, j=1, \dots, n}$ . As the gradient  $\nabla \mathcal{H}_\vartheta(\mathbf{q})$  vanishes at  $\mathbf{q}^*$ , we have that

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

We then can write

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

Normalization,  $\int_{\mathcal{Q}^n} \mu^\vartheta(\mathbf{q}) d\mathbf{q} = 1$ , implies that

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

The Laplace approximation of  $\mu^\vartheta(\mathbf{q})$  is then given by [Wong, 2001, Theorem 3, p. 495]

$$\mu^\vartheta(\mathbf{q}) = \left(\frac{2\pi}{\vartheta}\right)^{-\frac{n}{2}} |\Delta \mathcal{H}_\vartheta(\mathbf{q}^*)|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\vartheta(\mathbf{q} - \mathbf{q}^*)^\top (-\Delta \mathcal{H}_\vartheta(\mathbf{q}^*))(\mathbf{q} - \mathbf{q}^*)\right\} + o\left(\|\mathbf{q} - \mathbf{q}^*\|^2\right).$$

That is, in the limit of large  $\vartheta$ ,  $\mathbf{q}$  is asymptotically normally distributed with mean  $\mathbf{q}^*$  and variance  $-\frac{1}{\vartheta} \delta \mathcal{H}_\vartheta(\mathbf{q}^*)^{-1}$ , where  $\Delta \mathcal{H}_\vartheta(\mathbf{q}) = \left(\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j}\right)_{i, j=1, \dots, n}$  with  $\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i^2}$  given by Equation (45) while, for any  $i \neq j$ ,  $\frac{\partial^2 \mathcal{H}_\vartheta}{\partial q_i \partial q_j}$  is given by Equation (46).

We next turn to part (ii) of the proposition. We show that the networks  $G$  in the support of the stationary distribution  $\mu^\vartheta(\mathbf{q}, G)$  in the limit of vanishing noise  $\vartheta \rightarrow \infty$  is a nested split graph. A graph  $G$  is a nested split graph if for every node  $i \in \mathcal{N}$  there exist a weight  $x_i$  and a threshold  $T$  such that vertices  $i$  and  $j$  are linked if and only if  $x_i + x_j \geq T$  [Mahadev and Peled, 1995].

In the limit  $\vartheta \rightarrow \infty$  the conditional probability of the network  $G$  can be written as  $\mu^*(G|\mathbf{q}) = \lim_{\vartheta \rightarrow \infty} \mu^\vartheta(G|\mathbf{q}) = \prod_{i < j} \mathbf{1}_{\{\rho q_i q_j > \zeta\}}^{a_{ij}} \mathbf{1}_{\{\rho q_i q_j < \zeta\}}^{1-a_{ij}}$ . Assume that  $G$  is a stochastically stable network, that is for  $G \in \Omega^*$ , we must have that  $\mu^*(\mathbf{q}, G) = \lim_{\vartheta \rightarrow \infty} \mu^\vartheta(\mathbf{q}, G) > 0$ . Since,  $\mu^*(\mathbf{q}, G) = \mu^*(G|\mathbf{q})\mu^*(\mathbf{q})$  this implies that  $\mu^*(G|\mathbf{q}) > 0$ . It follows that  $\rho q_i q_j > \zeta$  for all  $a_{ij} = 1$  and  $\rho q_i q_j < \zeta$  for all  $a_{ij} = 0$ . We then define the weights  $x_i \equiv \log q_i$ ,  $x_j \equiv \log q_j$  and a threshold  $T \equiv \log \zeta - \log \rho$ , and conclude that  $G$  is a nested split graph (or threshold graph, see also supplementary Appendix B).

Moreover, the output distribution is given by  $\mu^\vartheta(\mathbf{q}) = \frac{1}{\mathcal{Z}_n^\vartheta} \sum_{G \in \mathcal{G}^n} e^{\vartheta \Phi(\mathbf{q}, G)} = \frac{1}{\mathcal{Z}_n^\vartheta} e^{\vartheta \mathcal{H}_\vartheta(\mathbf{q})}$ , where the Hamiltonian is given by Equation (41). The output profile that maximizes the Hamiltonian can be found from the FOC,  $\frac{\partial \mathcal{H}_\vartheta}{\partial q_i} = 0$ , from which we get

$$q_i = \frac{\eta_i}{2\nu} + \frac{1}{2\nu} \sum_{j \neq i}^n \left( \frac{\rho}{2} \left( 1 + \tanh \left( \frac{\vartheta}{2} (\rho q_i q_j - \zeta) \right) \right) - b \right) q_j.$$

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

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

we thus obtain<sup>55</sup>

$$q_i = \frac{\eta_i}{2\nu} + \frac{1}{2\nu} \sum_{j \neq i}^n q_j \left( \rho \mathbf{1}_{\{q_i q_j > \frac{\zeta}{\rho}\}} - b \right). \quad (57)$$

Note that for any profile of output levels  $\mathbf{q}$  there exists a unique nested split graph with adjacency matrix  $\mathbf{A} = (a_{ij})_{1 \leq i, j, n}$  whose elements are given by  $a_{ij} = \mathbf{1}_{\{q_i q_j > \frac{\zeta}{\rho}\}}$ . Then we can write Equation (57) as follows

$$q_i = \frac{\eta_i}{2\nu} + \frac{\rho}{2\nu} \sum_{j \neq i}^n a_{ij} q_j - \frac{b}{2\nu} \sum_{j \neq i}^n q_j. \quad (58)$$

Moreover, assume that  $\eta_i > \eta_j$ , then we want to show that  $q_j > q_i$  for the output profile  $\mathbf{q}$  solving Equation (57). For this purpose we consider the iteration

$$q_{i,t+1} = f_i(\mathbf{q}_t) \equiv \max \left\{ 0, \frac{\eta_i}{2\nu} + \frac{\rho}{2\nu} \sum_{j \neq i}^n q_{j,t} \mathbf{1}_{\{q_{i,t} q_{j,t} > \frac{\zeta}{\rho}\}} - \frac{b}{2\nu} \sum_{j \neq i}^n q_{j,t} \right\}, \quad (59)$$

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

$$q_{i,t+1} - q_{j,t+1} = f_i(\mathbf{q}_t) - f_j(\mathbf{q}_t) = \frac{b}{2\nu} (\eta_i - \eta_j) + \frac{\rho}{2\nu} \sum_{k \in \mathcal{N}_{i,t} \setminus \mathcal{N}_{j,t}} q_{k,t} + \frac{b}{2\nu} (q_{i,t} - q_{j,t}) > 0,$$

where we have used the fact that the condition  $q_{i,t} q_{j,t} > \frac{\zeta}{\rho}$  for  $i$  and  $j$  being linked represents a nested split graph, and for such a graph if  $q_{i,t} > q_{j,t}$  (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  $t \rightarrow \infty$  it holds for the fixed point  $\mathbf{q}$  solving Equation (57).

We now give a proof of part (iii) of the proposition. From Equation (17) we know that in the stochastically stable state the output levels satisfy the following equation

$$\mathbf{g}(\mathbf{q}) \equiv (\mathbf{I}_n + b\mathbf{B} - \rho\mathbf{A}) \mathbf{q} = \boldsymbol{\eta}, \quad (60)$$

where  $\mathbf{B}$  is a matrix of ones with zero diagonal and  $\mathbf{A}$  has elements  $a_{ij} = \mathbf{1}_{\{\rho q_i q_j > \zeta\}}$ . When the  $(\eta_i)_{i=1}^n$  are real valued random variables with probability density function  $f$ , then the probability density function  $\mu$  of  $\mathbf{q}$  is given by

$$\mu(\mathbf{q}) = \left| \det \left( \frac{d\mathbf{g}(\mathbf{q})}{d\mathbf{q}} \right) \right| f(\boldsymbol{\eta}),$$

where  $\left( \frac{d\mathbf{g}(\mathbf{q})}{d\mathbf{q}} \right)_{ij} = \frac{\partial \mathbf{g}_i(\mathbf{q})}{\partial q_j}$ . From Equation (60) we get  $\frac{d\mathbf{g}(\mathbf{q})}{d\mathbf{q}} = \mathbf{I}_n + b\mathbf{B} - \rho\mathbf{A}$ , and we denote this by  $\mathbf{M}$ . It then follows that  $\mu(\mathbf{q}) = |\det(\mathbf{M})| f(\mathbf{M}\mathbf{q})$ . In particular, if the  $(\eta_i)_{i=1}^n$  are identically and independently Pareto distributed with density function  $f(\eta) = (\gamma - 1)\eta^{-\gamma}$  for  $\eta \geq 1$  then

$$\mu(\mathbf{q}) = (\gamma - 1)^n |\det(\mathbf{M})| \prod_{i=1}^n (\mathbf{M}\mathbf{q})_i^{-\gamma}.$$

Next, consider  $\mathbf{q} = c\mathbf{u}$ ,  $c > 0$ , with  $\mathbf{u}$  being a vector of all ones. Then  $\mathbf{M} = \mathbf{I}_n + (b - \rho)\mathbf{B}$  for  $c$  large enough (because  $\rho q_i q_j = c^2 \rho > \zeta$  for  $c$  large enough), and  $\det(\mathbf{M}) = (1 + (n - 1)(b - \rho)) (1 +$

<sup>55</sup>Note that  $\mathcal{H}_\theta(\mathbf{q})$  is a real valued function that converges pointwise and whose derivatives converge uniformly on a closed interval  $[0, \bar{q}]$  so that we can exchange the derivative with the limit [Rudin, 1987].

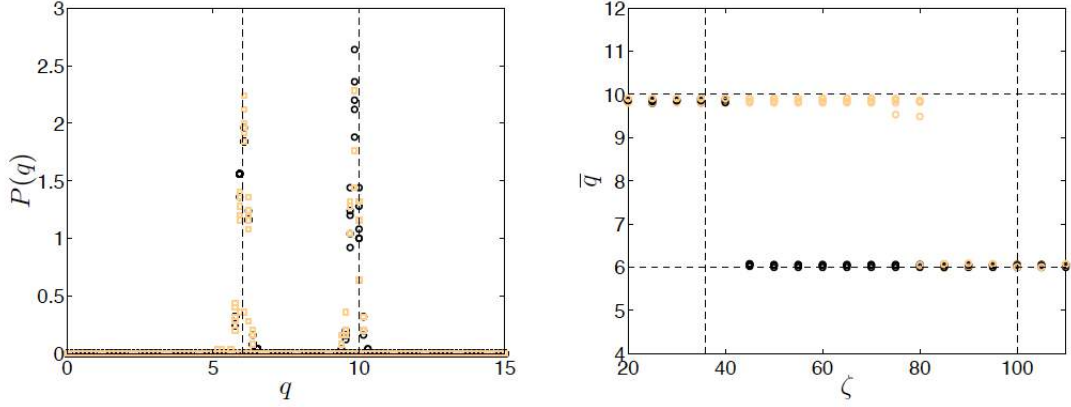


Figure 3: (Left panel) The stationary output distribution. The vertical dashed lines indicate the theoretical predictions from Equation (56). (Right panel) The average output level from numerical simulations with  $\vartheta = 1$  starting with different initial conditions (indicated with different colors). The horizontal dashed lines indicate the equilibrium quantities and the vertical dashed lines the threshold cost levels from Equation (13). In the region of the cost  $\zeta$  between the lower and upper thresholds two equilibria exist.

$b + \rho)^{n-1}$ . Further,  $(\mathbf{M}\mathbf{q})_i = 1 + (n-1)(b - \rho)$ , so that we can write

$$\mu(\mathbf{c}\mathbf{u}) = (1 + (n-1)(b - \rho)) (1 + b + \rho)^{n-1} (\gamma - 1)^n (1 + (n-1)(b - \rho))^{-n\gamma} c^{-n\gamma},$$

and we conclude that  $\mu(\mathbf{c}\mathbf{u}) \sim \prod_{i=1}^n O(c^{-\gamma})$  as  $c \rightarrow \infty$ .  $\square$

Figure 2 shows the output iteration of Eq. (59) over the firm fixed effects  $\eta_i$  together with the adjacency matrix  $\mathbf{A} = (a_{ij})_{1 \leq i, j, n}$  (see also Figure 4) whose elements are given by  $a_{ij} = \mathbf{1}_{\{q_i q_j > \frac{\zeta}{\rho}\}}$  and the vector  $\mathbf{q}$  is the fixed point of Equation (59). We observe that firms with higher  $\eta_i$  also have higher output 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  $\zeta$ .

An illustration with the average output level from numerical simulations starting with different initial conditions and a comparison with the predictions of Equation (56) can be seen in Figure 3.

**PROOF OF PROPOSITION 4.** We first give a proof of part (i) of the proposition. Welfare can be written as follows

$$\begin{aligned} W(\mathbf{q}) &= U(\mathbf{q}) + \Pi(\mathbf{q}, G) \\ &= \frac{1}{2} \sum_{i=1}^n q_i^2 + \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j + \sum_{i=1}^n \left( \eta q_i - \nu q_i^2 - b \sum_{j \neq i}^n q_i q_j + \rho \sum_{j=1}^n a_{ij} (q_i q_j - \zeta) \right) \\ &= \eta \sum_{i=1}^n q_i - \frac{2\nu - 1}{2} \sum_{i=1}^n q_i^2 - \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i}^n q_i q_j + \sum_{i=1}^n \sum_{j \neq i}^n a_{ij} (\rho q_i q_j - \zeta). \end{aligned}$$

The only network dependent part in  $W(\mathbf{q}, G)$  is the last term  $\sum_{i=1}^n \sum_{j \neq i}^n a_{ij} (\rho q_i q_j - \zeta)$ . For a given output vector  $\mathbf{q}$  the network that maximizes this term is a nested split graph  $G$  (see also supplementary Appendix B) where each link  $ij \in G$  if and only if  $\rho q_i q_j > \zeta$ .<sup>56</sup> Hence, we can

<sup>56</sup>A graph  $G$  is a nested split graph if for every node  $i \in \mathcal{N}$  there exist a weight  $x_i$  and a threshold  $T$  such that vertices  $i$  and  $j$  are linked if and only if  $x_i + x_j \geq T$  [Mahadev and Peled, 1995]. Then by letting  $\ln q_i = x_i$  and  $\ln(\zeta/\rho) = T$  yields  $\rho q_i q_j > \zeta$  iff  $x_i + x_j \geq T$  the conclusion follows.



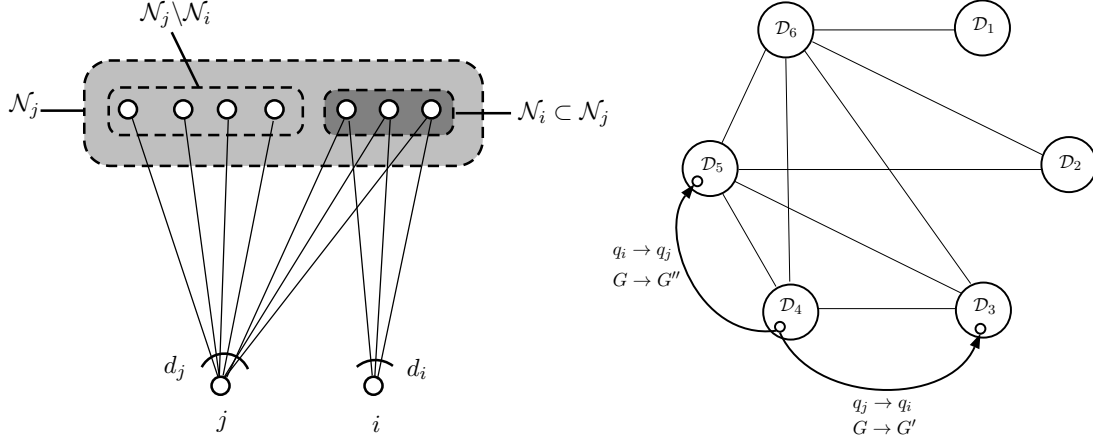


Figure 4: (Left panel) An illustration of the neighborhoods of  $i, j$  such that  $\mathcal{N}_i \subset \mathcal{N}_j$ . (Right panel) A schematic representation of a nested split graph  $G$  (see supplementary Appendix B).

write welfare reduced to this class of networks as follows

$$W(\mathbf{q}) = \eta \sum_{i=1}^n q_i - \frac{2\nu - 1}{2} \sum_{i=1}^n q_i^2 - \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i} q_i q_j + \sum_{i=1}^n \sum_{j \neq i} (\rho q_i q_j - \zeta) \mathbf{1}_{\{\rho q_i q_j > \zeta\}}.$$

The necessary first order condition (FOC) can be written as follows

$$\frac{\partial W}{\partial q_i} = \eta - (2\nu - 1)q_i - b \sum_{j \neq i} q_j + \rho \sum_{j \neq i} q_j \mathbf{1}_{\{\rho q_i q_j > \zeta\}} = 0. \quad (61)$$

We next consider a symmetric solution  $q_i = q$  for all  $i = 1, \dots, n$  with the property that  $\rho q^2 > \zeta$ . Then Equation (61) implies that

$$\eta - (2\nu - 1)q + (2\rho - b)(n - 1)q = 0,$$

from which we deduce that

$$q = \frac{\eta^*}{b + 2\nu^* - 2\rho - \frac{1}{n-1}},$$

where we have denoted by  $\eta^* = \eta/(n - 1)$  and  $\nu^* = \nu/(n - 1)$ . The corresponding network is a complete graph,  $K_n$ . Further, for a symmetric solution with  $\rho q^2 < \zeta$  from Equation (61) we must have that

$$\eta + (1 - 2\nu)q - b(n - 1)q = 0,$$

from which we deduce that

$$q = \frac{\eta^*}{b + 2\nu^* - \frac{1}{n-1}}.$$

The corresponding network is an empty graph,  $\overline{K}_n$ . Note that welfare in the complete network  $K_n$  and the empty network  $\overline{K}_n$  is the same if  $\zeta = \zeta^*$  where  $\zeta^*$  is given in Equation (21).

Consider a degree partition  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_m$  (see also supplementary Appendix B) in a nested split graph  $G$  such that  $d_i < d_j$  if  $i \in \mathcal{D}_k, j \in \mathcal{D}_{k'}$  with  $k' > k$ . From the symmetry of the FOC in Equation (61) it follows that for any  $i, j \in \mathcal{D}_k$  it must hold that  $q_i = q_j$ . Moreover, from the FOC we also observe that when  $d_i < d_j$  then  $q_i < q_j$ .

Further, we show that either the complete network  $K_n$  or the empty network  $\overline{K}_n$  are efficient. To do so, assume that  $G$  is efficient, and it is neither empty nor complete. Let the output profile in  $G$  be  $\mathbf{q}$ . We know that the efficient network is a nested split graph. Consider  $i, j$  such that  $q_j > q_i$  and  $\mathcal{N}_i \subset \mathcal{N}_j$ . Let  $G'$  be the graph obtained from  $G$  with the links of  $j$  in  $\mathcal{N}_j \setminus \mathcal{N}_i$  removed. Further, let  $\mathbf{q}' = (q_i, \dots, q_{j-1}, q_i, q_{j+1}, \dots, q_n)$ , that is,  $\mathbf{q}'$  is obtained from  $\mathbf{q}$  by replacing  $q_j$  with

$q_i$ . An illustration can be seen in the left panel in Figure 4. We then have that

$$W(\mathbf{q}', G') - W(\mathbf{q}, G) = (q_i - q_j) \left[ \eta + \frac{1 - 2\nu}{2}(q_i + q_j) - b \sum_{k \neq j} q_k + 2\rho \sum_{k \in \mathcal{N}_i} q_k \right] - 2 \sum_{k \in \mathcal{N}_j \setminus \mathcal{N}_i} (\rho q_j q_k - \zeta).$$

Similarly, consider the graph  $G''$  obtained from  $G$  with the links in  $\mathcal{N}_j \setminus \mathcal{N}_i$  added to  $i$ . Further, let  $\mathbf{q}'' = (q_i, \dots, q_{i-1}, q_j, q_{i+1}, \dots, q_n)$ , that is,  $\mathbf{q}''$  is obtained from  $\mathbf{q}$  by replacing  $q_i$  with  $q_j$ . Then we have that

$$W(\mathbf{q}'', G'') - W(\mathbf{q}, G) = (q_j - q_i) \left[ \eta + \frac{1 - 2\nu}{2}(q_i + q_j) - b \sum_{k \neq j} q_k + 2\rho \sum_{k \in \mathcal{N}_i} q_k \right] + 2 \sum_{k \in \mathcal{N}_j \setminus \mathcal{N}_i} (\rho q_j q_k - \zeta).$$

It follows that  $W(\mathbf{q}'', G'') - W(\mathbf{q}, G) + W(\mathbf{q}', G') - W(\mathbf{q}, G) = 0$ . There are three possible cases to consider such that this equality holds. First, if  $W(\mathbf{q}'', G'') - W(\mathbf{q}, G) < 0$  then we must have that  $W(\mathbf{q}', G') - W(\mathbf{q}, G) > 0$ . This means that  $W(\mathbf{q}', G') > W(\mathbf{q}, G)$  and  $(\mathbf{q}, G)$  is not efficient. Second, if  $W(\mathbf{q}', G') - W(\mathbf{q}, G) < 0$  then we must have that  $W(\mathbf{q}'', G'') - W(\mathbf{q}, G) > 0$ . This means that  $W(\mathbf{q}'', G'') > W(\mathbf{q}, G)$  and  $(\mathbf{q}, G)$  is not efficient. The third case to consider is  $W(\mathbf{q}'', G'') - W(\mathbf{q}, G) = 0$ . Then we must have that  $W(\mathbf{q}'', G'') = W(\mathbf{q}', G') = W(\mathbf{q}, G)$ .

From the definition of  $\mathbf{q}''$  and  $G''$  we see that the transition from  $G$  to  $G''$  corresponds to moving a node from a degree partition  $\mathcal{D}_k$  to a partition  $\mathcal{D}_l$  with  $l > k$ , while leaving welfare unchanged. An illustration can be seen in the right panel in Figure 4. In particular,  $(\mathbf{q}'', G'')$  must be efficient. We then can repeat this procedure to move up the node to the next higher partition while leaving welfare unchanged. Doing this across all partitions and for all nodes shows that we end up with the complete graph,  $K_n$ , with the same welfare as the original graph  $G$ . However, this is a contradiction to our initial assumption that the complete graph is not efficient. This shows that either the empty,  $\bar{K}_n$ , or the complete graph,  $K_n$ , must be efficient.

We next consider part (ii) of the proposition. In the heterogeneous firms case, welfare can be written as

$$W(\mathbf{q}) = \sum_{i=1}^n \eta_i q_i - \frac{2\nu - 1}{2} \sum_{i=1}^n q_i^2 - \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i} q_i q_j + \sum_{i=1}^n \sum_{j \neq i} a_{ij} (\rho q_i q_j - \zeta).$$

The only network dependent part in  $W(\mathbf{q}, G)$  is the last term  $\sum_{i=1}^n \sum_{j \neq i} a_{ij} (\rho q_i q_j - \zeta)$ , and, as in part (i), for a given output vector  $\mathbf{q}$  the network that maximizes this term is a nested split graph  $G$  where each link  $ij \in G$  if and only if  $\rho q_i q_j > \zeta$ . Moreover, from the necessary FOC we obtain

$$q_i = f_i(\mathbf{q}) \equiv \max \left\{ 0, \frac{\eta_i}{2\nu - 1} - \frac{b}{2\nu - 1} \sum_{j \neq i} q_j + \frac{\rho}{2\nu - 1} \sum_{j \neq i} q_j \mathbf{1}_{\{\rho q_i q_j > \zeta\}} \right\}. \quad (62)$$

We can compare this to the equilibrium output levels of Equation (57), which were given by

$$q_i = g_i(\mathbf{q}) \equiv \max \left\{ 0, \frac{\eta_i}{2\nu} - \frac{b}{2\nu} \sum_{j \neq i} q_j + \frac{\rho}{2\nu} \sum_{j \neq i} q_j \mathbf{1}_{\{\rho q_i q_j > \zeta\}} \right\}. \quad (63)$$

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

$$f_i(\mathbf{q}) - g_i(\mathbf{q}) = \left( \frac{1}{2\nu - 1} - \frac{1}{2\nu} \right) \left( \eta_i - b \sum_{j \neq i} q_j + \rho \sum_{j \neq i} q_j \mathbf{1}_{\{\rho q_i q_j > \zeta\}} \right) \geq 0.$$

Next, consider the differential equations  $\frac{dx}{dt} = \mathbf{f}(\mathbf{x}) - \mathbf{x}$  and  $\frac{dy}{dt} = \mathbf{g}(\mathbf{y}) - \mathbf{y}$ , both with initial condition  $\mathbf{x}_0 = \mathbf{y}_0 = (0, \dots, 0)^\top$ . Because  $\mathbf{f}(\mathbf{x}) > \mathbf{g}(\mathbf{x})$ , the comparison lemma implies that  $\mathbf{x}(t) > \mathbf{y}(t)$  for all  $t \geq 0$  (see Khalil [2002], Lemma 3.4). In particular, we can conclude that the

fixed point  $\mathbf{f}(\mathbf{q}) = \mathbf{q}$  must be higher than the fixed point  $\mathbf{g}(\mathbf{q}) = \mathbf{q}$ . That is, in the stochastically stable equilibrium output levels are too low compared to the social optimum. Moreover, because a link is only present if  $\rho q_i q_j > \zeta$  there are fewer links in the stochastically stable network than in the efficient network.  $\square$

## Supplement to “Network Formation with Local Complements and Global Substitutes: The Case of R&D Networks”

Chih-Sheng Hsieh<sup>a</sup>, Michael D. König<sup>b</sup>, Xiaodong Liu<sup>c</sup>

<sup>a</sup>*Department of Economics, Chinese University of Hong Kong, CUHK Shatin, Hong Kong, China.*

<sup>b</sup>*Department of Economics, University of Zurich, Schönberggasse 1, CH-8001 Zurich, Switzerland.*

<sup>c</sup>*Department of Economics, University of Colorado Boulder, Boulder, Colorado 80309-0256, United States.*

### B. Definitions and Characterizations

A *network (graph)*  $G$  is the pair  $(\mathcal{N}, \mathcal{E})$  consisting of a set of *nodes (vertices)*  $\mathcal{N} = \{1, \dots, n\}$  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  $\mathcal{N}_i^{(2)} = \bigcup_{j \in \mathcal{N}_i} \mathcal{N}_j \setminus (\mathcal{N}_i \cup \{i\})$  denote the second-order neighbors of node  $i$ . Similarly, the  $k$ -th order neighborhood of node  $i$  is defined recursively from  $\mathcal{N}_i^{(0)} = \{i\}$ ,  $\mathcal{N}_i^{(1)} = \mathcal{N}_i$  and  $\mathcal{N}_i^{(k)} = \bigcup_{j \in \mathcal{N}_i^{(k-1)}} \mathcal{N}_j \setminus \left( \bigcup_{l=0}^{k-1} \mathcal{N}_i^{(l)} \right)$ . A *walk* in  $G$  of length  $k$  from  $i$  to  $j$  is a sequence  $\langle i_0, i_1, \dots, i_k \rangle$  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  $\bar{K}_n$ ) when all its nodes are isolated.

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

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

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  $\mathcal{S}$  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  $d_{(1)} < d_{(2)} < \dots < d_{(k)}$ , 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  $i = 0, \dots, k$ . Then the set-valued vector  $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \dots, \mathcal{D}_k)$  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  $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \dots, \mathcal{D}_k)$  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  $\mathcal{D}_i$ ,  $i = 1, \dots, \lfloor \frac{k}{2} \rfloor$  and a dominating set  $\bigcup_{i=\lfloor \frac{k}{2} \rfloor + 1}^k \mathcal{D}_i$  in the graph  $G' = (\mathcal{N} \setminus \mathcal{D}_0, \mathcal{E})$ . Moreover, the

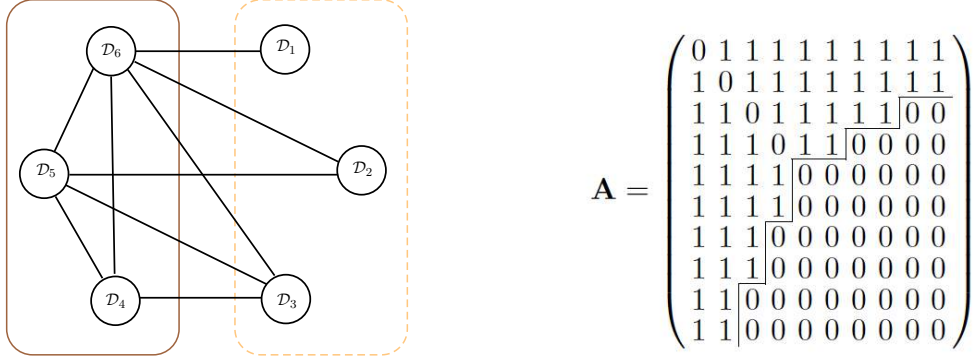


Figure 1: (Left panel) Representation of nested split graphs and their degree partitions  $\mathcal{D}$  with corresponding adjacency matrices  $\mathbf{A}$ . A line between  $\mathcal{D}_i$  and  $\mathcal{D}_j$  indicates that every node in  $\mathcal{D}_i$  is adjacent to every node in  $\mathcal{D}_j$ . The partitions included in the solid frame ( $\mathcal{D}_i$  with  $\lfloor \frac{k}{2} \rfloor + 1 \leq i \leq k$ ) are the dominating subsets while the partitions in the dashed frame ( $\mathcal{D}_i$  with  $1 \leq i \leq \lfloor \frac{k}{2} \rfloor$ ) are the independent sets with  $k = 6$ . The illustration follows Mahadev and Peled [1995, p. 11]. (Right panel) The corresponding stepwise adjacency matrix  $\mathbf{A}$  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  $h \leq i$ .

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  $i = 1, \dots, \lfloor \frac{k}{2} \rfloor$  if  $i = 1, \dots, k$ , while  $\mathcal{N}_v = \bigcup_{j=1}^i \mathcal{D}_{k+1-j} \setminus \{v\}$  if  $i = \lfloor \frac{k}{2} \rfloor + 1, \dots, k$ . See also the left panel in Figure 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  $\bar{K}_n$ . 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  $\bar{G}$  with the same nodes as  $G$  such that any two nodes of  $\bar{G}$  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  $\bar{K}_n$ .

Let  $\mathbf{A}$  be the symmetric  $n \times n$  *adjacency matrix* of the network  $G$ . The element  $a_{ij} \in \{0, 1\}$  indicates if there exists a link between nodes  $i$  and  $j$  such that  $a_{ij} = 1$  if  $(i, j) \in \mathcal{E}$  and  $a_{ij} = 0$  if  $(i, j) \notin \mathcal{E}$ . The  $k$ -th power of the adjacency matrix is related to walks of length  $k$  in the graph. In particular,  $(\mathbf{A}^k)_{ij}$  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  $\lambda_1, \lambda_2, \dots, \lambda_n$  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  $\lambda_i$  for  $i = 1, \dots, n$ . 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  $i = 1, \dots, n$ . Moreover, if  $\mathbf{v}_i$  and  $\mathbf{v}_j$  are eigenvectors for different eigenvalues,  $\lambda_i \neq \lambda_j$ , then  $\mathbf{v}_i$  and  $\mathbf{v}_j$  are orthogonal, i.e.  $\mathbf{v}_i^\top \mathbf{v}_j = 0$  if  $i \neq j$ . In particular,  $\mathbb{R}^n$  has an orthonormal basis consisting of eigenvectors of  $\mathbf{A}$ . Since  $\mathbf{A}$  is a real symmetric matrix, there exists an orthogonal matrix  $\mathbf{S}$  such that  $\mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{S} \mathbf{S}^\top = \mathbf{I}$  (that is  $\mathbf{S}^\top = \mathbf{S}^{-1}$ ) and  $\mathbf{S}^\top \mathbf{A} \mathbf{S} = \mathbf{D}$ , where  $\mathbf{D}$  is the diagonal matrix of eigenvalues of  $\mathbf{A}$  and the columns of  $\mathbf{S}$  are the corresponding eigenvectors. The *Perron-Frobenius eigenvalue*  $\lambda_{\text{PF}}(G)$  is the *largest real eigenvalue* of  $\mathbf{A}$  associated with  $G$ , i.e. all eigenvalues  $\lambda_i$  of  $\mathbf{A}$  satisfy  $|\lambda_i| \leq \lambda_{\text{PF}}(G)$  for  $i = 1, \dots, n$  and there exists an associated nonnegative eigenvector  $\mathbf{v}_{\text{PF}} \geq 0$  such that  $\mathbf{A}\mathbf{v}_{\text{PF}} = \lambda_{\text{PF}}(G)\mathbf{v}_{\text{PF}}$ . For a connected graph  $G$  the adjacency matrix  $\mathbf{A}$  has a unique largest real eigenvalue  $\lambda_{\text{PF}}(G)$  and a positive associated

eigenvector  $\mathbf{v}_{\text{PF}} > 0$ . 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  $(\mathbf{A}^k)_{ii}$  and the total number of closed walks of length  $k$  in  $G$  is  $\text{tr}(\mathbf{A}^k) = \sum_{i=1}^n (\mathbf{A}^k)_{ii} = \sum_{i=1}^n \lambda_i^k$ . We further have that  $\text{tr}(\mathbf{A}) = 0$ ,  $\text{tr}(\mathbf{A}^2)$  gives twice the number of links in  $G$  and  $\text{tr}(\mathbf{A}^3)$  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  $h \leq i$ . See also the right panel in Figure 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 which maximize the largest eigenvalue,  $\lambda_{\text{PF}}(G)$ , [Brualdi and Solheid, 1986], and they are the ones that maximize the degree variance [Peled et al., 1999]. See e.g. König et al. [2014a] 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  $G(n, (p_{ij}))$  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  $1 \leq i < j \leq n$  [Bollobás et al., 2007; Söderberg, 2002].

Given a set of attributes  $\mathcal{A} = \{a_1, a_2, \dots, a_m\}$ , a vertex  $v$  is associated with the set  $\mathcal{S}(v)$  of attributes selected by  $v$  from  $\mathcal{A}$ . Let the vertices  $v_1, \dots, v_n$  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  $\mathcal{N} = \{v_1, \dots, v_n\}$  defined by this adjacency relationship is then called the *random intersection graph*  $G(n, m, s)$  [Deijfen and Kets, 2009].

### C. Cournot Competition and Profits from R&D Collaborations

We consider a Cournot oligopoly game in which a set  $\mathcal{N} = \{1, \dots, n\}$  of firms is competing in a homogeneous product market.<sup>57</sup> We assume that firms are not only competitors in the product market, but they can also form pairwise collaborative agreements.<sup>58</sup> These pairwise links involve a commitment to share R&D results and thus lead to lower marginal cost of production of the collaborating firms.<sup>59</sup> 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

<sup>57</sup>Generalizations to Bertrand competition are straight forward [see König et al., 2014b; Westbrock, 2010].

<sup>58</sup>Such R&D collaborations often involve competing firms, as for example a strategic alliance between *Pfizer* and *Bayer*, both operating in the pharmaceuticals sector (with primary standard industry classification code 2834) to develop treatments for obesity, type 2 diabetes and other related disorders in the year 2006 illustrates.

<sup>59</sup>R&D partnerships have become a widespread phenomenon characterizing technological dynamics, especially in industries with rapid technological development such as, for instance, the pharmaceutical and chemical industries [see e.g. Hagedoorn, 2002; Powell et al., 2005; Roijakkers and Hagedoorn, 2006]. In these industries firms have become more specialized in specific domains of a technology and they tend to combine their knowledge with that of other firms that are specialized in different domains in order to jointly generate innovations that can help to develop new products or reduce their production costs [Ahuja, 2000; Powell et al., 1996]. For example, Bernstein [1988] finds that R&D spillovers decrease the unit costs of production for a sample of Canadian firms. Similarly Belderbos et al. [2004] find evidence for production cost reductions due to R&D collaborations using data on a large sample of Dutch innovating firms.

$n$  nodes, each firm sets an R&D effort level unilaterally.<sup>60</sup> Given the effort levels  $e_i \geq 0$ , marginal cost  $c_i \geq 0$  of firm  $i$  is given by [Spence, 1984]<sup>61,62</sup>

$$c_i(\mathbf{e}, G) = \bar{c}_i - \alpha e_i - \beta \sum_{j=1}^n a_{ij} e_j, \quad (64)$$

where  $a_{ij} = 1$  if firms  $i$  and  $j$  set up a collaboration (0 otherwise) and  $a_{ii} = 0$ . 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.<sup>63</sup> We further allow for ex ante heterogeneity among firms in the variable cost  $\bar{c}_i \geq 0$ , for  $i = 1, \dots, n$  [see also Banerjee and Duflo, 2005], expressing their different technological and organizational capabilities.<sup>64</sup>

Moreover, we also assume that firms incur a direct cost  $\gamma e_i^2$ ,  $\gamma \geq 0$ , for their R&D efforts and a fixed cost  $\zeta \geq 0$  for each R&D collaboration.<sup>65</sup> The profit of firm  $i$ , given the R&D network  $G$  and the quantities  $0 \leq q_i \leq \bar{q}$  and efforts  $0 \leq e_i \leq \bar{e}$ , is then given by

$$\pi_i(\mathbf{q}, \mathbf{e}, G) = (p_i - c_i)q_i - \gamma e_i^2 - \zeta d_i, \quad (65)$$

with vectors  $\mathbf{q} = (q_i)_{i=1}^n$  and  $\mathbf{e} = (e_i)_{i=1}^n$ . Inserting marginal costs from Equation (64) gives

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

The first-order condition with respect to R&D effort  $e_i$  is given by  $\frac{\partial \pi_i(\mathbf{q}, \mathbf{e}, G)}{\partial e_i} = \alpha q_i - 2\gamma e_i = 0$ . Solving for  $e_i$  and taking into account that  $e_i \in [0, \bar{e}]$  delivers  $e_i^* = \min\{\lambda q_i, \bar{e}\}$ , where we have denoted by  $\lambda = \frac{\alpha}{2\gamma}$ .<sup>66</sup> This equation can be viewed as reflecting learning-by-doing effects on R&D efforts. It reflects various empirical studies which have found that the R&D effort of a firm is correlated with its output or size [Cohen and Klepper, 1996a,b]. We then can write marginal

<sup>60</sup>See also Kamien et al. [1992] for a similar model of competitive R&D in which firms unilaterally choose their R&D effort levels.

<sup>61</sup>Note that we have neglected spillovers among non-collaborating firms. For an extension incorporating this additional spillover channel see König et al. [2014b].

<sup>62</sup>This generalizes earlier studies such as the one by D'Aspremont and Jacquemin [1988] where spillovers were assumed to take place between all firms in the industry and no distinction between collaborating and non-collaborating firms was made.

<sup>63</sup>Note that in this model, firms are exposed to business stealing effects if their rivals increase their output via cost reducing R&D collaborations.

<sup>64</sup>Blundell et al. [1995] argued that because the main source of unobserved heterogeneity in models of innovation lies in the different knowledge stocks with which firms enter a sample, a variable that approximates the build-up of firm knowledge at the time of entering the sample is a particularly good control for unobserved heterogeneity.

<sup>65</sup>In Section 2.4 we discuss several extensions of the model including heterogeneous linking costs.

<sup>66</sup>König et al. [2014b] show that with  $q_i \in [0, \bar{q}]$  we must have that  $0 \leq e_i \leq q_i \leq \bar{q}$ , and requiring that  $\min_{i \in \mathcal{N}} \bar{c}_i > \bar{q}(1 + \beta(n-1))$ , implies that the best response effort level of firm  $i$  is given by  $e_i^* = \lambda q_i$ .

costs from Equation (64) as follows<sup>67</sup>

$$c_i(\mathbf{e}^*(\mathbf{q}), G) = \bar{c} - \lambda\alpha q_i - \lambda\beta \sum_{j=1}^n a_{ij}q_j. \quad (66)$$

Profits can be written as

$$\pi_i(\mathbf{q}, G) \equiv \pi_i(\mathbf{q}, \mathbf{e}^*(\mathbf{q}), G) = p_i q_i - \bar{c}q_i - \lambda\alpha q_i^2 + \lambda\beta q_i \sum_{j=1}^n a_{ij}q_j - \lambda^2\gamma q_i^2 - \zeta d_i. \quad (67)$$

Next we consider the demand for goods produced by firm  $i$ . A representative consumer maximizes [Singh and Vives, 1984]

$$U(I, q_1, \dots, q_n) = I + a \sum_{i=1}^n q_i - \frac{1}{2} \sum_{i=1}^n q_i^2 - \frac{b}{2} \sum_{i=1}^n \sum_{j \neq i} q_i q_j, \quad (68)$$

with the budget constraint  $I + \sum_{i=1}^n q_i \leq E$  and endowment  $E$ . The parameter  $a$  captures the total size of the market, whereas  $b \in (0, 1]$ , measures the degree of substitutability between products. In particular,  $b = 1$  depicts a market of perfect substitutable goods, while  $b \rightarrow 0$  represents the case of almost independent markets.<sup>68</sup> The constraint is binding and the utility maximization of the representative consumer gives the inverse demand function for firm  $i$

$$p_i = a - q_i - b \sum_{j \neq i} q_j. \quad (69)$$

Firm  $i$  then sets its quantity,  $q_i$ , in order to maximize its profit,  $\pi_i$ , given by Equation (67). We also assume that there is a maximum production capacity  $\bar{q}$  such that  $q_i \leq \bar{q}$  for all  $i \in \mathcal{N}$ . Inserting marginal cost from Equation (66) and inverse demand from Equation (69) we can write firm  $i$ 's profit as

$$\pi_i(\mathbf{q}, G) = (a - \bar{c}_i)q_i - (1 - \lambda\alpha + \lambda^2\gamma)q_i^2 - bq_i \sum_{j \neq i} q_j + \lambda\beta \sum_{j=1}^n a_{ij}q_i q_j - \zeta d_i. \quad (70)$$

In the following we will denote by  $\eta_i \equiv a - \bar{c}_i$ ,  $\nu \equiv 1 - \lambda\alpha + \lambda^2\gamma$  and  $\rho \equiv \lambda\beta$ , so that Equation (70) becomes [Ballester et al., 2006]

$$\pi_i(\mathbf{q}, G) = \underbrace{\eta_i q_i - \nu q_i^2}_{\text{own concavity}} \quad \underbrace{- bq_i \sum_{j \neq i} q_j}_{\text{global substitutability}} \quad + \quad \underbrace{\rho q_i \sum_{j=1}^n a_{ij} q_j}_{\text{local complementarity}} \quad - \zeta d_i,$$

which is exactly Equation (1) in Section 2.

<sup>67</sup>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. In Section 2.2 we will, however, introduce noise in the optimal output and collaboration decisions, since these depend on the decisions of all other firms in the industry and their characteristics, which might be harder to observe.

<sup>68</sup>Observe that we do not ex ante impose any restrictions on the parameter  $b$ , in particular, we do not require that  $b = 1$  nor  $b = 0$ .



## D. Multinomial Logit Output Choice

Consider a discretization (partition)  $\mathcal{Q}_N = \{0, \Delta q, 2\Delta q, \dots, \bar{q}\}$  of the interval  $\mathcal{Q} = [0, \bar{q}]$  into  $N$  subintervals with length  $\Delta q = \bar{q}/N$  for some (large)  $N$ . Let the profit of firm  $i$  from choosing an output level  $q_i \in \mathcal{Q}_N$  be given by  $\pi_i(q_i, \mathbf{q}_{-i}, G) + \varepsilon_i$ . When the error term  $\varepsilon_i$  is independently and identically type-I extreme value distributed with parameter  $\vartheta$  we get [Anderson et al., 1992]:<sup>69,70</sup>

$$\mathbb{P} \left( q_i = \operatorname{argmax}_{q'_i \in \mathcal{Q}_N} \{ \pi_i(q_i, \mathbf{q}_{-i}, G) + \varepsilon_i \} \right) = \frac{e^{\vartheta \pi_i(q_i, \mathbf{q}_{-i}, G)}}{\sum_{q'_i \in \mathcal{Q}_N} e^{\vartheta \pi_i(q'_i, \mathbf{q}_{-i}, G)}}.$$

Assume that the output adjustment rate is given by  $\chi N/\bar{q} > 0$ . Then

$$\begin{aligned} \mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = (q_i, \mathbf{q}_{-it}, G_t) | \boldsymbol{\omega}_t = (\mathbf{q}_t, G_t)) &= \frac{\chi N}{\bar{q}} \mathbb{P} \left( q_i = \operatorname{argmax}_{q'_i \in \mathcal{Q}_N} \{ \pi_i(q'_i, \mathbf{q}_{-i}, G) + \varepsilon \} \right) \Delta t + o(\Delta t) \\ &= \frac{\chi N}{\bar{q}} \frac{e^{\vartheta \pi_i(q_i, \mathbf{q}_{-it}, G_t)}}{\sum_{q'_i \in \mathcal{Q}_N} e^{\vartheta \pi_i(q'_i, \mathbf{q}_{-it}, G_t)}} \Delta t + o(\Delta t) \\ &= \chi \frac{e^{\vartheta \pi_i(q_i, \mathbf{q}_{-it}, G_t)}}{\sum_{q'_i \in \mathcal{Q}_N} e^{\vartheta \pi_i(q'_i, \mathbf{q}_{-it}, G_t)} \Delta q} \Delta t + o(\Delta t). \end{aligned} \quad (71)$$

Using the trapezoidal rule we can write the sum in the denominator in the last line of Equation (71) as follows [see e.g. Atkinson, 1989]

$$\sum_{q'_i \in \mathcal{Q}_N} e^{\vartheta \pi_i(q'_i, \mathbf{q}_{-it}, G_t)} \Delta q = \int_{\mathcal{Q}} e^{\vartheta \pi_i(q', \mathbf{q}_{-it}, G_t)} dq' + \frac{\Delta q}{2} \left( e^{\vartheta \pi_i(0, \mathbf{q}_{-it}, G_t)} + e^{\vartheta \pi_i(\bar{q}, \mathbf{q}_{-it}, G_t)} \right) + o(\Delta q). \quad (72)$$

Then in the limit of  $N \rightarrow \infty$ , respectively,  $\Delta q \downarrow 0$ , we can write

$$\mathbb{P}(\boldsymbol{\omega}_{t+\Delta t} = (q_i, \mathbf{q}_{-it}, G_t) | \boldsymbol{\omega}_t = (\mathbf{q}_t, G_t)) = \chi \frac{e^{\vartheta \pi_i(q, \mathbf{q}_{-it}, G_t)}}{\int_{\mathcal{Q}} e^{\vartheta \pi_i(q', \mathbf{q}_{-it}, G_t)} dq'} \Delta t + o(\Delta t),$$

which is exactly Equation (3). See also Anderson et al. [2004, 2001]; Ben-Akiva and Watanatada [1981]; McFadden [1976] for further discussion. The transition probability then states that with increasing values of  $\vartheta$  (lower levels of noise), firms choose output levels with higher probability that yield higher profits.

## E. Firm Heterogeneity

In the following sections we will discuss three possible extensions of the model that incorporate firm heterogeneity. First, in Section E.1 we allow for heterogeneous collaborations costs. Second, in Section E.2 we incorporate heterogeneous spillovers between collaborating firms.

<sup>69</sup>For a type-I extreme value distributed random variable  $\varepsilon$  we have that  $\mathbb{P}(\varepsilon \leq c) = e^{-e^{c/\zeta-\gamma}}$ , where  $\gamma \approx 0.58$  is Euler's constant. The mean is  $\mathbb{E}(\varepsilon) = 0$  and the variance is given by  $\operatorname{Var}(\varepsilon) = \frac{\pi^2 \zeta^2}{6}$ .

<sup>70</sup>For a computationally efficient implementation of the multinomial logit model see D'Haultfoeuille and Iaria [2016].

## 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  $\zeta_i(G) = \sum_{j=1}^n a_{ij}\psi_{ij}$  (see Equation (23)), where the function  $\psi_{ij}$  is additively separable  $\psi_{ij} = s_i + s_j$ , and we assume that the cost  $s_i \geq 0$  is proportional to the inverse of the firm's productivity, that is  $s_i = \frac{1}{\phi_i}$ , where  $\phi_i > 0$  is the productivity (or efficiency) of firm  $i$  [similar to e.g. Melitz, 2003; Melitz et al., 2008]. This implies that firms with higher productivity incur lower collaboration costs. The probability of a link between firms  $i$  and  $j$  is then given by Equation (16):<sup>71</sup>

$$p^\vartheta(q_i, s_i, q_j, s_j) = \frac{e^{\vartheta(\rho q_i q_j - s_i - s_j)}}{1 + e^{\vartheta(\rho q_i q_j - s_i - s_j)}}. \quad (73)$$

Next, we assume that the firms' productivities,  $\phi_i \geq c > 0$ , are Pareto distributed [König et al., 2016; Melitz et al., 2008], with density  $f(x) = \frac{\gamma}{c} \left(\frac{c}{x}\right)^{\gamma+1}$  with  $x > c$ , where  $c > 0$  is a lower-cut-off and  $\gamma > 0$  is a positive parameter. The complementary distribution function is then given by  $F(x) = 1 - \left(\frac{c}{x}\right)^\gamma$ . It follows that the cost  $s = \frac{1}{\phi}$  has the density  $f(s) = \gamma c^\gamma s^{\gamma-1}$  for  $s \in (0, \frac{1}{c})$ , and the cumulative distribution function  $F(s) = (cs)^\gamma$ .<sup>72</sup>

**Proposition 5.** *Assume that the firms output levels are concentrated on  $q^*$  in the limit of  $\vartheta \rightarrow \infty$ , then under the continuum approximation,<sup>73</sup> the degree distribution is given by*

$$P(k) = \frac{c^\gamma}{k} \left( \rho(q^*)^2 - \left( \frac{k}{(n-1)c^\gamma} \right)^{\frac{1}{\gamma}} \right)^{\gamma-1} \left( \frac{k}{(n-1)c^\gamma} \right)^{\frac{1}{\gamma}}, \quad (74)$$

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

PROOF OF PROPOSITION 5. The generating function of the degree  $d_1(G)$  is given by

$$\begin{aligned} \mathbb{E}\left(x_1^{d_1(G)}\right) &= \mathbb{E}\left(\mathbb{E}\left(x_1^{d_1(G)} \mid q_1, s_1\right)\right) \\ &= \mathbb{E}\left(\left(\mathbb{E}\left(\frac{1 + p(q_1, s_1, q_2, s_2)x_1}{1 + p(q_1, s_1, q_2, s_2)} \mid q_1, s_1\right)\right)^{n-1}\right), \end{aligned}$$

<sup>71</sup>Similar specifications can be found in the empirical literature on network formation [Graham, 2015]. For example, Graham [2014] 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  $\mathbb{P}(a_{ij} = 1) = \frac{e^{X_i + X_j + \mathbf{Z}_{ij}^\top \beta}}{1 + e^{X_i + X_j + \mathbf{Z}_{ij}^\top \beta}}$  where  $X_i$  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  $\mathbb{P}(a_{ij} = 1) = \frac{e^{X_i + X_j}}{1 + e^{X_i + X_j}}$ .

<sup>72</sup>In the following propositions we will assume that the firms output levels are concentrated on  $q^*$  in the limit of  $\vartheta \rightarrow \infty$ , and 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 that has been analyzed in the main part of the paper.

<sup>73</sup>This is an approximation which has shown to be accurate in various network formation models as the network size becomes large [Dorogovtsev and Mendes, 2013, pp. 117].

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

$$\begin{aligned} \mathbb{E} \left( x_1^{d_1(G)} \middle| q_1 = q, s_1 = s \right) &= \mathbb{E} \left( \frac{1 + p(q, s, q_2, s_2) x_1}{1 + p(q, s, q_2, s_2)} \middle| q, s \right)^{n-1} \\ &= (1 + (x_1 - 1) \mathbb{E}(p(q, s, q_2, s_2) | q_1 = q, s_1 = s))^{n-1} \\ &= \left( 1 + (x_1 - 1) \int_{\mathcal{Q}} dq' \mu^\vartheta(q') \int ds' \gamma c^\gamma (s')^{\gamma-1} p(q, s, q', s') \right)^{n-1} \\ &= \left( 1 + (x_1 - 1) \int_{\mathcal{Q}} dq' \mu^\vartheta(q') \int ds' \gamma c^\gamma (s')^{\gamma-1} \frac{e^{\vartheta(\rho q q' - s - s')}}{1 + e^{\vartheta(\rho q q' - s - s')}} \right)^{n-1}. \end{aligned}$$

In the limit of  $\vartheta \rightarrow \infty$  in Equation (73) we obtain

$$\lim_{\vartheta \rightarrow \infty} p^\vartheta(q, s, q', s') = \lim_{\vartheta \rightarrow \infty} \frac{e^{\vartheta(\rho q q' - s - s')}}{1 + e^{\vartheta(\rho q q' - s - s')}} = \mathbf{1}_{\{\rho q q' > s + s'\}},$$

so that we can write

$$\begin{aligned} \mathbb{E} \left( x_1^{d_1(G)} \middle| q_1 = q, s_1 = s \right) &= \left( 1 + (x_1 - 1) \int_{\mathcal{Q}} dq' \mu^\vartheta(q') \int ds' \gamma c^\gamma (s')^{\gamma-1} \mathbf{1}_{\{\rho q q' > s + s'\}} \right)^{n-1} \\ &= \left( 1 + (x_1 - 1) \gamma c^\gamma \int_0^{\rho q q^* - s} ds' (s')^{\gamma-1} \right)^{n-1} \\ &= (1 + (x_1 - 1) c^\gamma (\rho q q^* - s)^\gamma)^{n-1} \\ &= e^{(x_1 - 1)(n-1) c^\gamma (\rho q q^* - s)^\gamma}. \end{aligned}$$

This is the generating function of a Poisson random variable with expectation and variance given by  $\bar{d}(q, s) \equiv (n-1) c^\gamma (\rho q q^* - s)^\gamma$ . 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

$$\mathbb{P}(d_1(G) = k | q_1 = q, s_1 = s) = \delta(k - \bar{d}(q, s)) = \delta(k - (n-1) c^\gamma (\rho q q^* - s)^\gamma).$$

Note that under the continuum approximation there exists a one-to-one mapping from the degree  $k$  to the cost  $s$ , where for a given  $k$  and output  $q$ , the cost  $s$  is given by

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

Using the fact that<sup>74</sup>

$$\delta(k - (n-1) c^\gamma (\rho(q^*)^2 - s)^\gamma) = \delta \left( s - \left( \rho(q^*)^2 - \left( \frac{k}{(n-1) c^\gamma} \right)^{\frac{1}{\gamma}} \right) \right) \frac{1}{\gamma k} \left( \frac{k}{(n-1) c^\gamma} \right)^{\frac{1}{\gamma}}. \quad (75)$$

---

<sup>74</sup>When  $g(x)$  is a continuously differentiable function in  $\mathbb{R}$  it holds that  $\delta(g(x)) = \sum_{i=1}^m \frac{\delta(x-x_i)}{|g'(x_i)|}$  where the  $m$  roots  $x_i$  satisfy  $g(x_i) = 0$  for all  $i = 1, \dots, m$ .

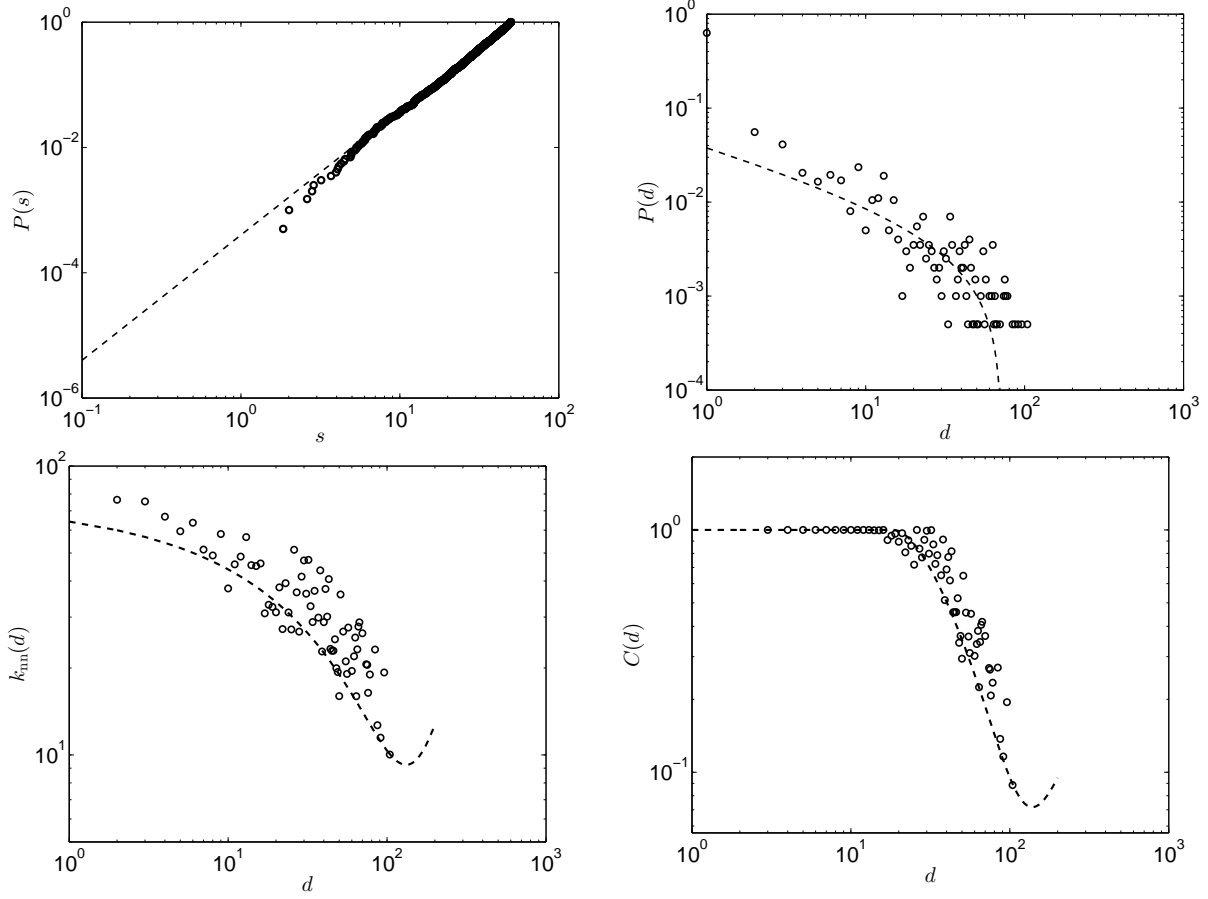


Figure 1: (Top left panel) The empirical and the theoretical cumulative cost distribution  $F(s) = (cs)^\gamma$  with  $\gamma = 2$  and  $c = 0.02$ . The empirical distribution is indicated with circles and the theoretical distribution with a dashed line. (Top right panel) The degree distribution  $P(d)$ . The dashed line indicates the theoretical prediction of Equation (74). (Bottom left panel) The average nearest neighbor degree distribution  $k_{nn}(d)$ , decreasing with increasing degrees  $d$  and thus indicating a dissortative network. The dashed line indicates the theoretical prediction of Equation (76). (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.

and assuming that the output distribution concentrates on  $q^*$ , the degree distribution is given by

$$\begin{aligned}
P(k) &= \int ds \mathbb{P}(d_1(G) = k | q_1 = q^*, s_1 = s) f(s) \\
&= \gamma c^\gamma \int ds \delta \left( s - (n-1)c^\gamma (\rho(q^*)^2 - s)^\gamma \right) s^{\gamma-1} \\
&= \gamma c^\gamma \int ds \delta \left( k - \left( \rho(q^*)^2 - \left( \frac{k}{(n-1)c^\gamma} \right)^{\frac{1}{\gamma}} \right) \right) \frac{1}{\gamma k} \left( \frac{k}{(n-1)c^\gamma} \right)^{\frac{1}{\gamma}} s^{\gamma-1} \\
&= \frac{c^\gamma}{k} \left( \rho(q^*)^2 - \left( \frac{k}{(n-1)c^\gamma} \right)^{\frac{1}{\gamma}} \right)^{\gamma-1} \left( \frac{k}{(n-1)c^\gamma} \right)^{\frac{1}{\gamma}} \\
&= O \left( k^{-\frac{\gamma-1}{\gamma}} \right).
\end{aligned}$$

□

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 1 for the case of  $\gamma = 2$  and  $n = 200$  firms.

**Proposition 6.** *Assume that the firms output levels are concentrated on  $q^*$  in the limit of  $\vartheta \rightarrow \infty$ , then under the continuum approximation, the average nearest neighbor degree distribution is given by*

$$k_{\text{nn}}(k) = 1 + \frac{(n-1)^2 \gamma c^{2\gamma}}{k} \int_0^{\left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}} ds' (s'^{\gamma-1} (\rho(q^*)^2 - s'^\gamma)). \quad (76)$$

PROOF OF PROPOSITION 6. Next we compute the average nearest neighbor degree distribution

$$k_{\text{nn}}(k) = 1 + \frac{1}{P(k)} \int ds \int_{\mathcal{Q}} dq f(s) \mu^\vartheta(q) g(k|q, s) \tilde{k}_{\text{nn}}(q, s),$$

where

$$\begin{aligned} \tilde{k}_{\text{nn}}(q, s) &= \int ds' \int_{\mathcal{Q}} dq' p(q', s'|q, s) \bar{d}(q', s'), \\ g(k|q, s) &= \mathbb{P}(d_1(G) = k | q_1 = q, s_1 = s) = \delta(k - \bar{d}(q, s)), \\ \bar{d}(q, s) &= (n-1)c^\gamma (\rho q q^* - s)^\gamma \\ p(q', s'|q, s) &= \frac{(n-1)p(q, s, q', s') f(s') \mu^\vartheta(q')}{\bar{d}(q, s)}, \\ \lim_{\vartheta \rightarrow \infty} p^\vartheta(q, s, q', s') &= \lim_{\vartheta \rightarrow \infty} \frac{e^{\vartheta(\rho q q' - s - s')}}{1 + e^{\vartheta(\rho q q' - s - s')}} = \mathbf{1}_{\{\rho q q' > s + s'\}} \\ \mu^\vartheta(q) &= \delta(q - q^*) \\ f(s) &= \gamma c^\gamma s^{\gamma-1}. \end{aligned}$$

It then follows that

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

and therefore

$$\begin{aligned} \tilde{k}_{\text{nn}}(q, s) &= \frac{n-1}{\bar{d}(q, s)} \int ds' f(s') \int dq' \delta(q' - q^*) \mathbf{1}_{\{\rho q q' > s + s'\}} \bar{d}(q', s') \\ &= \frac{n-1}{\bar{d}(q, s)} \int ds' f(s') \mathbf{1}_{\{\rho q q^* > s + s'\}} \bar{d}(q^*, s') \\ &= \frac{n-1}{\bar{d}(q, s)} \int_0^{\rho q q^* - s} ds' f(s') \bar{d}(q^*, s') \\ &= c^{-\gamma} (\rho q q^* - s)^{-\gamma} \int_0^{\rho q q^* - s} ds' \gamma c^\gamma (s')^{\gamma-1} (n-1) c^\gamma (\rho(q^*)^2 - s')^{-\gamma} \\ &= \frac{(n-1) \gamma c^\gamma J(q, s)}{(\rho q q^* - s)^\gamma}, \end{aligned}$$

where we have denoted by

$$J(q, s) \equiv \int_0^{\rho q q^* - s} ds' (\rho(q^*)^2 - s')^\gamma (s')^{\gamma-1}.$$

We then get

$$\begin{aligned}
k_{\text{nn}}(k) &= 1 + \frac{1}{P(k)} \int ds \int_{\mathcal{Q}} dq f(s) \delta(q - q^*) \delta(k - \bar{d}(q, s)) \tilde{k}_{\text{nn}}(q, s) \\
&= 1 + \frac{1}{P(k)} \int ds \gamma c^\gamma s^{\gamma-1} \delta(k - \bar{d}(q^*, s)) \tilde{k}_{\text{nn}}(q^*, s) \\
&= 1 + \frac{1}{P(k)} \int ds f(s) \delta(k - \bar{d}(q^*, s)) \frac{(n-1)\gamma c^\gamma J(q^*, s)}{(\rho q q^* - s)^\gamma}.
\end{aligned}$$

Using Equation (75) we can write this as

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

□

Figure 1 shows the results from numerical simulations compared with the theoretical prediction of Equation (76).

**Proposition 7.** *Assume that the firms output levels are concentrated on  $q^*$  in the limit of  $\vartheta \rightarrow \infty$ , then under the continuum approximation, the clustering coefficient is given by*

$$\begin{aligned}
C(k) &= \mathbf{1}_{\{k < (n-1)\left(\frac{\rho(q^*)^2 c}{2}\right)^\gamma\}} + \mathbf{1}_{\{k > (n-1)\left(\frac{\rho(q^*)^2 c}{2}\right)^\gamma\}} \frac{(n-1)c^\gamma}{k} \\
&\quad \times \left( 1 + \gamma \frac{(n-1)c^\gamma}{k} \int_{\rho(q^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}}^{\left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}} ds' (s')^{\gamma-1} (\rho(q^*)^2 - s')^\gamma \right), \quad (77)
\end{aligned}$$

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

**PROOF OF PROPOSITION 7.** Next we analyze the clustering coefficient of a firm with degree  $k$ , which can be written as

$$\begin{aligned}
C(k) &= \frac{1}{P(k)} \int ds \int_{\mathcal{Q}} dq f(s) \delta(q - q^*) g(k|q, s) \tilde{C}(q, s) \\
&= \frac{1}{P(k)} \int ds f(s) g(k|q^*, s) \tilde{C}(q^*, s),
\end{aligned}$$

where

$$\tilde{C}(q^*, s) = \int ds' \int ds'' \int_{\mathcal{Q}} dq' \int_{\mathcal{Q}} dq'' p(q', s', q'', s'') p(q', s'|q^*, s) p(q'', s''|q^*, s).$$

This can further be written as follows

$$\begin{aligned}
\tilde{C}(q^*, s) &= \int ds' \int ds'' \int_{\mathcal{Q}} dq' \int_{\mathcal{Q}} dq'' \mathbf{1}_{\{\rho q' q'' > s' + s''\}} \\
&\times \frac{(n-1) \mathbf{1}_{\{\rho q^* q' > s + s'\}} f(s') \delta(q' - q^*)}{\bar{d}(q^*, s)} \frac{(n-1) \mathbf{1}_{\{\rho q^* q'' > s + s''\}} f(s'') \delta(q'' - q^*)}{\bar{d}(q^*, s)} \\
&= \frac{(n-1)^2}{\bar{d}(q^*, s)^2} \int ds' f(s') \int ds'' f(s'') \mathbf{1}_{\{\rho(q^*)^2 > s' + s''\}} \mathbf{1}_{\{\rho(q^*)^2 > s + s'\}} \mathbf{1}_{\{\rho(q^*)^2 > s + s''\}} \\
&= \frac{(n-1)^2}{\bar{d}(q^*, s)^2} \left( \mathbf{1}_{\left\{s > \frac{\rho(q^*)^2}{2}\right\}} \int_0^{\rho(q^*)^2 - s} ds' f(s') \int_0^{\rho(q^*)^2 - s} ds'' f(s'') \right. \\
&\quad \left. + \mathbf{1}_{\left\{s < \frac{\rho(q^*)^2}{2}\right\}} \left( \int_0^s ds' f(s') \int_0^{\rho(q^*)^2 - s} ds'' f(s'') + \int_s^{\rho(q^*)^2 - s} ds' f(s') \int_0^{\rho(q^*)^2 - s'} ds'' f(s'') \right) \right).
\end{aligned}$$

We then get (see also Figure 2)

$$\begin{aligned}
\tilde{C}(q^*, s) &= \frac{\gamma^2 c^{2\gamma} (n-1)^2}{\bar{d}(q^*, s)} \left( \mathbf{1}_{\left\{s < \frac{\rho(q^*)^2}{2}\right\}} \int_0^{\rho(q^*)^2 - s} ds'(s')^{\gamma-1} \int_0^{\rho(q^*)^2 - s} ds''(s'')^{\gamma-1} \right. \\
&\quad \left. + \mathbf{1}_{\left\{s > \frac{\rho(q^*)^2}{2}\right\}} \left( \int_0^s ds'(s')^{\gamma-1} \int_0^{\rho(q^*)^2 - s} ds''(s'')^{\gamma-1} + \int_s^{\rho(q^*)^2 - s} ds'(s')^{\gamma-1} \int_0^{\rho(q^*)^2 - s'} ds''(s'')^{\gamma-1} \right) \right) \\
&= \frac{\gamma c^{2\gamma} (n-1)^2}{\bar{d}(q^*, s)} \left( \mathbf{1}_{\left\{s < \frac{\rho(q^*)^2}{2}\right\}} \int_0^{\rho(q^*)^2 - s} ds'(s')^{\gamma-1} (\rho(q^*)^2 - s)^\gamma \right. \\
&\quad \left. + \mathbf{1}_{\left\{s > \frac{\rho(q^*)^2}{2}\right\}} \left( \int_0^s ds'(s')^{\gamma-1} (\rho(q^*)^2 - s)^\gamma + \int_s^{\rho(q^*)^2 - s} ds'(s')^{\gamma-1} (\rho(q^*)^2 - s')^\gamma \right) \right) \\
&= \frac{\gamma c^{2\gamma} (n-1)^2}{\bar{d}(q^*, s)} \left( \mathbf{1}_{\left\{s < \frac{\rho(q^*)^2}{2}\right\}} \frac{1}{\gamma} (\rho(q^*)^2 - s)^{2\gamma} \right. \\
&\quad \left. + \mathbf{1}_{\left\{s > \frac{\rho(q^*)^2}{2}\right\}} \left( \frac{1}{\gamma} s^\gamma (\rho(q^*)^2 - s)^\gamma + \int_s^{\rho(q^*)^2 - s} ds'(s')^{\gamma-1} (\rho(q^*)^2 - s')^\gamma \right) \right) \\
&= \frac{c^{2\gamma} (n-1)^2}{\bar{d}(q^*, s)} \left( \mathbf{1}_{\left\{s < \frac{\rho(q^*)^2}{2}\right\}} (\rho(q^*)^2 - s)^{2\gamma} + \mathbf{1}_{\left\{s > \frac{\rho(q^*)^2}{2}\right\}} (s^\gamma (\rho(q^*)^2 - s)^\gamma + \gamma J(s)) \right), \tag{78}
\end{aligned}$$

where we have denoted by

$$J(s) \equiv \int_s^{\rho(q^*)^2 - s} ds'(s')^{\gamma-1} (\rho(q^*)^2 - s')^\gamma.$$

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

$$\tilde{C}(q^*, s) = \frac{1}{(\rho(q^*)^2 - s)^{2\gamma}} \left( \mathbf{1}_{\left\{s < \frac{\rho(q^*)^2}{2}\right\}} (\rho(q^*)^2 - s)^{2\gamma} + \mathbf{1}_{\left\{s > \frac{\rho(q^*)^2}{2}\right\}} (s^\gamma (\rho(q^*)^2 - s)^\gamma + \gamma J(s)) \right). \tag{79}$$

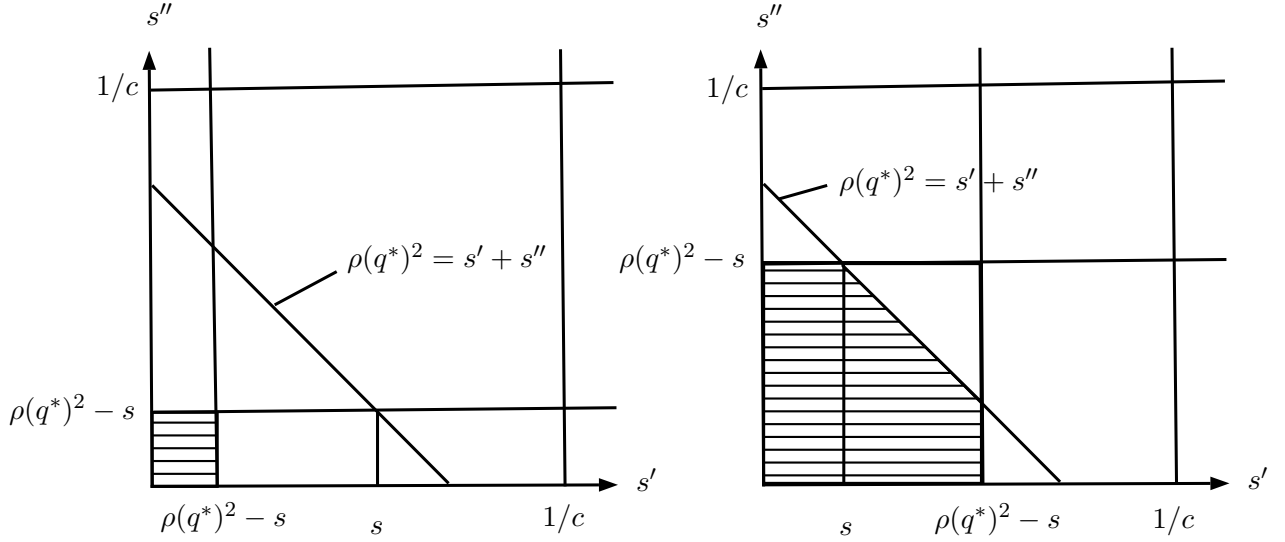


Figure 2: The area of integration in Equation (78) for the case of  $s < \frac{\rho(q^*)^2}{2}$  (left panel) and for the case of  $s > \frac{\rho(q^*)^2}{2}$  (right panel).

Hence we get

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

Inserting Equation (79) this gives

$$\begin{aligned}
C(k) &= \mathbf{1}_{\{k < (n-1)\left(\frac{\rho(q^*)^2 c}{2}\right)^\gamma\}} + \mathbf{1}_{\{k > (n-1)\left(\frac{\rho(q^*)^2 c}{2}\right)^\gamma\}} \left(\frac{(n-1)c^\gamma}{k} + \gamma \left(\frac{(n-1)c^\gamma}{k}\right)^2 J\left(\rho(q^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}\right)\right) \\
&= \mathbf{1}_{\{k < (n-1)\left(\frac{\rho(q^*)^2 c}{2}\right)^\gamma\}} + \mathbf{1}_{\{k > (n-1)\left(\frac{\rho(q^*)^2 c}{2}\right)^\gamma\}} \left(\frac{(n-1)c^\gamma}{k} \right. \\
&\quad \left. + \gamma \left(\frac{(n-1)c^\gamma}{k}\right)^2 \int_{\rho(q^*)^2 - \left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}}^{\left(\frac{k}{(n-1)c^\gamma}\right)^{\frac{1}{\gamma}}} ds' (s')^{\gamma-1} (\rho(q^*)^2 - s')^\gamma\right)
\end{aligned}$$

□

Figure 1 shows the results from numerical simulations compared with the theoretical prediction of Equation (77). The figure further illustrates that the model can generate two-vertex and three-vertex degree correlations, such as a decreasing average nearest neighbor connectivity,  $k_{nn}(d)$ , indicating a dissortative 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



[Griffith et al., 2003]. We assume that the technologies embodied in a firm  $i \in \mathcal{N} = \{1, \dots, n\}$  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  $k \in \{1, \dots, N\}$  or not. We introduce a spillover function  $f : \mathcal{H}^N \times \mathcal{H}^N \rightarrow \mathbb{R}$  capturing the potential technology transfer between any pairs of firms. A possible specification is one in which  $f(\mathbf{h}_i, \mathbf{h}_j) = \mathbf{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle > s\}}$ , 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  $s > 0$  technologies in common. This is an instance of a *random intersection graph* [Deijfen and Kets, 2009] (see also Appendix B).<sup>75</sup>

Given the spillover function  $f(\mathbf{h}_i, \mathbf{h}_j)$ , the marginal cost of production of a firm  $i$  becomes

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

and profits of firm  $i$  are given by

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

The optimal effort levels are given by  $e_i = \frac{\alpha}{2\gamma} q_i = \lambda q_i$ . Inserting into profits yields

$$\begin{aligned} \pi_i &= (a - \bar{c})q_i - (1 - \lambda\alpha + \lambda^2\gamma)q_i^2 - bq_i \sum_{j \neq i} q_j + \lambda\beta q_i \sum_{j=1}^n a_{ij} f(\mathbf{h}_i, \mathbf{h}_j) q_j - \zeta d_i \\ &= \eta q_i - \nu q_i^2 - bq_i \sum_{j \neq i} q_j + \rho q_i \sum_{j=1}^n a_{ij} f(\mathbf{h}_i, \mathbf{h}_j) q_j - \zeta d_i. \end{aligned}$$

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

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

The stationary distribution (see Theorem 1) is given by

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

The probability of observing a network  $G \in \mathcal{G}^n$ , given an output distribution  $\mathbf{q} \in [0, \bar{q}]^n$  and

<sup>75</sup>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.  $d(\mathbf{h}_i, \mathbf{h}_j) = |\mathbf{h}_i \setminus \mathbf{h}_j| \times |\mathbf{h}_j \setminus \mathbf{h}_i| = |\mathbf{h}_i \cap \mathbf{h}_j^c| \times |\mathbf{h}_j^c \cap \mathbf{h}_i| = \sum_{k=1}^N h_{ik}(1-h_{jk}) \sum_{k=1}^N (1-h_{ik})h_{jk}$ , where  $\mathbf{u} = (1, \dots, 1)^\top$  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  $a_1, a_2 \geq 0$ .

technology portfolios  $\mathbf{h} \in \mathcal{H}^N$  is determined by the conditional distribution

$$\mu^\vartheta(G|\mathbf{q}, \mathbf{h}) = \prod_{i < j} \frac{e^{\vartheta a_{ij}(\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}{1 + e^{\vartheta(\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}, \quad (80)$$

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

$$p^\vartheta(q_i, \mathbf{h}_i, q_j, \mathbf{h}_j) \equiv \frac{e^{\vartheta(\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}{1 + e^{\vartheta(\rho f(\mathbf{h}_i, \mathbf{h}_j) q_i q_j - \zeta)}}. \quad (81)$$

In the following we consider a particularly simple specification in which each firm  $i$  is assigned a technology  $k \in \{1, \dots, N\}$  uniformly at random so that  $h_{ik} = 1$  and  $h_{il} = 0$  for all  $l \neq k$ . Moreover, let  $f(\mathbf{h}_i, \mathbf{h}_j) = \mathbf{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle \geq 1\}}$ , that is, firms  $i$  and  $j$  can only benefit from a collaboration if they have a technology in common.

**Proposition 8.** *Assume that each firm  $i$  is assigned a technology  $k \in \{1, \dots, N\}$  uniformly at random and let  $f(\mathbf{h}_i, \mathbf{h}_j) = \mathbf{1}_{\{\langle \mathbf{h}_i, \mathbf{h}_j \rangle \geq 1\}}$ .*

(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*

$$k_{nn}(k) = \frac{k(1 - \frac{1}{N})(1 + n\frac{1}{N} - (n+1)\frac{1}{N})}{\frac{1}{N}(1 + n - k)}, \quad (82)$$

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 8.** We first prove part (i) of the proposition. If technologies are assigned uniformly at random then

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

Due to symmetry the firms quantities in the stationary state when  $\vartheta \rightarrow \infty$  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 a proof of part (ii) of the proposition. The average nearest neighbor degree distribution is then given by

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

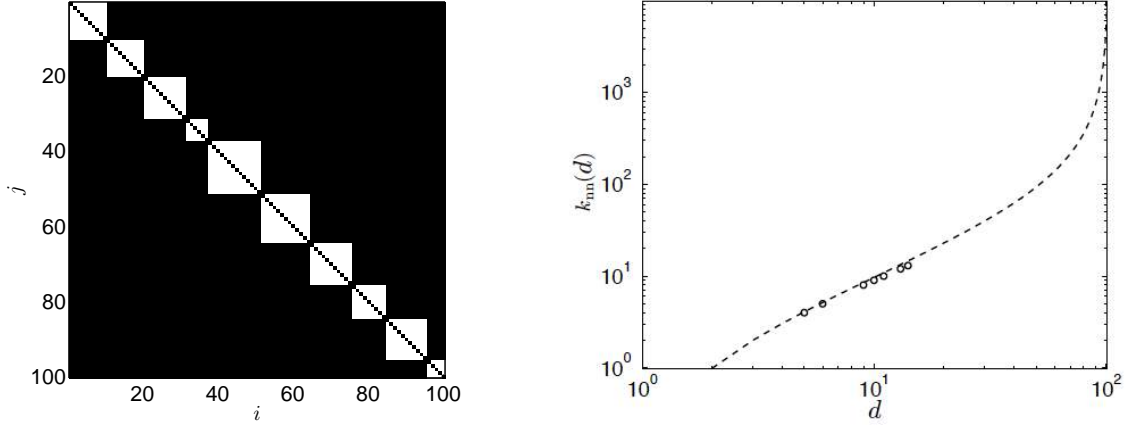


Figure 3: (Left panel) Illustration of the matrix with elements  $\mathbf{1}_{\{(h_i, h_j) > 0\}}$  for  $n = 100$  firms and  $N = 10$  technologies. (Right panel) The average nearest neighbor degree distribution,  $k_{\text{nn}}(d)$ , for the same parameters. The dashed line represents the solution from Equation (82) while the circles correspond to a numerical simulation.

where

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

We then get

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

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

Finally, we give a proof of part (iii) of the proposition. The clustering coefficient is simply given by  $C(k) = \mathbb{P}(a_{23} = 1 | a_{12} = 1, a_{23} = 1, d_1(G) = k) = 1$ . 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.  $\square$

An illustration of the average nearest neighbor degree  $k_{\text{nn}}(k)$  can be seen in Figure 3.

## 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 information), it gives us a sample of 2,033 firms and a total of 720 collaborations in the year 2006. The average degree of the firms in this sample is 0.71 with a standard deviation of 2.01, and the maximum degree is 25 attained by *Pfizer Inc.* Figure 1 shows the largest connected component of the R&D collaboration network, while Figure 3.1 shows the corresponding component with firms in the SIC-28 sector (see Appendix B for the definition of a connected component). The figure indicates two clusters appearing which are related to the different industries in which firms are operating.

Figure 2, and Tables 1 and 2 show the 10, 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 534 firms (26.27 % of the total), followed by the sector electronic and other electric equipment, with 306 firms (15.05 % of the total). At the 3-digit SIC level the largest sector is the drugs development sector, with 416 firms (20.46 % of the total), and the second largest sector is computer and data processing services with 193 firms (9.49 % of the total).

Figure 3 shows the degree distribution,  $P(d)$ , the average nearest neighbor connectivity,  $k_{nn}(d)$ , the clustering degree distribution,  $C(d)$ , 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 network 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 1) 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

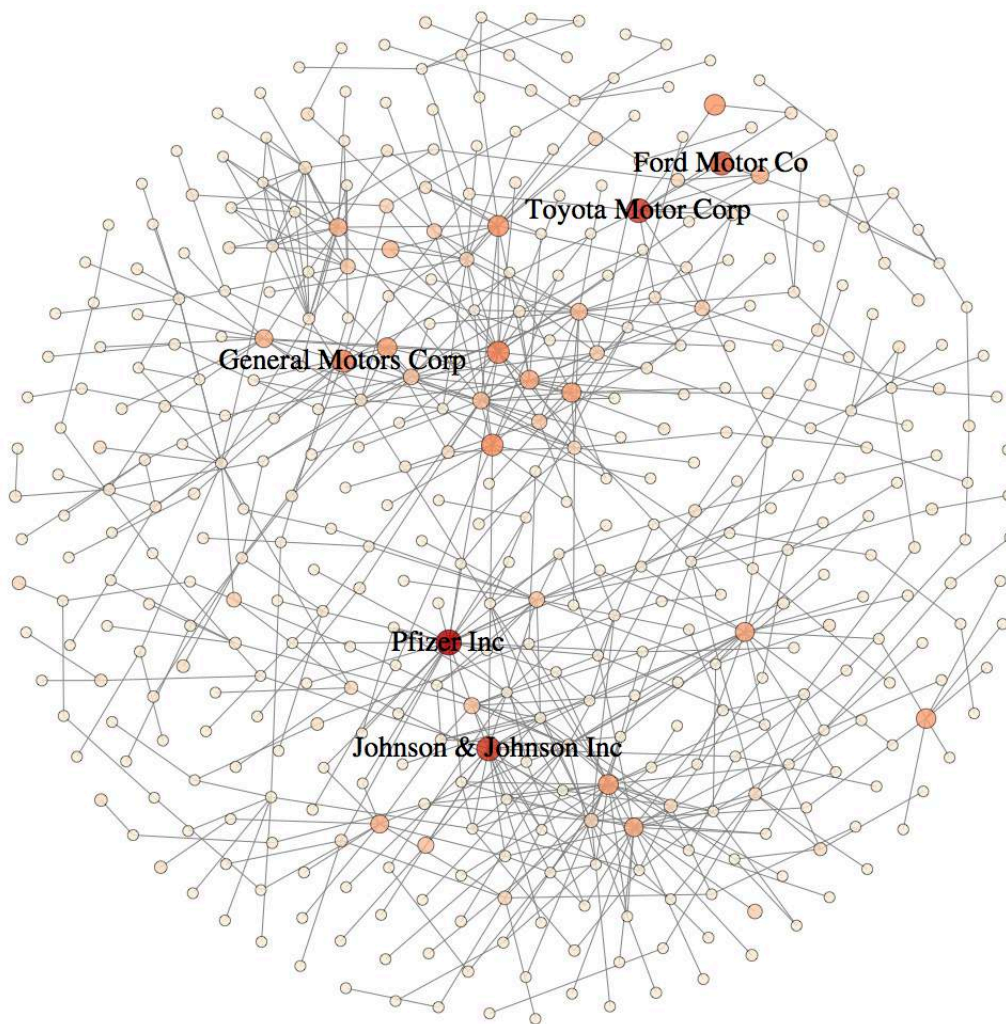


Figure 1: The largest connected component in the observed network of R&D collaborations in the year 2006 for the firms without missing observations on R&D expenditures and industry classifications. The shade and size of a node indicates its R&D expenditures. The 5 largest firms in terms of their R&D expenditures are mentioned in the graph. The number of firms is 2033 and the number of firms in the largest connected component is 431. The figure indicates a separation between the manufacturing and the chemicals and pharmaceuticals sectors.

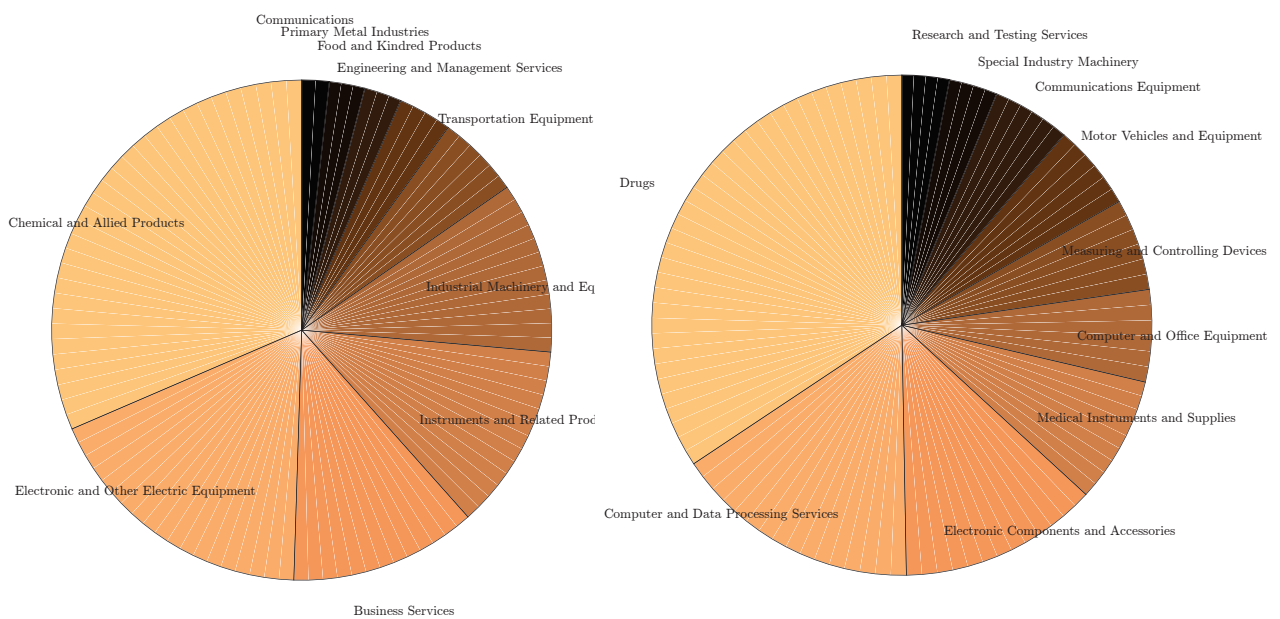


Figure 2: The shares of the ten largest sectors at the 2-digit (left panel) and 3-digit (right panel) SIC levels. See also Tables 1 and 2, respectively.

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

Sector	2-dig SIC	# firms	% of tot.	Rank
Chemical and Allied Products	28	534	26.27	1
Electronic and Other Electric Equipment	36	306	15.05	2
Business Services	73	205	10.08	3
Instruments and Related Products	38	204	10.03	4
Industrial Machinery and Equipment	35	188	9.25	5
Transportation Equipment	37	91	4.48	6
Engineering and Management Services	87	59	2.90	7
Food and Kindred Products	20	41	2.02	8
Primary Metal Industries	33	39	1.92	9
Communications	48	30	1.48	10
Electric Gas and Sanitary Services	49	24	1.18	11
Fabricated Metal Products	34	21	1.03	12
Health Services	80	20	0.98	13
Miscellaneous Manufacturing Industries	39	18	0.89	14
Paper and Allied Products	26	16	0.79	15
Rubber and Miscellaneous Plastics Products	30	16	0.74	16
Petroleum and Coal Products	29	14	0.69	17
Stone Clay and Glass Products	32	14	0.69	18
Wholesale Trade - Durable Goods	50	13	0.64	19
Textile Mill Products	22	12	0.59	20

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

Sector	3-dig SIC	# firms	% of tot.	Rank
Drugs	283	416	20.46	1
Computer and Data Processing Services	737	193	9.49	2
Electronic Components and Accessories	367	156	7.67	3
Medical Instruments and Supplies	384	99	4.87	4
Computer and Office Equipment	357	72	3.54	5
Measuring and Controlling Devices	382	72	3.54	6
Motor Vehicles and Equipment	371	67	3.30	7
Communications Equipment	366	61	3.00	8
Special Industry Machinery	355	38	1.87	9
Research and Testing Services	873	37	1.82	10
Misc. Electrical Equipment and Supplies	369	28	1.38	11
Chemicals & Allied Products	280	26	1.28	12
Plastics Materials and Synthetic	282	25	1.23	13
General Industrial Machinery	356	25	1.23	14
Electrical Industrial Apparatus	362	17	0.84	15
Aircraft and Parts	372	16	0.79	16
Misc. Chemical Products	289	15	0.74	17
Blast Furnace and Basic Steel Products	331	15	0.74	18
Agricultural Chemicals	287	14	0.69	19
Metalworking Machinery	354	14	0.69	20

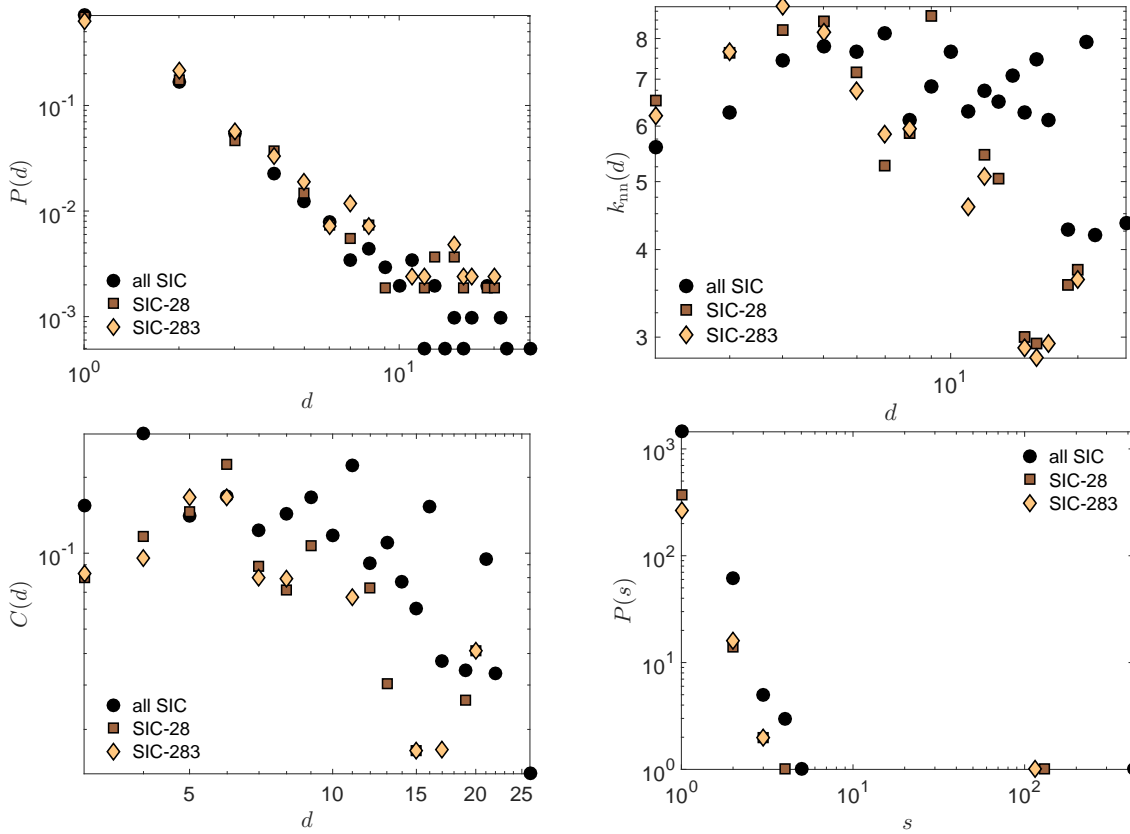


Figure 3: The degree distribution,  $P(d)$ , the average nearest neighbor connectivity,  $k_{nn}(d)$ , the clustering degree distribution,  $C(d)$ , and the component size distribution,  $P(s)$ .

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,  $k_{nn}(d)$ . While the average nearest neighbor connectivity  $k_{nn}(d)$  is decreasing with 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 2.4 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 datasources to obtain information about M&As. The first is the Thomson Reuters’ Securities Data Company (SDC) M&A database, which has historically been the most widely used database for empirical research in the field of M&As. Data in SDC dates back to 1965 with a slightly more complete coverage of deals starting in the early 1980s. The second database with information about M&As is Bureau van Dijk’s (BvD) Zephyr database, which is a recent alternative to the SDC M&As database. The history of deals recorded in Zephyr goes back to 1997. In 1997 and 1998 only European deals are recorded, while international deals are included starting from 1999. According to Huyghebaert and Luypaert [2010], Zephyr “covers deals of smaller value and has a better coverage of European transactions”. A comparison and more detailed discussion of the two databases can be found in Bollaert and Delanghe [2015] and Bena et al. [2008].

We merged the SDC and Zephyr databases (with the above mentioned name matching algorithm; see also 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 from the network.

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

The combined CATI-SDC alliance database provides the names for each firm in an alliance, but it does not contain information about the firms’ output levels or R&D expenses. We therefore matched the firms’ names in the combined CATI-SDC database with the firms’ names in Standard & Poor’s Compustat U.S. 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 statement 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 contains 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



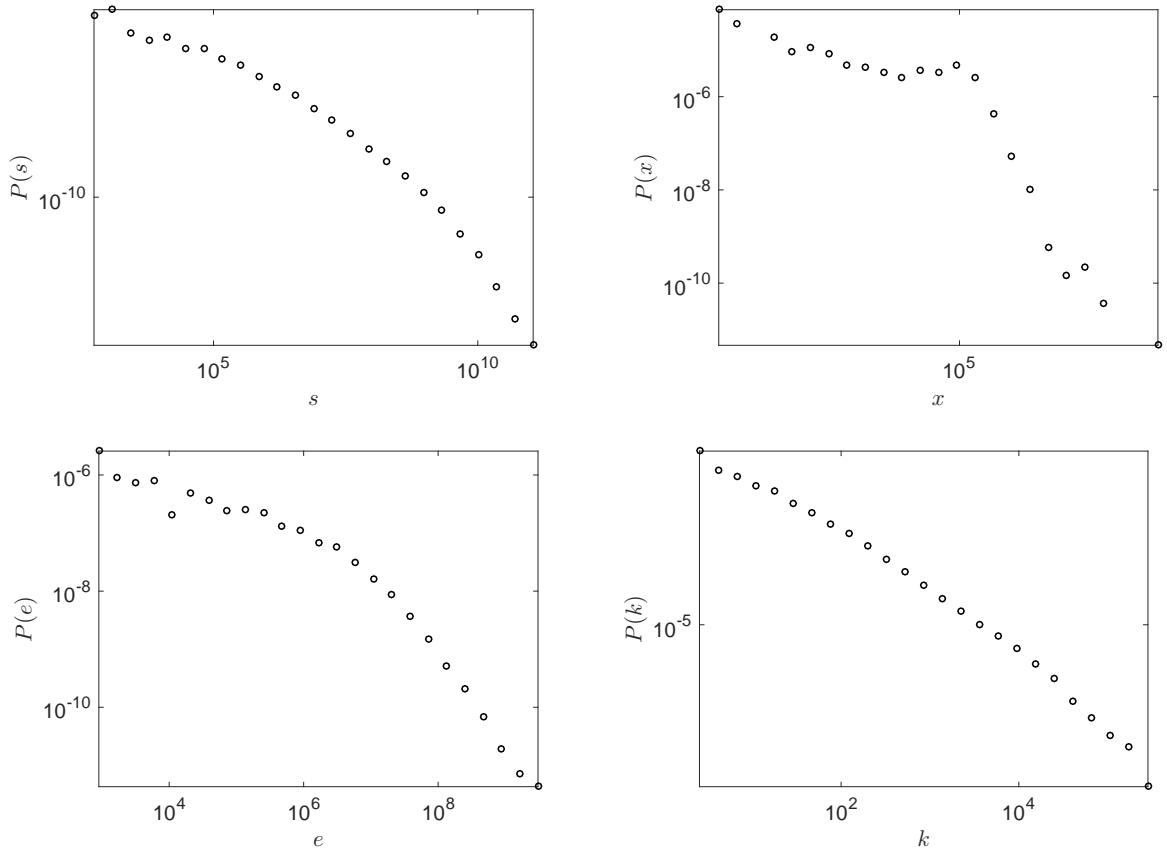


Figure 4: The sales distribution,  $P(s)$ , the productivity distribution,  $P(x)$ , the R&D expenditures distribution,  $P(e)$ , and the patent stock distribution,  $P(k)$  using a logarithmic binning of the data [McManus et al., 1987].

the local Chambers of Commerce to fulfil 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.

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

For the matching of firms across datasets we adopted the name matching algorithm developed as part of the NBER patent data project [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 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 2,033 firms, on which the empirical analysis in Section 3 is based.

The empirical distributions for sales,  $P(s)$ , productivity,  $P(x)$ , R&D expenditures,  $P(e)$ , and the patent stocks,  $P(k)$  (using a logarithmic binning of the data with 100 bins [McManus et al., 1987]) are shown in Figure 4. All distributions are highly skewed, indicating a large degree of inequality in firms' sizes, productivity and patent activities. Moreover, Figure 5 shows a correlation scatter plot for sales, productivity, R&D expenditures and the patent stocks. All are highly correlated, with a Spearman correlation coefficient between sales and R&D expenditures

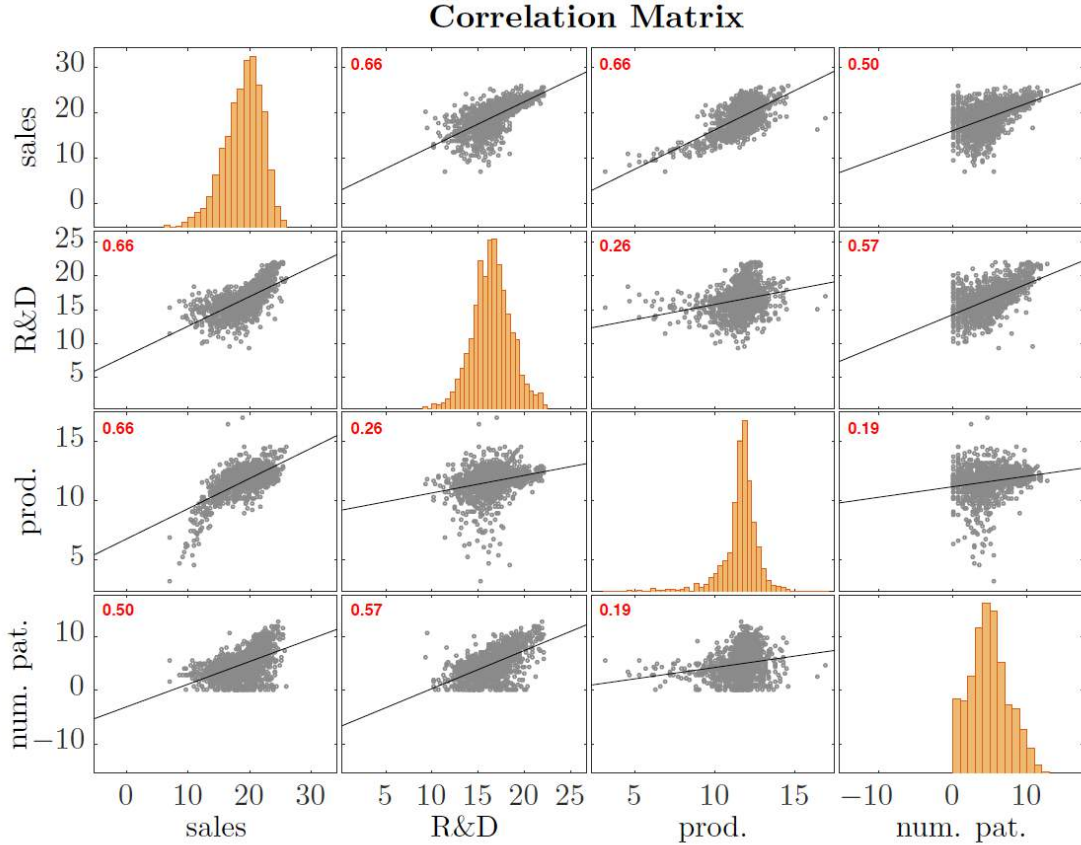


Figure 5: Correlation scatter plot for sales, productivity, R&D expenditures and the patent stocks.

of  $\rho = 0.69$  between sales and productivity of  $\rho = 0.54$ , and sales and the number of patents of  $\rho = 0.53$ . The correlation between R&D expenditure and productivity is  $\rho = 0.29$  and R&D expenditures and the number of patents is  $\rho = 0.56$ . Finally, the correlation between productivity and the number of patents is  $\rho = 0.22$ .

### F.3. Geographic Location and Distance

The number of firms in each country is shown in Figures 6 and 7, respectively, while Table 3 shows the 25 countries with the largest numbers of firms. The dominant role of the U.S. with 989 collaborations making up 48.65% of the total number of collaborations is clearly visible. The second largest country in terms of R&D collaborations is Japan with 408, which comprises 20.07% of all collaborations. The U.S. and Japan then together account for 68.72%, that is, more than two thirds of all R&D collaborations 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 the Google Maps Geocoding API.<sup>76</sup> 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 9, while Figure 8 shows the locations (at the city level) and collaborations of the firms in the

<sup>76</sup>See <https://developers.google.com/maps/documentation/geocoding/intro>.

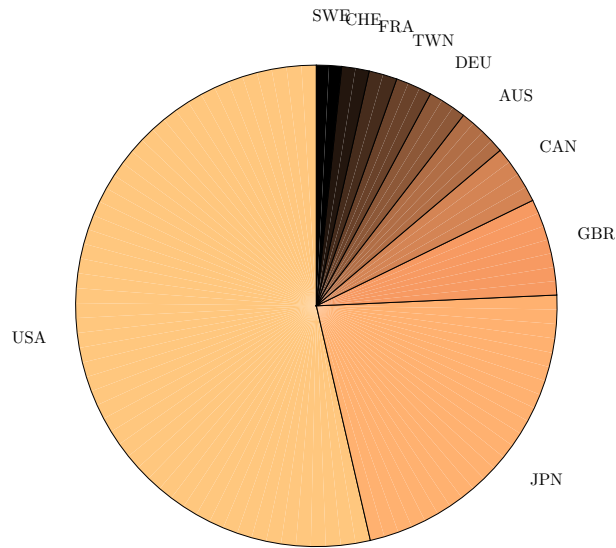


Figure 6: The number of firms in each country.

Table 3: The 25 countries with the largest numbers of firms.

Name	Code	# firms	% of tot.	Rank
United States	USA	989	48.65	1
Japan	JPN	408	20.07	2
United Kingdom	GBR	120	5.90	3
Canada	CAN	73	3.59	4
Australia	AUS	62	3.05	5
Germany	DEU	48	2.36	6
Taiwan	TWN	45	2.21	7
France	FRA	35	1.72	8
Switzerland	CHE	34	1.67	9
Sweden	SWE	31	1.52	10
India	IND	21	1.03	11
Finland	FIN	16	0.79	12
Netherlands	NLD	14	0.69	13
Iceland	ISL	13	0.64	14
Slovakia	SVK	13	0.64	15
Denmark	DNK	12	0.59	16
Belgium	BEL	11	0.54	17
Italy	ITA	11	0.54	18
Israel	ISR	10	0.49	19
Morocco	MAR	9	0.44	20
Norway	NOR	8	0.39	21
China	CHN	7	0.34	22
Singapore	SGP	7	0.34	23
Spain	ESP	4	0.20	24
Hong Kong	HKG	4	0.20	25

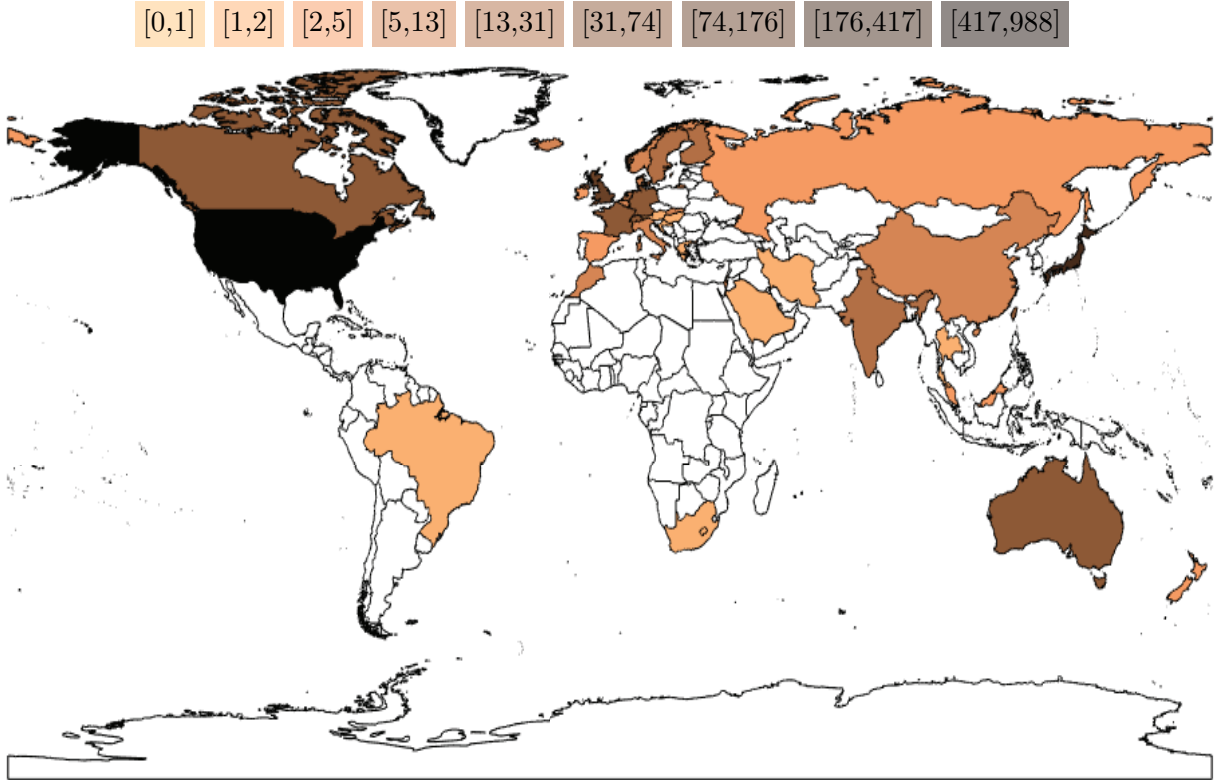


Figure 7: The number of firms in each country.

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 this spillovers, in line with previous empirical studies [Lychagin et al., 2016].

#### F.4. Patents and Technological Similarity

We identified the patent portfolios of the firms in our dataset using the EPO Worldwide Patent Statistical Database (PATSTAT) [Jaffe and Trajtenberg, 2002; Thoma et al., 2010]. It includes bibliographic details on patents filed to 80 patent offices worldwide, covering more than 60 million documents. Hence filings in all major countries and the the World International Patent Office are covered. We matched the firms in our data with the assignees in the PATSTAT database using the above mentioned name matching algorithm. 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]. We obtained matches for roughly 30% of the firms in the data. The technology classes were identified using the main international patent classification (IPC) numbers at the 4 digit level.

The technological proximity  $f_{ij}^k = f_{ji}^k$ ,  $k \in \{J, M\}$  (see Equation (23)), 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

$$f_{ij}^J = \frac{\mathbf{P}_i^\top \mathbf{P}_j}{\sqrt{\mathbf{P}_i^\top \mathbf{P}_i} \sqrt{\mathbf{P}_j^\top \mathbf{P}_j}}. \quad (83)$$

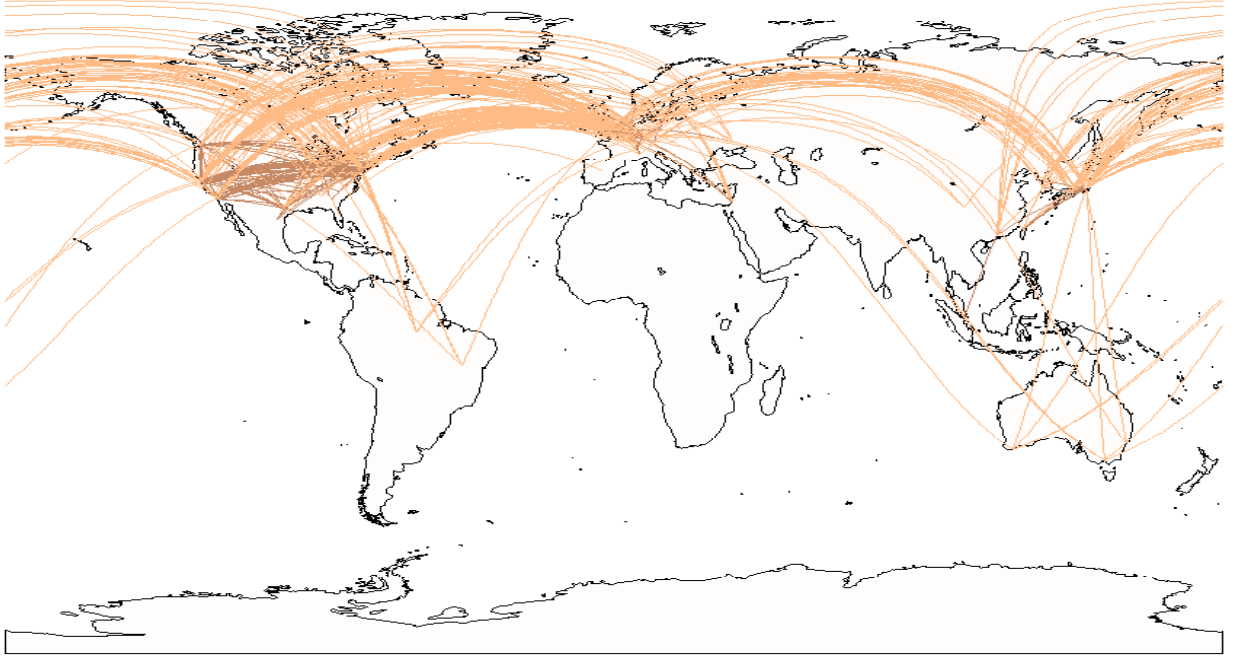


Figure 8: The locations (at the city level) and collaborations of the firms in the combined CATI-SDC database.

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

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  $1 \leq i \leq n$  and  $1 \leq j \leq N$ . Further, we construct the  $(N \times n)$  normalized patent shares matrix  $\tilde{\mathbf{T}}$  with elements  $\tilde{T}_{ji} = T_{ji} / \sqrt{\sum_{k=1}^n T_{ki}^2}$ , 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  $\mathbf{F}^M = (f_{ij}^M)_{1 \leq i, j \leq n}$  is defined as

$$\mathbf{F}^M = \tilde{\mathbf{T}}^\top \mathbf{\Omega} \tilde{\mathbf{T}}. \quad (84)$$

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 (23). Both measures are highly correlated. The Spearman correlation coefficient between the Jaffe and the Mahalanobis proximity metrics is 0.91882, and a correlation plot can be seen in the right panel of Figure 9.

## G. Estimation Algorithms

In the following appendices we provide additional details regarding our estimation algorithms. First, in Appendix G.1 we give a detailed explanation for how the DMH algorithm is implemented. Second, in Appendix G.2 we discuss the implementation of the AEX algorithm. Further, in Appendix G.3 we outline the convergence proof of the AEX algorithm.

### G.1. Implementation of the DMH Algorithm

In our empirical model, the unknown parameters are denoted by  $\boldsymbol{\theta} = (\rho, b, \boldsymbol{\delta}^\top, \boldsymbol{\gamma}^\top, \varkappa)$ , where  $\rho$  captures the technology spillover effect;  $b$  captures the competition effect;  $\boldsymbol{\delta}$  captures effects of firm's exogenous characteristics in terms of productivity and sector fixed effect;  $\boldsymbol{\gamma}$  captures effects

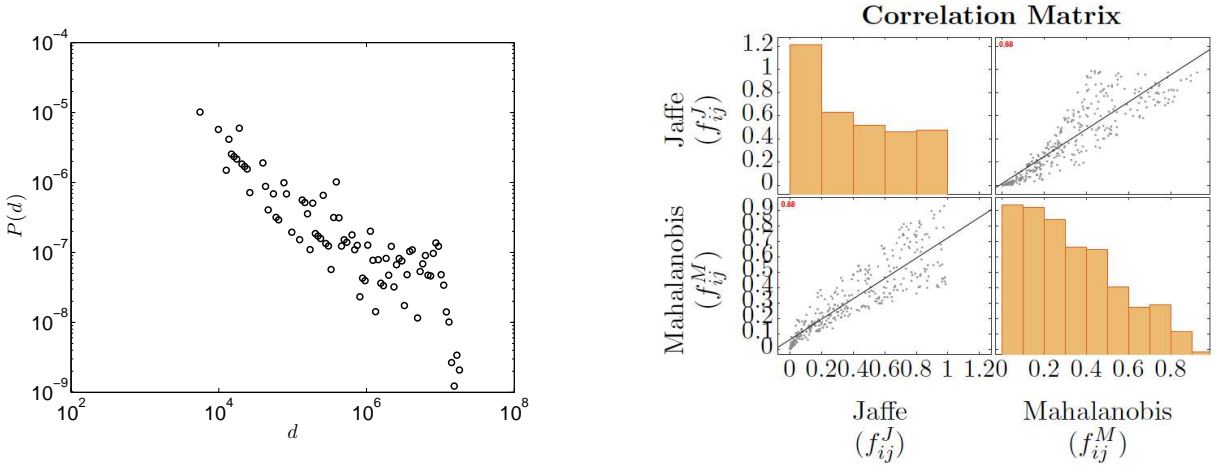


Figure 9: (Left panel) The distance distribution,  $P(d)$ , across collaborating firms in the combined CATI-SDC database. (Right panel) Correlation plot for the Jaffe ( $f_{ij}^J$ ) and the Mahalanobis ( $f_{ij}^M$ ) technology proximity metrics across pairs of firms  $1 \leq i, j \leq n$ .

of dyad-specific exogenous factors in the link cost function; and  $\varkappa$  captures the cyclic triangles effect. For the Bayesian posterior analysis, the joint posterior (probability) density function of  $\theta$  can be constructed by

$$P(\theta|\mathbf{q}, G) \propto \pi(\theta) \cdot \mu(\mathbf{q}|G, \theta), \quad (85)$$

where  $\pi(\cdot)$  represents the prior density function and we assume independence between prior distributions. We specify prior distributions  $(\rho, b) \sim \mathcal{U}_2(\mathcal{O})$ ,  $\delta \sim \mathcal{N}_i(\delta_0, \Delta_0)$ ,  $\gamma \sim \mathcal{N}_r(\gamma_0, \Gamma_0)$ , and  $\varkappa \sim \mathcal{N}(\varkappa_0, \mathbf{K}_0)$ . The prior of  $(\rho, b)$  is a multivariate uniform (uninformative) distribution with the values restricted to the compact subset  $\mathcal{O}$  of  $\mathbb{R}^2$ , in which the matrix  $\mathbf{M}(G) = \mathbf{I}_n + b\mathbf{B} - \rho(\mathbf{A} \circ \mathbf{F})$  is positive definite.<sup>77</sup> The priors for  $\delta$ ,  $\gamma$ , and  $\varkappa$  are normal distributions, which are conjugate priors commonly used in the Bayesian literature [Koop et al., 2007; Robert and Casella, 2004]. In our applications, we fix  $\delta_0 = 0$ ,  $\Delta_0 = 100\mathbf{I}_i$ ,  $\gamma_0 = 0$ ,  $\Gamma_0 = 100\mathbf{I}_r$ ,  $\varkappa_0 = 0$ , and  $\mathbf{K}_0 = 100$  to ensure prior density is relatively flat over the range of the data.

Directly drawing samples from the posterior distribution,  $P(\theta|\mathbf{q}, G)$ , would be difficult due to high dimensionality. Therefore, we apply the Metropolis-within-Gibbs sampling procedure: we first divide unknown parameters into blocks so that drawing from their conditional posterior distributions is feasible. When we make draws sequentially from these conditional posterior distributions by cycling, in the limit these draws can be treated as draws from the joint posterior distribution. For the conditional posterior distribution of  $(\rho, b)$ , which is not available in a closed form, we use the MH algorithm to draw from this conditional distribution. It has been shown in Tierney [1994] and Chib and Greenberg [1996] that the combination of Markov chains is still a Markov chain with the invariant distribution equal to the correct objective distribution.

We apply the Gibbs sampler to update the following set of conditional posterior distributions:

- (i)  $P(\rho, b|\mathbf{q}, G, \theta \setminus (\rho, b)) \propto \mu(\mathbf{q}, G|\theta)I((\rho, b) \in \mathcal{O})$ , where  $\theta \setminus (\rho, b)$  stands for  $\theta$  with  $(\rho, b)$  ex-

<sup>77</sup>This requirement implies that  $\mathbf{M}(G)$  is invertible and  $\mathbf{M}(G)^{-1}$  is also positive definite.  $\mathbf{M}(G)$  is positive definite if and only if all of its eigenvalues are positive.

cluded.

(ii)  $P(\boldsymbol{\delta}|\mathbf{q}, G, \boldsymbol{\theta} \setminus \boldsymbol{\delta}) \propto \phi_l(\boldsymbol{\delta}; \boldsymbol{\delta}_0, \Delta_0) \cdot \mu(\mathbf{q}|G, \boldsymbol{\theta})$ , where  $\phi$  denotes the normal density function.

(iii)  $P(\boldsymbol{\gamma}|\mathbf{q}, G, \boldsymbol{\theta} \setminus \boldsymbol{\gamma}) \propto \phi_r(\boldsymbol{\gamma}; \boldsymbol{\gamma}_0, \Gamma_0) \cdot \mu(\mathbf{q}, G|\boldsymbol{\theta})$ .

(iv)  $P(\boldsymbol{\varkappa}|\mathbf{q}, G, \boldsymbol{\theta} \setminus \boldsymbol{\varkappa}) \propto \phi(\boldsymbol{\varkappa}; \boldsymbol{\varkappa}_0, \mathbf{K}_0) \cdot \mu(\mathbf{q}, G|\boldsymbol{\theta})$ .

We apply the DMH algorithm to simulate draws from (i) to (iv) because their distributions are not available in a closed form. At each  $t^{\text{th}}$  MCMC iteration, the implementation steps are illustrated as follows:

Step I. Simulate  $(\rho^{(t)}, b^{(t)})$  from  $P(\rho, b | \mathbf{q}, G, \Upsilon^{(t-1)})$  by the DMH algorithm, where  $\Upsilon^{(t-1)}$  denotes the rest of parameters evaluated at the  $(t-1)^{\text{th}}$  iteration.

(a) propose  $(\tilde{\rho}, \tilde{b})$  from a random walk proposal density  $T_1(\rho, b|b^{(t-1)}, \rho^{(t-1)})$ .

(b) simulate an auxiliary data  $\mathbf{q}'$  and  $G'$  by  $M$  runs of the MH algorithm based on

$$\mu(\mathbf{q}, G|\tilde{\rho}, \tilde{b}, \Upsilon^{(t-1)}) = \frac{\exp\left(\Phi\left(\mathbf{q}, G, \tilde{\rho}, \tilde{b}, \Upsilon^{(t-1)}\right)\right)}{\sum_{\mathbf{g} \in \mathcal{G}^n} \int_{\mathcal{Q}^n} \exp\left(\Phi\left(\mathbf{q}, \mathbf{g}, \tilde{\rho}, \tilde{b}, \Upsilon^{(t-1)}\right)\right) d\mathbf{q}}$$

starting from the observed network  $G$ .<sup>78</sup> The details are as follows:

First, let the initial auxiliary network  $G^{(0)}$  equal to the observed network  $G$ , i.e., set the auxiliary adjacency matrix  $\mathbf{A}^{(0)} = \mathbf{A}$ . At the  $m^{\text{th}}$  run,  $m = 1, \dots, M$ , we propose  $\tilde{G}$  from  $G^{(m-1)}$  by selecting node  $i$  with probability  $\frac{1}{n}$  and node  $j \neq i$  with a probability  $\frac{1}{n-1}$  randomly and change the value  $a_{ij}^{(m-1)}$  and  $a_{ji}^{(m-1)}$  in matrix  $\mathbf{A}^{(m-1)}$  to  $1 - a_{ij}^{(m-1)}$  and  $1 - a_{ji}^{(m-1)}$  and propose it as  $\tilde{\mathbf{A}}$ . With a certain probability, we flip all elements in  $\mathbf{A}^{(m-1)}$  from 0 to 1 (or 1 to 0) and propose it as  $\tilde{\mathbf{A}}$ . Denote  $\tilde{\mathbf{M}} = \mathbf{I}_n + \tilde{b}\mathbf{B} - \tilde{\rho}(\tilde{\mathbf{A}} \circ \mathbf{F})$  and calculate  $\tilde{\mathbf{q}}^* = \tilde{\mathbf{M}}^{-1}(\mathbf{X}\delta^{(t-1)})$ . Simulate  $\tilde{\mathbf{q}}$  from  $\mathcal{N}(\tilde{\mathbf{q}}^*, \tilde{\mathbf{M}}^{-1})$ . Then, with the acceptance probability

$$\tilde{\alpha}(\tilde{G}|G^{(m-1)}) = \min \left\{ \frac{\exp\left(\Phi(\tilde{G}, \tilde{\mathbf{q}})\right)}{\exp\left(\Phi(G^{(m-1)}, \mathbf{q}^{(m-1)})\right)}, 1 \right\},$$

set  $G^{(m)}$  to  $\tilde{G}$  and  $\mathbf{q}^{(m)} = \tilde{\mathbf{q}}$ , otherwise, set  $G^{(m)} = G^{(m-1)}$  and  $\mathbf{q}^{(m)} = \mathbf{q}^{(m-1)}$ . After the  $M$  runs, collect  $G^{(M)}$  and  $\mathbf{q}' = \mathbf{q}^{(M)}$ .

(c) With the acceptance probability equals to

$$\begin{aligned} & \alpha(\tilde{\rho}, \tilde{b}|\rho^{(t-1)}, b^{(t-1)}, G', \mathbf{q}') \\ &= \min \left\{ \frac{\mu\left(\mathbf{q}, G, \Upsilon^{(t-1)}, \tilde{\rho}, \tilde{b}\right)}{\mu\left(\mathbf{q}, G, \Upsilon^{(t-1)}, \rho^{(t-1)}, b^{(t-1)}\right)} \cdot \frac{\mu\left(\mathbf{q}', G'^{(t-1)}, \rho^{(t-1)}, b^{(t-1)}\right)}{\mu\left(\mathbf{q}', G'^{(t-1)}, \tilde{\rho}, \tilde{b}\right)} \cdot \frac{I((\tilde{\rho}, \tilde{b}) \in O)}{I((\rho^{(t-1)}, b^{(t-1)}) \in O)}, 1 \right\} \\ &= \min \left\{ \frac{\exp\left(\Phi(\mathbf{q}, G, \Upsilon^{(t-1)}, \tilde{\rho}, \tilde{b})\right)}{\exp\left(\Phi(\mathbf{q}, G, \Upsilon^{(t-1)}, \rho^{(t-1)}, b^{(t-1)})\right)} \cdot \frac{\exp\left(\Phi(\mathbf{q}', G'^{(t-1)}, \rho^{(t-1)}, b^{(t-1)})\right)}{\exp\left(\Phi(\mathbf{q}', G'^{(t-1)}, \tilde{\rho}, \tilde{b})\right)} \cdot \frac{I((\tilde{\rho}, \tilde{b}) \in O)}{I((\rho^{(t-1)}, b^{(t-1)}) \in O)}, 1 \right\}, \end{aligned}$$

<sup>78</sup>This step mimics the MH sampler in Snijders [2002] and Mele [2017]. In practice, we set  $M = n^2$  where  $n$  is the size of the network.

set  $(\rho^{(t)}, b^{(t)}) = (\tilde{\rho}, \tilde{b})$ . Otherwise, set  $(\rho^{(t)}, b^{(t)}) = (\rho^{(t-1)}, b^{(t-1)})$ .

Step II. Simulate  $\delta^{(t)}$  from

$$P(\delta|\mathbf{q}, G, \rho^{(t)}, b^{(t)}, \Upsilon^{(t-1)}) \propto \phi_l(\delta; \delta_0, \Delta_0) \cdot \mu(\mathbf{q}|G, \rho^{(t)}, b^{(t)}, \delta, \Upsilon^{(t-1)}).$$

(a) propose  $\tilde{\delta}$  from a random walk proposal density  $T_2(\delta|\delta^{(t-1)})$ .

(b) simulate an auxiliary data  $\mathbf{q}'$  and  $G'$  by  $M$  runs of the MH algorithm based on

$$\mu(\mathbf{q}, G|\tilde{\delta}, \rho^{(t)}, b^{(t)}, \Upsilon^{(t-1)}) = \frac{\exp\left(\Phi\left(\mathbf{q}, G, \tilde{\delta}, \rho^{(t)}, b^{(t)}, \Upsilon^{(t-1)}\right)\right)}{\sum_{\mathbf{g} \in \mathcal{G}^n} \int_{\mathcal{Q}^n} \exp\left(\Phi\left(\mathbf{q}, \mathbf{g}, \tilde{\delta}, \rho^{(t)}, b^{(t)}, \Upsilon^{(t-1)}\right)\right) d\mathbf{q}}$$

starting from the observed network  $G$ .

(c) With the acceptance probability equals to

$$\begin{aligned} & \alpha(\tilde{\delta}|\delta^{(t-1)}, G', \mathbf{q}') \\ &= \min \left\{ \frac{\mu\left(\mathbf{q}, G, \tilde{\delta}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)}{\mu\left(\mathbf{q}, G, \delta^{(t-1)}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)} \cdot \frac{\mu\left(\mathbf{q}', G', \delta^{(t-1)}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)}{\mu\left(\mathbf{q}', G', \tilde{\delta}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)} \cdot \frac{\phi_l(\tilde{\delta}; \delta_0, \Delta_0)}{\phi_l(\delta^{(t-1)}; \delta_0, \Delta_0)}, 1 \right\} \\ &= \min \left\{ \frac{\exp\left(\Phi\left(\mathbf{q}, G, \tilde{\delta}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)\right)}{\exp\left(\Phi\left(\mathbf{q}, G, \delta^{(t-1)}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)\right)} \cdot \frac{\exp\left(\Phi\left(\mathbf{q}', G', \delta^{(t-1)}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)\right)}{\exp\left(\Phi\left(\mathbf{q}', G', \tilde{\delta}, \Upsilon^{(t-1)}, \rho^{(t)}, b^{(t)}\right)\right)} \cdot \frac{\phi_l(\tilde{\delta}; \delta_0, \Delta_0)}{\phi_l(\delta^{(t-1)}; \delta_0, \Delta_0)}, 1 \right\}, \end{aligned}$$

set  $\delta^{(t)} = \tilde{\delta}$ . Otherwise, set  $\delta^{(t)} = \delta^{(t-1)}$ .

Step III. Simulate  $\gamma^{(t)}$  from  $P(\gamma|\mathbf{q}, G, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})$  by the DMH algorithm.

(a) propose  $\tilde{\gamma}$  from a random walk proposal density  $T_3(\gamma|\gamma^{(t-1)})$

(b) simulate an auxiliary data  $\mathbf{q}'$  and  $G'$  by  $M$  runs of the MH algorithm based on

$$\mu(\mathbf{q}, G|\tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}) = \frac{\exp\left(\Phi\left(\mathbf{q}, G, \tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)\right)}{\sum_{\mathbf{g} \in \mathcal{G}^n} \int_{\mathcal{Q}^n} \exp\left(\Phi\left(\mathbf{q}, \mathbf{g}, \tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)\right) d\mathbf{q}}$$

starting from the observed network  $G$ .

(d) With the acceptance probability equal to

$$\begin{aligned} & \alpha(\tilde{\gamma}|\gamma^{(t-1)}, G', \mathbf{q}') \\ &= \min \left\{ \frac{\mu\left(G, \mathbf{q}, \tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)}{\mu\left(G, \mathbf{q}, \gamma^{(t-1)}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)} \cdot \frac{\mu\left(G', \mathbf{q}', \gamma^{(t-1)}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)}{\mu\left(G', \mathbf{q}', \tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)} \cdot \frac{\phi_r(\tilde{\gamma}|\gamma_0, \Gamma_0)}{\phi_r(\gamma^{(t-1)}|\gamma_0, \Gamma_0)}, 1 \right\} \\ &= \min \left\{ \frac{\exp\left(\Phi\left(G, \mathbf{q}, \tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)\right)}{\exp\left(\Phi\left(G, \mathbf{q}, \gamma^{(t-1)}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)\right)} \cdot \frac{\exp\left(\Phi\left(G', \mathbf{q}', \gamma^{(t-1)}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)\right)}{\exp\left(\Phi\left(G', \mathbf{q}', \tilde{\gamma}, \Upsilon^{(t-1)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}\right)\right)} \cdot \frac{\phi_r(\tilde{\gamma}|\gamma_0, \Gamma_0)}{\phi_r(\gamma^{(t-1)}|\gamma_0, \Gamma_0)}, 1 \right\}, \end{aligned}$$

set  $\gamma^{(t)} = \tilde{\gamma}$ . Otherwise, set  $\gamma^{(t)} = \gamma^{(t-1)}$ .



Step IV. Simulate  $\varkappa^{(t)}$  from  $P(\varkappa|\mathbf{q}, G, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})$  by the DMH algorithm.

(a) propose  $\tilde{\varkappa}$  from a random walk proposal density  $T_4(\varkappa|\varkappa^{(t-1)})$

(b) simulate an auxiliary data  $\mathbf{q}'$  and  $G'$  by  $M$  runs of the MH algorithm based on

$$\mu(\mathbf{q}, G|\tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}) = \frac{\exp(\Phi(\mathbf{q}, G, \tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}))}{\sum_{\mathbf{g} \in \mathcal{G}^n} \int_{\mathcal{Q}^n} \exp(\Phi(\mathbf{q}, \mathbf{g}, \tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})) d\mathbf{q}}$$

starting from the observed network  $G$ .

(d) With the acceptance probability equal to

$$\begin{aligned} & \alpha(\tilde{\varkappa}|\varkappa^{(t-1)}, G', \mathbf{q}') \\ &= \min \left\{ \frac{\mu(G, \mathbf{q}, \tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})}{\mu(G, \mathbf{q}, \varkappa^{(t-1)}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})} \cdot \frac{\mu(G', \mathbf{q}', \varkappa^{(t-1)}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})}{\mu(G', \mathbf{q}', \tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)})} \cdot \frac{\phi_r(\tilde{\varkappa}|\varkappa_0, \mathbf{K}_0)}{\phi_r(\varkappa^{(t-1)}|\varkappa_0, \mathbf{K}_0)}, 1 \right\} \\ &= \min \left\{ \frac{\exp(\Phi(G, \mathbf{q}, \tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}))}{\exp(\Phi(G, \mathbf{q}, \varkappa^{(t-1)}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}))} \cdot \frac{\exp(\Phi(G', \mathbf{q}', \varkappa^{(t-1)}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}))}{\exp(\Phi(G', \mathbf{q}', \tilde{\varkappa}, \gamma^{(t)}, \delta^{(t)}, \rho^{(t)}, b^{(t)}))} \cdot \frac{\phi_r(\tilde{\varkappa}|\varkappa_0, \mathbf{K}_0)}{\phi_r(\varkappa^{(t-1)}|\varkappa_0, \mathbf{K}_0)}, 1 \right\}, \end{aligned}$$

set  $\gamma^{(t)} = \tilde{\gamma}$ . Otherwise, set  $\gamma^{(t)} = \gamma^{(t-1)}$ .

## G.2. Implementation of the AEX Algorithm

The AEX algorithm consists of two Markov chains running in parallel. Let the subscript  $t$  denote the  $t^{\text{th}}$  iteration of two chains. In the first chain, we simulate auxiliary network and output sample  $(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)})$  from  $\mu(\tilde{G}, \tilde{\mathbf{q}}|\tilde{\boldsymbol{\theta}}^{(t)})$ , where  $\tilde{\boldsymbol{\theta}}^{(t)}$  is sampled from the set of pre-specified parameter points  $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m)$ . In our applications, we set  $m = 50$  and  $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m)$  are chosen from the DMH draws using the Max-Min procedure suggested in Liang et al. [2015]. The index function  $H(\tilde{\boldsymbol{\theta}}^{(t)})$  denotes the order index of  $\tilde{\boldsymbol{\theta}}^{(t)}$ , i.e.,  $H(\tilde{\boldsymbol{\theta}}^{(t)}) = i$  if  $\tilde{\boldsymbol{\theta}}^{(t)} = \boldsymbol{\theta}_i$ . Let  $\mathbf{p} = (p_1, \dots, p_m)$  be the desired sampling frequencies from the respective distributions  $\mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_1), \dots, \mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_m)$ , where  $0 < p_i < 1$  and  $\sum_{i=1}^m p_i = 1$ . We follow Liang et al. [2015] to set  $p_1 = \dots = p_m = \frac{1}{m}$ . As required by the SAMC algorithm, we specify a *gain factor* sequence,  $\{a_t\}$ , which is a positive, nonincreasing sequence satisfying the following condition:

$$(A_1) \lim_{t \rightarrow \infty} a_t = 0, \quad \sum_{t=1}^{\infty} a_t = \infty, \quad \sum_{t=1}^{\infty} a_t^k < \infty \text{ for some } k \in (1, 2].$$

Following Liang et al. [2015], we set  $a_t = \frac{t_0}{\max(t_0, t)}$  with a known constant  $t_0 > 1$ . Since a larger value of  $t_0$  will help the auxiliary chain to reach each distribution  $\mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_i)$  more quickly, we set  $t_0 = 20,000$  in this paper. Also, we let  $w_i^{(t)}$  denote an *abundance factor* attached to each distribution  $\mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_i)$  at iteration  $t$  and  $w^{(t)} = (w_1^{(t)}, \dots, w_m^{(t)})$ . We set the initial values  $w_1^{(0)} = \dots = w_m^{(0)} = 1$  in the simulation. Finally, let  $\Omega^{(t)} = (\tilde{G}^{(1)}, \dots, \tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(1)}, \dots, \tilde{\mathbf{q}}^{(t)}, \tilde{\boldsymbol{\theta}}^{(1)}, \dots, \tilde{\boldsymbol{\theta}}^{(t)}, H(\tilde{\boldsymbol{\theta}}^{(1)}), \dots, H(\tilde{\boldsymbol{\theta}}^{(t)}), w^{(1)}, \dots, w^{(t)})$  denote the information in the auxiliary chain that we collect up to iteration  $t$ . In the second chain, we draw  $\boldsymbol{\theta}^{(t)}$  by the exchange algorithm from the target posterior distribution  $\mu(\boldsymbol{\theta}|\mathbf{q}, G)$ . Explicitly, the AEX algorithm is implemented by Part I and Part II at each  $t^{\text{th}}$  iteration described as follows:

### Part I. Auxiliary network simulation

1. Choose to update  $\tilde{\boldsymbol{\theta}}^{(t)}$  or  $\tilde{G}^{(t)}$  with a pre-specified probability. In our application, we choose 75% for updating  $\tilde{\boldsymbol{\theta}}^{(t)}$  and 25% for updating  $\tilde{G}^{(t)}$ .

- (a) Update  $\tilde{\boldsymbol{\theta}}^{(t)}$ : Select  $\tilde{\boldsymbol{\theta}}'$  from the set  $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m)$  according to a proposal distribution  $T_1(\cdot|\tilde{\boldsymbol{\theta}}^{(t-1)})$ . With the probability

$$\tilde{\alpha}(\tilde{\boldsymbol{\theta}}'|\tilde{\boldsymbol{\theta}}^{(t-1)}) = \min \left\{ \frac{w_{H(\tilde{\boldsymbol{\theta}}^{(t-1)})}^{(t-1)} \exp(\Phi(\tilde{G}^{(t-1)}, \tilde{\mathbf{q}}^{(t-1)}, \tilde{\boldsymbol{\theta}}')) T_1(\tilde{\boldsymbol{\theta}}^{(t-1)}|\tilde{\boldsymbol{\theta}}')}{w_{H(\tilde{\boldsymbol{\theta}}')}^{(t-1)} \exp(\Phi(\tilde{G}^{(t-1)}, \tilde{\mathbf{q}}^{(t-1)}, \tilde{\boldsymbol{\theta}}^{(t-1)})) T_1(\tilde{\boldsymbol{\theta}}|\tilde{\boldsymbol{\theta}}^{(t-1)})}, 1 \right\}, \quad (86)$$

update  $(\tilde{\boldsymbol{\theta}}^{(t)}, \tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}) = (\tilde{\boldsymbol{\theta}}', \tilde{G}^{(t-1)}, \tilde{\mathbf{q}}^{(t-1)})$ . Otherwise, set  $(\tilde{\boldsymbol{\theta}}^{(t)}, \tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}) = (\tilde{\boldsymbol{\theta}}^{(t-1)}, \tilde{G}^{(t-1)}, \tilde{\mathbf{q}}^{(t-1)})$ .

- (b) Update  $(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)})$ : simulate  $\tilde{G}^{(t)}$  and  $\tilde{\mathbf{q}}^{(t)}$  from  $\mu(\tilde{G}, \tilde{\mathbf{q}}|\tilde{\boldsymbol{\theta}}^{(t-1)})$  via few MH updates starting from  $(\tilde{G}^{(t-1)}, \tilde{\mathbf{q}}^{(t-1)})$ . Then set  $\tilde{\boldsymbol{\theta}}^{(t)} = \tilde{\boldsymbol{\theta}}^{(t-1)}$ .

2. For  $i = 1, \dots, m$ , update the abundance factor  $w_i^{(t)}$  by

$$\log(w_i^{(t)}) = \log(w_i^{(t-1)}) + a_t(e_{i,t} - p_i),$$

where  $e_{i,t} = 1$  if  $\tilde{\boldsymbol{\theta}}^{(t)} = \boldsymbol{\theta}_i$  and 0 otherwise.

3. Append  $(\tilde{\boldsymbol{\theta}}^{(t)}, \tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, H(\tilde{\boldsymbol{\theta}}^{(t)}), w^{(t)})$  to the collection  $\Omega^{(t-1)}$  to form  $\Omega^{(t)}$ .

## Part II. Adaptive Exchange algorithm for target parameter

4. Propose a candidate  $\boldsymbol{\theta}'$  from a random walk proposal distribution  $T_2(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t-1)})$ .
5. Re-sample an auxiliary network and output data  $(G', \mathbf{q}')$  from the collection  $\Omega^{(t)}$  via an importance sampling procedure. With the probability

$$P((G', \mathbf{q}') = (\tilde{G}^{(i)}, \tilde{\mathbf{q}}^{(i)})) = \frac{\sum_{j=1}^t \left\{ w_{H(\tilde{\boldsymbol{\theta}}^{(j)})}^{(j)} \frac{\exp(\Phi(\tilde{G}^{(j)}, \tilde{\mathbf{q}}^{(j)}, \boldsymbol{\theta}'))}{\exp(\Phi(\tilde{G}^{(j)}, \tilde{\mathbf{q}}^{(j)}, \tilde{\boldsymbol{\theta}}^{(j)}))} I(\tilde{G}^{(j)} = \tilde{G}^{(i)}) \right\}}{\sum_{j=1}^t \left\{ w_{H(\tilde{\boldsymbol{\theta}}^{(j)})}^{(j)} \frac{\exp(\Phi(\tilde{G}^{(j)}, \tilde{\mathbf{q}}^{(j)}, \boldsymbol{\theta}'))}{\exp(\Phi(\tilde{G}^{(j)}, \tilde{\mathbf{q}}^{(j)}, \tilde{\boldsymbol{\theta}}^{(j)}))} \right\}}, \quad (87)$$

choose  $(G', \mathbf{q}') = (\tilde{G}^{(i)}, \tilde{\mathbf{q}}^{(i)})$ ,  $i \in (1, 2, \dots, t)$ .

6. Implement the exchange algorithm. With the probability

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t-1)}, G', \mathbf{q}') = \min \left\{ \frac{\exp(\Phi(G, \mathbf{q}, \boldsymbol{\theta}'))}{\exp(\Phi(G, \mathbf{q}, \boldsymbol{\theta}^{(t-1)}))} \cdot \frac{\exp(\Phi(G', \mathbf{q}', \boldsymbol{\theta}^{(t-1)}))}{\exp(\Phi(G', \mathbf{q}', \boldsymbol{\theta}'))} \cdot \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(t-1)})}, 1 \right\}, \quad (88)$$

set  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}'$ . Otherwise, set  $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$ .

### G.3. Convergence of the AEX Algorithm

In this appendix, we outline the ideas and main results behind the convergence proof of the AEX algorithm. Interested readers are referred to [Jin et al. \[2013\]](#) and the supplementary material of [Liang et al. \[2015\]](#) for the completed version of proof. From Part II of the AEX algorithm, one can see that auxiliary sample  $(G', \mathbf{q}')$  is drawn via a dynamic importance sampling procedure in Equation (87), which implies that the underlying proposal distribution for auxiliary networks changes from iteration to iteration. Therefore, AEX falls into the class of adaptive MCMC algorithms with varying stationary distributions and requires an unconventional convergence theory [[Fort et al., 2011](#); [Roberts and Rosenthal, 2007](#)].

Liang et al. [2015] provided an ergodicity theorem for adaptive Markov chains with varying stationary distributions, which can be used to prove ergodicity of the AEX algorithm. Consider a state space,  $(\mathbb{X}, \mathcal{H})$ , where  $X_t \in \mathbb{X}$  denotes the state of the Markov chain at iteration  $t$  and  $\mathcal{H}$  denotes the Borel set defined on  $\mathbb{X}$ . Let  $\gamma_t$  denote realization of the adoption index  $\Gamma_t \in \mathbb{Y}$  and  $\mathcal{F}_t = \sigma(X_0, \dots, X_t, \Gamma_0, \dots, \Gamma_t)$  be the filtration generated by  $\{X_i, \Gamma_i\}_{i=0}^t$ . Let  $P_{\gamma_t}$  denote the transition kernel at iteration  $t$  and thus,

$$P(X_{t+1} \in B | X_t = x, \Gamma_t = \gamma_t, \mathcal{F}_{t-1}) = P_{\gamma_t}(x, B), \quad x \in \mathbb{X}, \gamma_t \in \mathbb{Y}, B \in \mathcal{H}.$$

Let  $P_\gamma^h(x, B) = P_\gamma(X_h \in B | X_0 = x)$  denote the  $h$ -step transition probability for the Markov chain with the fixed transition kernel  $P_\gamma$  and the initial condition  $X_0 = x$ . Also let  $P^h((x, \gamma), B) = P(X_h \in B | X_0 = x, \Gamma_0 = \gamma, \mathcal{F}_{h-1})$ ,  $B \in \mathcal{H}$ , denote the  $h$ -step transition probability for the adaptive Markov chain with the initial conditions  $X_0 = x$  and  $\Gamma_0 = \gamma$ . An adaptive Markov chain is called ergodic if  $\lim_{t \rightarrow \infty} V(x, \gamma, t) = 0$  for all  $x \in \mathbb{X}$  and  $\gamma \in \mathbb{Y}$ , where

$$V(x, \gamma, t) = \| P^t((x, \gamma), \cdot) - \pi(\cdot) \| = \sup_{B \in \mathcal{H}} \| P^t((x, \gamma), B) - \pi(B) \|$$

denotes the total variation distance between the distribution of the adaptive chain at iteration  $t$  and the target distribution  $\pi(\cdot)$ .

**Theorem 2.** *Consider an adaptive Markov chain defined on the state space  $(\mathbb{X}, \mathcal{H})$  with the adoption index  $\Gamma_t \in \mathbb{Y}$ . The adaptive Markov chain is ergodic if the following conditions are satisfied:*

- (a) (Stationarity) *There exist a stationary distribution  $\pi_{\gamma_t}(\cdot)$  for each transition kernel  $P_{\gamma_t}$ .*
- (b) (Asymptotic Simultaneous Uniform Ergodicity) *For any  $u > 0$ , there exist a measurable set  $E(u)$  in the probability space such that  $\Pr(E(u)) \geq 1 - u$  and on this set  $E(u)$ , for any  $\epsilon > 0$ , there exist  $t(\epsilon) \in \mathbb{N}$  and  $n(\epsilon) \in \mathbb{N}$  such that*

$$\sup_{x \in \mathbb{X}} \| P_{\gamma_t}^n(x, \cdot) - \pi(\cdot) \| \leq \epsilon,$$

*for all  $t > t(\epsilon)$  and  $n > n(\epsilon)$ .*

- (c) (Diminishing Adoption) *let  $D_t = \sup_{x \in \mathbb{X}} \| P_{\Gamma_{t+1}}(x, \cdot) - P_{\Gamma_t}(x, \cdot) \|$  and  $\lim_{t \rightarrow \infty} D_t = 0$  in probability.*

From conditions (a) and (b) of Theorem 2, it is implied that on the set  $E(u)$ , for any  $\epsilon$ ,

$$\| \pi_{\gamma_t}(\cdot) - \pi(\cdot) \| = \| \pi_{\gamma_t}(\cdot) P_{\gamma_t}^h - \pi(\cdot) \| \leq \sup_{x \in \mathbb{X}} \| P_{\gamma_t}^n(x, \cdot) - \pi(\cdot) \| \leq \epsilon,$$

for  $t > t(\epsilon)$  and  $n > n(\epsilon)$ . Furthermore, by the triangular inequality,

$$\| P_{\gamma_t}^n(x, \cdot) - \pi_{\gamma_t}(\cdot) \| < \| P_{\gamma_t}^n(x, \cdot) - \pi(\cdot) \| + \| \pi_{\gamma_t}(\cdot) - \pi(\cdot) \| < 2\epsilon,$$

for  $t > t(\epsilon)$  and  $n > n(\epsilon)$ . Thus, the simultaneous uniform ergodicity of  $\{P_{\gamma_t}\}$  on the set  $E(u)$  is established. Given condition (c) of Theorem 2 and the simultaneous uniform ergodicity of  $\{P_{\gamma_t}\}$ , Liang et al. [2015] show the weak law of large numbers for an adaptive Markov chain in the following theorem:

**Theorem 3.** Consider an adaptive Markov chain defined on the state space  $(\mathbb{X}, \mathcal{H})$ . Let  $g(\cdot)$  be a bounded measurable function. Suppose that conditions (a), (b), and (c) of Theorem 2 hold, then  $\frac{1}{T} \sum_{t=1}^T g(X_t) \rightarrow \pi(g)$ , in probability as  $T \rightarrow \infty$ , where  $\pi(g) = \int_{\mathbb{X}} g(x) \pi(dx)$ .

The rest of procedure is to show the AEX algorithm satisfies three conditions of Theorem 2 and thus it is ergodic and the weak law of large number holds for the average of sample path  $\{\boldsymbol{\theta}^{(t)}\}$ . To begin with, we assume that in the auxiliary chain, the Markov transition kernel,  $P_w$ , for updating the state variable  $X_t = (\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \tilde{\boldsymbol{\theta}}^{(t)}) \in \mathbb{X}$  via the SAMC algorithm satisfies the following Doeblin condition:

(A<sub>2</sub>) For any given  $w \in \varpi$ , the Markov transition kernel  $P_w$  is irreducible and aperiodic. In addition, there exist an integer  $h$ , a real number  $0 < \delta < 1$ , and a probability measure  $\nu$  such that for any compact subset  $\mathfrak{K} \subset \varpi$ ,

$$\inf_{w \in \mathfrak{K}} P_w^h(x, B) \geq \delta \nu(B), \quad \forall x \in \mathbb{X}, \forall B \in \mathcal{H},$$

where  $\mathcal{H}$  denotes the Borel set of  $\mathbb{X}$ .

The condition (A<sub>2</sub>) will be assured by Theorem 2.2 of Roberts and Tweedie [1996] if  $\mathbb{X}$  is compact, the potential function  $\Phi(G', \epsilon(\boldsymbol{\theta}), \boldsymbol{\theta})$  is bounded away from 0 and  $\infty$  on  $\mathbb{X}$ , and the proposal distribution  $T(y|x)$  satisfies the local positive condition:

(local positive condition) For every  $x \in \mathbb{X}$ , there exist  $\epsilon_1 > 0$ , and  $\epsilon_2 > 0$  such that  $|x - y| \leq \epsilon_1 \Rightarrow T(y|x) \geq \epsilon_2$ .

The following Lemma 2 shows that draws of  $(G', \mathbf{q}')$  from the dynamic importance sampler of Equation (87) converges to the distribution of  $\mu(\cdot|\boldsymbol{\theta}')$  almost surely when the number of iterations goes to infinity.

**Lemma 2.** Assume (i) conditions (A<sub>1</sub>) and (A<sub>2</sub>) are satisfied; (ii)  $\mathbb{X}$  is compact; (iii)  $\exp(\Psi(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}))$  is bounded away from 0 and  $\infty$  on  $\mathbb{X} \times \Theta$ ; Given a total of  $N$  iterations, let  $\{\mathbf{g}^{(k)}, \mathbf{q}^{(k)}\}_{k=1}^n$  be  $n$  distinct values of  $(\tilde{G}, \tilde{\mathbf{q}})$ . Re-sample a random sample  $(G', \mathbf{q}')$  from  $\Omega^{(N)}$  such that

$$P((G', \mathbf{q}') = (\mathbf{g}^{(k)}, \mathbf{q}^{(k)})|\boldsymbol{\theta}') = \frac{\sum_{t=1}^N \sum_{i=1}^m \left\{ w_i^{(t)} \frac{\exp(\Psi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}'))}{\exp(\Psi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}_i))} I\left(H(\tilde{\boldsymbol{\theta}}^{(t)}) = i \text{ and } (\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}) = (\mathbf{g}^{(k)}, \mathbf{q}^{(k)})\right) \right\}}{\sum_{t=1}^N \sum_{i=1}^m \left\{ w_i^{(t)} \frac{\exp(\Psi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}'))}{\exp(\Psi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}_i))} I\left(H(\tilde{\boldsymbol{\theta}}^{(t)}) = i\right) \right\}}, \quad (89)$$

for  $k = 1, \dots, n$ , then the distribution of  $(G', \mathbf{q}')$  converges to  $\mu(G', \mathbf{q}'|\boldsymbol{\theta}')$  almost surely as  $N \rightarrow \infty$ .

**PROOF OF LEMMA 2.** By the assumption that  $\mathbb{X}$  is compact and  $\exp(\Psi(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}))$  is bounded away from 0 and  $\infty$ , it follows from the convergence and the strong law of large numbers of SAMC [Liang et al., 2015] that

$$\begin{aligned} & \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^m \left\{ w_i^{(t)} \frac{\exp(\Psi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}'))}{\exp(\Psi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}_i))} I\left(H(\tilde{\boldsymbol{\theta}}^{(t)}) = i\right) \right\} \\ & \rightarrow \sum_{i=1}^m \int_{\mathbb{X}} \frac{\kappa(\boldsymbol{\theta}_i)}{p_i} \frac{\exp(\Psi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}'))}{\exp(\Psi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}_i))} p_i \mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_i) d\tilde{G} d\tilde{\mathbf{q}} \\ & = m\kappa(\boldsymbol{\theta}'), \quad a.s., \end{aligned} \quad (90)$$

where  $\mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_i) = \frac{\exp(\Phi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}_i))}{\kappa(\boldsymbol{\theta}_i)}$  and  $\kappa(\boldsymbol{\theta}_i) = \int_{\mathbb{X}} \exp(\Phi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}_i))$ . Similarly, for any Borel set  $A \in \mathbb{X}$ ,

$$\begin{aligned} & \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^m \left\{ w_i^{(t)} \frac{\exp(\Phi(\tilde{G}_t, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}'))}{\exp(\Phi(\tilde{G}_t, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}_i))} I \left( H(\tilde{\boldsymbol{\theta}}^{(t)}) = i \text{ and } (\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}) \in A \right) \right\} \\ & \rightarrow \sum_{i=1}^m \int_A \frac{\kappa(\boldsymbol{\theta}_i) \exp(\Phi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}'))}{p_i \exp(\Phi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}_i))} p_i \mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}_i) d\tilde{G} d\tilde{\mathbf{q}} \\ & = m \int_A \exp(\Phi(\tilde{G}, \tilde{\mathbf{q}}, \boldsymbol{\theta}')) d\tilde{G} d\tilde{\mathbf{q}}. \end{aligned} \quad (91)$$

Putting Equations (90) and (91) together, as  $N \rightarrow \infty$ , we have

$$\begin{aligned} & P((G', \mathbf{q}') \in A | \Omega^{(N)}, \boldsymbol{\theta}') \quad (92) \\ & = \frac{\sum_{t=1}^N \sum_{i=1}^m \left\{ w_i^{(t)} \frac{\exp(\Phi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}'))}{\exp(\Phi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}_i))} I \left( H(\tilde{\boldsymbol{\theta}}^{(t)}) = i \text{ and } (\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}) \in A \right) \right\}}{\sum_{t=1}^N \sum_{i=1}^m \left\{ w_i^{(t)} \frac{\exp(\Phi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}'))}{\exp(\Phi(\tilde{G}^{(t)}, \tilde{\mathbf{q}}^{(t)}, \boldsymbol{\theta}_i))} I \left( H(\tilde{\boldsymbol{\theta}}^{(t)}) = i \right) \right\}} \\ & \rightarrow \int_A \mu(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}') d\tilde{G} d\tilde{\mathbf{q}}, \quad a.s. \end{aligned} \quad (93)$$

Finally, by Lebesgue's dominated convergence theorem,

$$P((G', \mathbf{q}') \in A | \boldsymbol{\theta}') = E[P((G', \mathbf{q}') \in A | \Omega^{(N)}; \boldsymbol{\theta}')] \rightarrow \int_A f(\tilde{G}, \tilde{\mathbf{q}}|\boldsymbol{\theta}') d\tilde{G} d\tilde{\mathbf{q}}. \quad (94)$$

□

Notice that  $\{\boldsymbol{\theta}^{(t)}\}$  from the AEX algorithm form an adaptive Markov chain with the transition kernel given by

$$\begin{aligned} & \tilde{P}_l(\boldsymbol{\theta}, d\boldsymbol{\theta}') = \\ & \int_{\mathbb{X}} \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, G', \mathbf{q}') T_1(d\boldsymbol{\theta}'|\boldsymbol{\theta}) \nu_l(G', \mathbf{q}'|\boldsymbol{\theta}') dG' d\mathbf{q}' + \delta_{\boldsymbol{\theta}}(d\boldsymbol{\theta}') \left[ 1 - \int_{\Theta \times \mathbb{X}} \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, G', \mathbf{q}') T_1(d\boldsymbol{\theta}'|\boldsymbol{\theta}) \nu_l(G', \mathbf{q}'|\boldsymbol{\theta}') dG' d\mathbf{q}' \right], \end{aligned}$$

where  $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, G', \mathbf{q}')$  is defined in Equation (88),  $\delta_{\boldsymbol{\theta}}(d\boldsymbol{\theta}') = 1$  if  $\boldsymbol{\theta} \in d\boldsymbol{\theta}'$  and 0 otherwise,  $T_1(d\boldsymbol{\theta}'|\boldsymbol{\theta})$  is the proposal distribution,  $l$  denotes the cardinality of the set  $\Omega^{(l)}$ , i.e.,  $l = |\Omega^{(l)}|$ , and  $\nu_l(G', \mathbf{q}'|\boldsymbol{\theta}')$  denotes the distribution of  $(G', \mathbf{q}')$  re-sampled from  $\Omega^{(l)}$ . Different from the transition kernel of AEX, the transition kernel of the exchange algorithm depends on the perfect draw of  $(G', \mathbf{q}')$  from  $\mu(\cdot|\boldsymbol{\theta}')$ , which is given by

$$\begin{aligned} & P(\boldsymbol{\theta}, d\boldsymbol{\theta}') = \\ & \int_{\mathbb{X}} \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, G', \mathbf{q}') T_1(d\boldsymbol{\theta}'|\boldsymbol{\theta}) \mu(G', \mathbf{q}'|\boldsymbol{\theta}') dG' d\mathbf{q}' + \delta_{\boldsymbol{\theta}}(d\boldsymbol{\theta}') \left[ 1 - \int_{\Theta \times \mathbb{X}} \alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, G', \mathbf{q}') T_1(d\boldsymbol{\theta}'|\boldsymbol{\theta}) \mu(G', \mathbf{q}'|\boldsymbol{\theta}') dG' d\mathbf{q}' \right], \end{aligned}$$

where  $\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}, G', \mathbf{q}')$  is defined in Equation (31)

**Lemma 3.** Assume that (i) conditions  $(A_1)$  and  $(A_2)$  are satisfied; (ii) both  $\Theta$  and  $\mathbb{X}$  are compact; (iii) the potential function  $\Phi(G', \boldsymbol{\epsilon}(\boldsymbol{\theta}), \boldsymbol{\theta})$  is continuously differentiable in  $\boldsymbol{\theta}$  for all  $G' \in \mathbb{X}$  and bounded away from 0 and  $\infty$  on  $\mathbb{X} \times \Theta$ ; (iv)  $\pi(\boldsymbol{\theta})$  and  $q(\boldsymbol{\theta}'|\boldsymbol{\theta})$  are continuously differentiable in  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}'$ . Define  $\tilde{D}_l = \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{P}_l(\boldsymbol{\theta}, \cdot) - P(\boldsymbol{\theta}, \cdot)\|$ , then  $\tilde{D}_l \rightarrow 0$  almost surely as  $l \rightarrow \infty$ .

Define  $D_l = \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{P}_{l+1}(\boldsymbol{\theta}, \cdot) - \tilde{P}_l(\boldsymbol{\theta}, \cdot)\|$ . Since  $D_l \leq \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{P}_{l+1}(\boldsymbol{\theta}, \cdot) - P(\boldsymbol{\theta}, \cdot)\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\tilde{P}_l(\boldsymbol{\theta}, \cdot) - P(\boldsymbol{\theta}, \cdot)\|$ , we have  $\lim_{l \rightarrow \infty} D_l = 0$  almost surely by Lemma 3. Thus,  $\tilde{P}_l$

satisfies condition (c) of Theorem D1. It is known that the transition kernel of the exchange algorithm,  $P(\boldsymbol{\theta}, d\boldsymbol{\theta}')$ , is irreducible and aperiodic and admits an invariant limit distribution. Lemma 4 shows that the transition kernel of the AEX algorithm,  $\tilde{P}_l(\boldsymbol{\theta}, d\boldsymbol{\theta}')$ , also has these properties and thus satisfies condition (a) of Theorem 2.

**Lemma 4.** *Assume (i) the conditions of Lemma 3 are satisfied; (ii)  $P$  is irreducible and aperiodic and admits an invariant distribution, then  $\tilde{P}_l$  is irreducible and aperiodic, and hence exists a stationary distribution  $\pi_l(\boldsymbol{\theta}|G', \mathbf{q}')$  such that for any  $\boldsymbol{\theta}_0 \in \Theta$ ,*

$$\lim_{k \rightarrow \infty} \|\tilde{P}_l^k(\boldsymbol{\theta}_0, \cdot) - \pi_l(\cdot|G', \mathbf{q}')\| = 0.$$

Lemma 5 establishes the asymptotic simultaneous uniform ergodicity of the kernel  $\tilde{P}_l$ 's.

**Lemma 5.** *Assume the conditions of Lemma 4 are satisfied. If the proposal  $q(\cdot, \cdot)$  satisfies the local positive condition, then for any  $e > 0$ , there exist a measurable set  $E(e)$  in the probability space such that  $P(E(e)) > 1 - e$  and on this set  $E(e)$ , for any  $\epsilon > 0$ , there exist  $n(\epsilon) \in \mathbb{N}$  and  $l(\epsilon) \in \mathbb{N}$  such that for any  $n > n(\epsilon)$  and  $l > l(\epsilon)$ ,*

$$\|\tilde{P}_l^n(\boldsymbol{\theta}_0, \cdot) - \pi(\cdot|G, \mathbf{q})\| \leq \epsilon, \quad \text{for all } \boldsymbol{\theta}_0 \in \Theta.$$

Putting Lemmas 3, 4, and 5 altogether, we have the following theorem in regard to ergodicity and the weak law of large number for AEX.

**Theorem 4.** *Assume the conditions of Lemma 5 hold. If the proposal  $T_1(\cdot, \cdot)$  satisfies the local positive condition and the unnormalized density function  $\exp(\Phi(G', \mathbf{q}', \boldsymbol{\theta}))$  is upper semi-continuous in  $\boldsymbol{\theta}$  for all  $(G', \mathbf{q}') \in \mathbb{X}$ , then the adaptive exchange algorithm is ergodic and for any bounded measurable function  $f$ ,*

$$\frac{1}{T} \sum_{t=1}^T f(\boldsymbol{\theta}^{(t)}) \rightarrow \pi(f|G, \mathbf{q}), \quad \text{in probability,}$$

as  $T \rightarrow \infty$ , where  $\pi(f|G, \mathbf{q}) = \int_{\Theta} f(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|G, \mathbf{q})d\boldsymbol{\theta}$ .

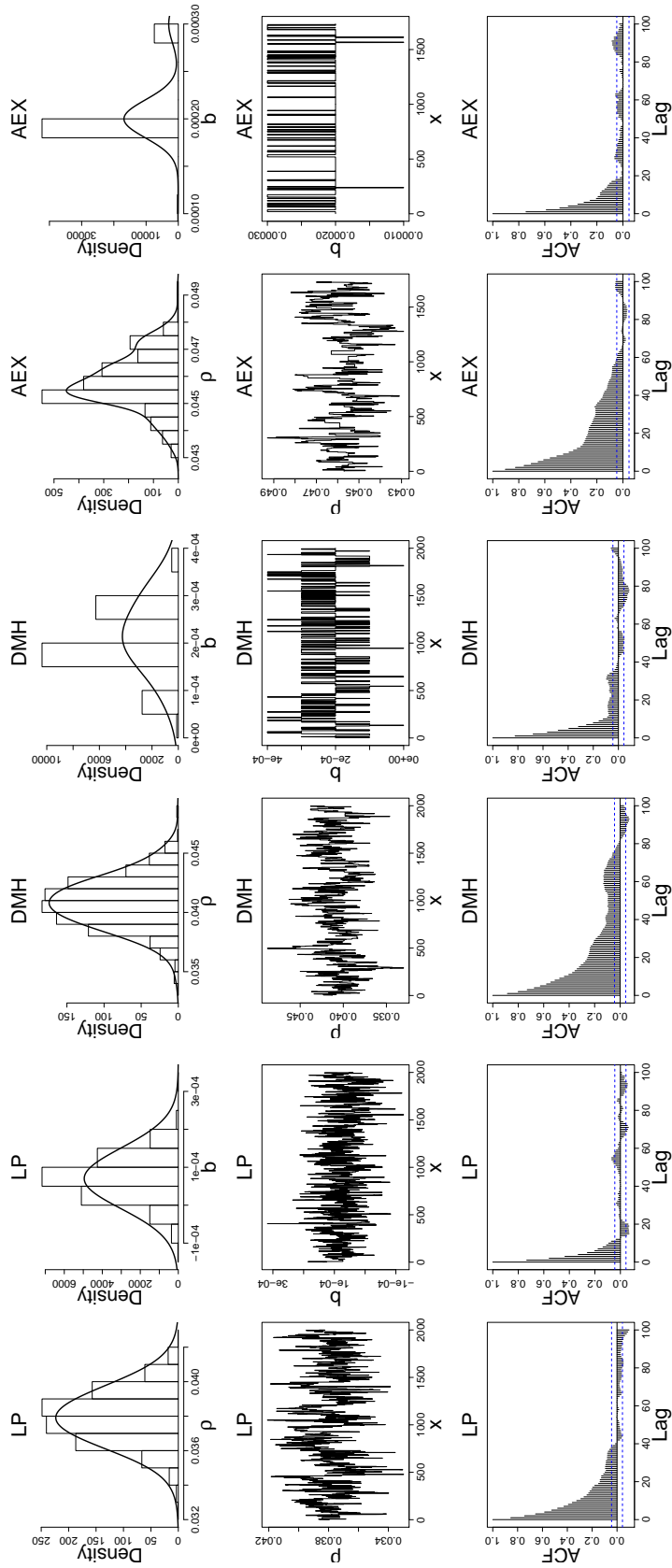


Figure 1: Plots of MCMC draws for  $\rho$  and  $b$  under the SIC-28 subsample.

#### G.4. Fixed Network

When the network link update rates vanish (by setting the rates  $\tau = \xi = 0$  of the stochastic process in Definition 1), i.e. the network is fixed, or just looking at a time interval between any two link changes where all firms adjust output levels instantaneously, we can study firms' output decisions conditional on a given fixed network configuration  $G$ . Assuming that the inverse noise parameter  $\vartheta$  is large, output levels chosen by firms will be close to the potential maximizer  $\mathbf{q}^*$ . We can then make the following Taylor expansion for the potential function

$$\Phi(\mathbf{q}, G) = \Phi(\mathbf{q}^*, G) + (\mathbf{q} - \mathbf{q}^*)^\top \nabla \Phi(\mathbf{q}^*, G) + \frac{1}{2}(\mathbf{q} - \mathbf{q}^*)^\top \Delta \Phi(\mathbf{q}^*, G)(\mathbf{q} - \mathbf{q}^*) + o(\|\mathbf{q} - \mathbf{q}^*\|).$$

Noting that the gradient vanishes at  $\mathbf{q}^*$ , i.e.,  $\nabla \Phi(\mathbf{q}^*, G) = 0$ , and dropping terms of the order  $o(\|\mathbf{q} - \mathbf{q}^*\|)$  we can write

$$\Phi(\mathbf{q}, G) \approx \Phi(\mathbf{q}^*, G) + \frac{1}{2}(\mathbf{q} - \mathbf{q}^*)^\top \Delta \Phi(\mathbf{q}^*, G)(\mathbf{q} - \mathbf{q}^*).$$

Inserting the above equation into the Gibbs measure gives

$$\mu^\vartheta(\mathbf{q}, G) = \frac{1}{\mathcal{Z}_\vartheta} e^{\vartheta \Phi(\mathbf{q}, G)} \approx \frac{1}{\mathcal{Z}_\vartheta} e^{\vartheta \Phi(\mathbf{q}^*, G)} \exp \left\{ \frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top \Delta \Phi(\mathbf{q}^*, G)(\mathbf{q} - \mathbf{q}^*) \right\}.$$

The partition function can be written as

$$\begin{aligned} \mathcal{Z}_\vartheta &= \int_{\mathcal{Q}^n} e^{\vartheta \Phi(\mathbf{q}, G)} d\mathbf{q} \\ &\approx e^{\vartheta \Phi(\mathbf{q}^*, G)} \int_{\mathcal{Q}^n} \exp \left\{ \frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top \Delta \Phi(\mathbf{q}^*, G)(\mathbf{q} - \mathbf{q}^*) \right\} d\mathbf{q} \\ &= e^{\vartheta \Phi(\mathbf{q}^*, G)} \left( \frac{2\pi}{\vartheta} \right)^{\frac{n}{2}} |\nabla \Phi(\mathbf{q}^*, G)|^{-\frac{1}{2}}. \end{aligned}$$

Therefore, the above Laplace approximation of the Gibbs measure yields [Wong, 2001, Theorem 3, p. 495]

$$\mu^\vartheta(\mathbf{q}, G) \approx \left( \frac{2\pi}{\vartheta} \right)^{-\frac{n}{2}} |\nabla \Phi(\mathbf{q}^*, G)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \vartheta (\mathbf{q} - \mathbf{q}^*)^\top (-\Delta \Phi(\mathbf{q}^*, G))(\mathbf{q} - \mathbf{q}^*) \right\}.$$

where the potential function  $\Phi(\mathbf{q}, G)$  is given by Equation (25). The gradient can be written as  $\nabla \Phi(\mathbf{q}, G) = \boldsymbol{\eta} - \mathbf{M}(G)\mathbf{q}$  and the Hessian is given by  $\Delta \Phi(\mathbf{q}, G) = -\mathbf{M}(G)$ . The FOCs are then given by  $\nabla \Phi(\mathbf{q}^*, G) = \boldsymbol{\eta} - \mathbf{M}(G)\mathbf{q}^* = 0$  so that  $\mathbf{q}^* = \mathbf{M}(G)^{-1}\mathbf{X}\boldsymbol{\delta}$  when the matrix  $\mathbf{M}(G)$  is invertible and  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\delta}$ .<sup>79</sup> Hence, the density of output levels  $\mathbf{q}$  with the (exogenous) network  $G$  can be rewritten as

$$\mu^\vartheta(\mathbf{q}|G) \approx \left( \frac{2\pi}{\vartheta} \right)^{-\frac{n}{2}} |\mathbf{M}(G)|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{q} - \mathbf{M}(G)^{-1}\mathbf{X}\boldsymbol{\delta})^\top \mathbf{M}(G)(\mathbf{q} - \mathbf{M}(G)^{-1}\mathbf{X}\boldsymbol{\delta}) \right\}, \quad (95)$$

which implies that the output  $\mathbf{q}$ , conditional on the R&D network  $G$ , follows a Gaussian normal density function with mean  $\mathbf{M}(G)^{-1}\mathbf{X}\boldsymbol{\delta}$  and variance  $\mathbf{M}(G)^{-1}$ .

<sup>79</sup>The existence of a unique potential maximizer  $\mathbf{q}^*$  is guaranteed when the output levels are bounded and the matrix  $\mathbf{M}(G)$  is positive definite [Byong-Hun, 1983; König et al., 2014b].



## G.5. Matrix Perturbation

For the computational implementation of the MCMC algorithm we will have to evaluate the inverse of  $\mathbf{M}(G)$  whenever a link has been added or removed in the network  $G$ . To do this in an efficient way, the following lemma will be helpful.

**Lemma 6.** *Let  $\mathbf{e}_i$  and  $\mathbf{e}_j$  be the  $i$ -th and  $j$ -th unit basis vectors in  $\mathbb{R}^n$ . The matrix  $\mathbf{e}_i\mathbf{e}_j^\top$  has a one in the  $ij$ -th position and zeros everywhere else. Define the matrix*

$$\mathbf{B}_{ij} \equiv \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}.$$

*Adding a perturbation  $\alpha$  to the matrix  $\mathbf{A}$  in the  $ij$ -th and the  $ji$ -th position can be written as  $\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \rho\mathbf{e}_j\mathbf{e}_i^\top$ .*

(i) *The inverse of the perturbed matrix can be written as*

$$(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top)^{-1} = \mathbf{A}^{-1} - \alpha\mathbf{B}_{ij} - \alpha \frac{(\mathbf{A}^{-1} - \alpha\mathbf{B}_{ij})\mathbf{e}_j\mathbf{e}_i^\top(\mathbf{A}^{-1} - \alpha\mathbf{B}_{ij})}{1 + \alpha\mathbf{e}_i^\top(\mathbf{A}^{-1} - \alpha\mathbf{B}_{ij})\mathbf{e}_j}. \quad (96)$$

(ii) *The determinant of the perturbed matrix can be written as*

$$\det(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top) = \left(1 + \alpha\mathbf{e}_i^\top(\mathbf{A}^{-1} - \alpha\mathbf{B}_{ij})\mathbf{e}_j\right) \left(1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i\right) \det(\mathbf{A}). \quad (97)$$

PROOF OF LEMMA 6. We first give a proof of part (i) of the lemma. The Sherman–Morrison formula states that [Meyer, 2000, page 124]

$$(\mathbf{A} + \alpha\mathbf{u}\mathbf{v}^\top)^{-1} = \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u}}.$$

Let  $\mathbf{c} = \mathbf{e}_i$  and  $\mathbf{d} = \mathbf{e}_j$ , where  $\mathbf{e}_i$  and  $\mathbf{e}_j$  are the  $i$ -th and  $j$ -th unit basis vectors, respectively. The matrix  $\mathbf{c}\mathbf{d}^\top$  then has a one in the  $(i, j)$ -position and zeros elsewhere, so that adding a one to the matrix  $\mathbf{A}$  in the  $(i, j)$ -position and the  $(j, i)$ -position (resulting from adding a link  $ij$  to the adjacency matrix  $\mathbf{A}$ ) yields a perturbed matrix  $\mathbf{B}$  which can be written as

$$\mathbf{B} = \mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top = \mathbf{C} + \alpha\mathbf{e}_j\mathbf{e}_i^\top,$$

where we have denoted by  $\mathbf{C} \equiv \mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top$ . Using the Sherman–Morrison formula we then can write

$$\mathbf{B}^{-1} = (\mathbf{C} + \alpha\mathbf{e}_j\mathbf{e}_i^\top)^{-1} = \mathbf{C}^{-1} - \alpha \frac{\mathbf{C}^{-1}\mathbf{e}_j\mathbf{e}_i^\top\mathbf{C}^{-1}}{1 + \alpha\mathbf{e}_i^\top\mathbf{C}^{-1}\mathbf{e}_j}, \quad (98)$$

while applying Sherman–Morrison to  $\mathbf{C}^{-1}$  gives

$$\mathbf{C}^{-1} = (\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top)^{-1} = \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}.$$

Inserting  $\mathbf{C}^{-1}$  into Equation (98) yields

$$\mathbf{B}^{-1} = \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i} - \alpha \frac{\left(\mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)}{1 + \alpha\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j}.$$

We next give a proof of part (ii) of the lemma. We have that  $\det(\mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top + \alpha \mathbf{e}_j \mathbf{e}_i^\top) = \det(\mathbf{C} + \alpha \mathbf{e}_j \mathbf{e}_i^\top)$ , where we have denoted by  $\mathbf{C} \equiv \mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top$ . The matrix determinant lemma states that [Horn and Johnson, 1990]

$$\det(\mathbf{A} + \mathbf{u} \mathbf{v}^\top) = (1 + \mathbf{v}^\top \mathbf{A}^{-1} \mathbf{u}) \det(\mathbf{A}).$$

It then follows that

$$\det(\mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top + \alpha \mathbf{e}_j \mathbf{e}_i^\top) = \det(\mathbf{C} + \alpha \mathbf{e}_j \mathbf{e}_i^\top) = (1 + \alpha \mathbf{e}_i^\top \mathbf{C}^{-1} \mathbf{e}_j) \det(\mathbf{C}),$$

Similarly, from the matrix determinant lemma, we have that

$$\det(\mathbf{C}) = \det(\mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top) = (1 + \alpha \mathbf{e}_j^\top \mathbf{A}^{-1} \mathbf{e}_i) \det(\mathbf{A}).$$

Further, the Sherman–Morrison formula states that [Meyer, 2000, page 124]

$$(\mathbf{A} + \alpha \mathbf{c} \mathbf{d}^\top)^{-1} = \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1} \mathbf{u} \mathbf{v}^\top \mathbf{A}^{-1}}{1 + \alpha \mathbf{v}^\top \mathbf{A}^{-1} \mathbf{u}}.$$

It follows that

$$\mathbf{C}^{-1} = (\mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top)^{-1} = \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1} \mathbf{e}_i \mathbf{e}_j^\top \mathbf{A}^{-1}}{1 + \alpha \mathbf{e}_j^\top \mathbf{A}^{-1} \mathbf{e}_i}.$$

Putting the above results together, we find that

$$\begin{aligned} \det(\mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top + \alpha \mathbf{e}_j \mathbf{e}_i^\top) &= (1 + \alpha \mathbf{e}_i^\top \mathbf{C}^{-1} \mathbf{e}_j) \det(\mathbf{C}) \\ &= \left( 1 + \alpha \mathbf{e}_i^\top \left( \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1} \mathbf{e}_i \mathbf{e}_j^\top \mathbf{A}^{-1}}{1 + \alpha \mathbf{e}_j^\top \mathbf{A}^{-1} \mathbf{e}_i} \right) \mathbf{e}_j \right) (1 + \alpha \mathbf{e}_j^\top \mathbf{A}^{-1} \mathbf{e}_i) \det(\mathbf{A}). \end{aligned}$$

□

Note that due to Lemma 6 the determinant and the inverse of the perturbed matrix  $\mathbf{A} + \alpha \mathbf{e}_i \mathbf{e}_j^\top + \alpha \mathbf{e}_j \mathbf{e}_i^\top$  can be efficiently computed if the determinant,  $\det(\mathbf{A})$ , and the inverse,  $\mathbf{A}^{-1}$ , of  $\mathbf{A}$  are already known.

## H. Performance of Estimation Methods: Consistency, computation time and missing data

In this paper we consider three types of Bayesian MCMC estimation methods, the likelihood partition (LP), the double Metropolis-Hastings (DMH), and the adaptive exchange (AEX) algorithm to estimate our model. In order to demonstrate the performance of each method, we conduct a small-scale simulation study to show that (i) the true parameter values can be obtained from each method when estimating the correct model, and (ii) to illustrate the relative computation cost of each method. For the first part, as known from the Bayesian identification literature [see, e.g., Berger, 1985, p.224], the Bayesian asymptotic theory related to the posterior distribution is identical to the asymptotic distribution theory for the maximum likelihood estimator (MLE). Therefore, the posterior distribution of parameters should be concentrated at the true value, i.e., consistent when the sample size goes to infinity (where the role of the prior vanishes), just like MLE does. Regarding the latter, since estimating ERGMs is usually computationally costly, we report the computation time (under our simulation design) so that potential users can better gauge the feasibility of our methods when applied to their own data.

In this simulation, we generate an artificial network ( $G$ ) and output data ( $\mathbf{q}$ ) from the data generating process (DGP) described in Definition 1. We consider two network sizes ( $n$ ) – 100 and 200 – for comparing posterior distributions under different sample sizes. In the DGP, we capture individual exogenous heterogeneity in the output process by  $\beta x$ , where the variable  $x$  is generated from a normal distribution,  $\mathcal{N}(0, 4)$ , and the coefficient  $\beta$  is set to one. The true value of the complementarity effect  $\rho$  is set to 0.05 and the true value of the substitution effect  $b$  is set to 0.01. Exogenous linking costs for each network link ( $a_{ij}$ ) are captured by a constant term and two dyad-specific exogenous variables  $c_1$  and  $c_2$  generated as follows: For  $c_1$ , we first draw  $n$  discrete uniform variables  $H$  from  $[1, 4]$ . If  $H_{1i}$  and  $H_{1j}$  are the same, then we set  $c_{1,ij} = 1$ . Otherwise, we set it to zero. For  $c_2$ , we use the exogenous variable  $x$  from the output process and define  $c_{2,ij} = |x_i - x_j|$ . The parameters assigned to the cost function  $\gamma_0 + \gamma_1 c_{1,ij} + \gamma_2 c_{2,ij}$  are set to  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2) = (-7, 2, 1)$ . We normalize other parameters in Definition 1 as follows:  $\vartheta = 1$ ,  $\nu = 0.5$ ,  $\lambda = 1$ ,  $\xi = 1$ ,  $\chi = 1$ .

Given the artificial data, in terms of the output levels  $\mathbf{q}$  and the network  $G$ , we first estimate the exact same model in the DGP and simulate 50,000 MCMC draws from the conditional posterior distribution  $P(\rho, b, \boldsymbol{\gamma}', \beta | \mathbf{q}, G)$  under the LP, DMH, and AEX methods discussed in Section 3.3 of the paper. We initially code all of the three methods in Matlab. However, we found Matlab was not ideal for handling the heavy sequential computations that are part of the DMH and AEX algorithms. Therefore, we recode the DMH and AEX methods in Fortran. The average computation time for a single MCMC iteration (measured in seconds), which is executed on a single workstation with dual Intel Xeon 2.60 GHz CPUs, is reported in Figure 1. One can see that when the network size increases, the computation time required by the DMH and AEX algorithms increases dramatically compared to the LP algorithm, even though we have adopted a high level programming language such as Fortran in order to gain computational efficiency. Moreover, we illustrate the distributions of the posterior draws in Figure 2. One can see that when the sample size increases from 100 to 200, the draws of  $(\rho, b, \boldsymbol{\gamma}')$  become more concentrated around the true values.

Furthermore, we use this simulation environment to examine the impact of missing observations on the estimation results. We take the artificial network and output sample with size equal to 200 generated from the DGP described above and conduct 100 Monte Carlo repetitions. In each repetition, we remove different amounts of data, 25%, 50%, and 75% of the total observations, respectively, at random and estimate the model using the LP approach. We compute the posterior mean from MCMC draws as the point estimate for each parameter and report the mean and the standard deviation across the Monte Carlo repetitions in Table 1. We find that the bias and volatility of the point estimates increase with increasing amounts of missing observations. However, from the most severe case (75% missing) that we consider, in which only 50 nodes were left in our sample, the average biases are maintained below 15% for the complementarity effect  $\rho$ ; below 72% for the substitution effect  $b$ ; and generally below 12% for the other parameters in the linking cost function. From this simulation exercise we therefore find that the bias incurred from missing data is relatively weak.

## I. Analysis of the SIC-36 Sector – Electronic and Other Electric Equipment

The SIC-36 sector is the second largest 2-digit SIC sector in our sample. It consists of 198 firms and 29 R&D collaboration links. Excluding the sectors without any observation, we obtain 8 sub-

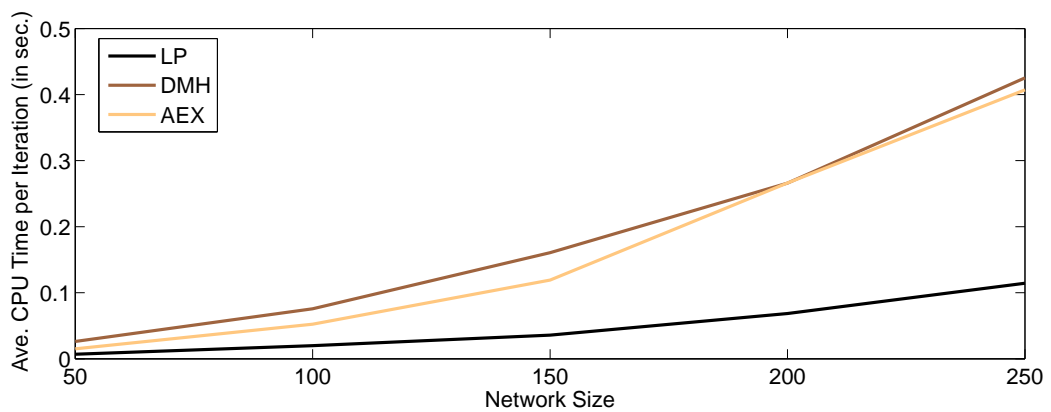


Figure 1: The average computation time for a single MCMC iteration (measured in seconds).

Table 1: Monte Carlo simulation results based on increasing levels of missing data.

	DGP	25% missing		50%missing		75% missing	
		mean	s.d.	mean	s.d.	mean	s.d.
$\rho$	0.0500	0.0578	0.0070	0.0677	0.0108	0.0414	0.0156
$b$	0.0100	0.0098	0.0020	0.0123	0.0033	0.0172	0.0040
$\gamma_0$	-7.0000	-7.0992	0.1374	-7.1463	0.2371	-7.4734	0.7513
$\gamma_1$	2.0000	1.9797	0.0807	1.9851	0.1847	2.0692	0.3760
$\gamma_2$	1.0000	1.0372	0.0373	1.0450	0.0623	1.1196	0.1894

Note: The number of repetitions for each simulation is set to 100. The true parameters are provided in the first column. We consider different levels of missing data: 25%, 50%, and 75%. The mean and the standard deviation of the point estimates across 100 repetitions are reported.

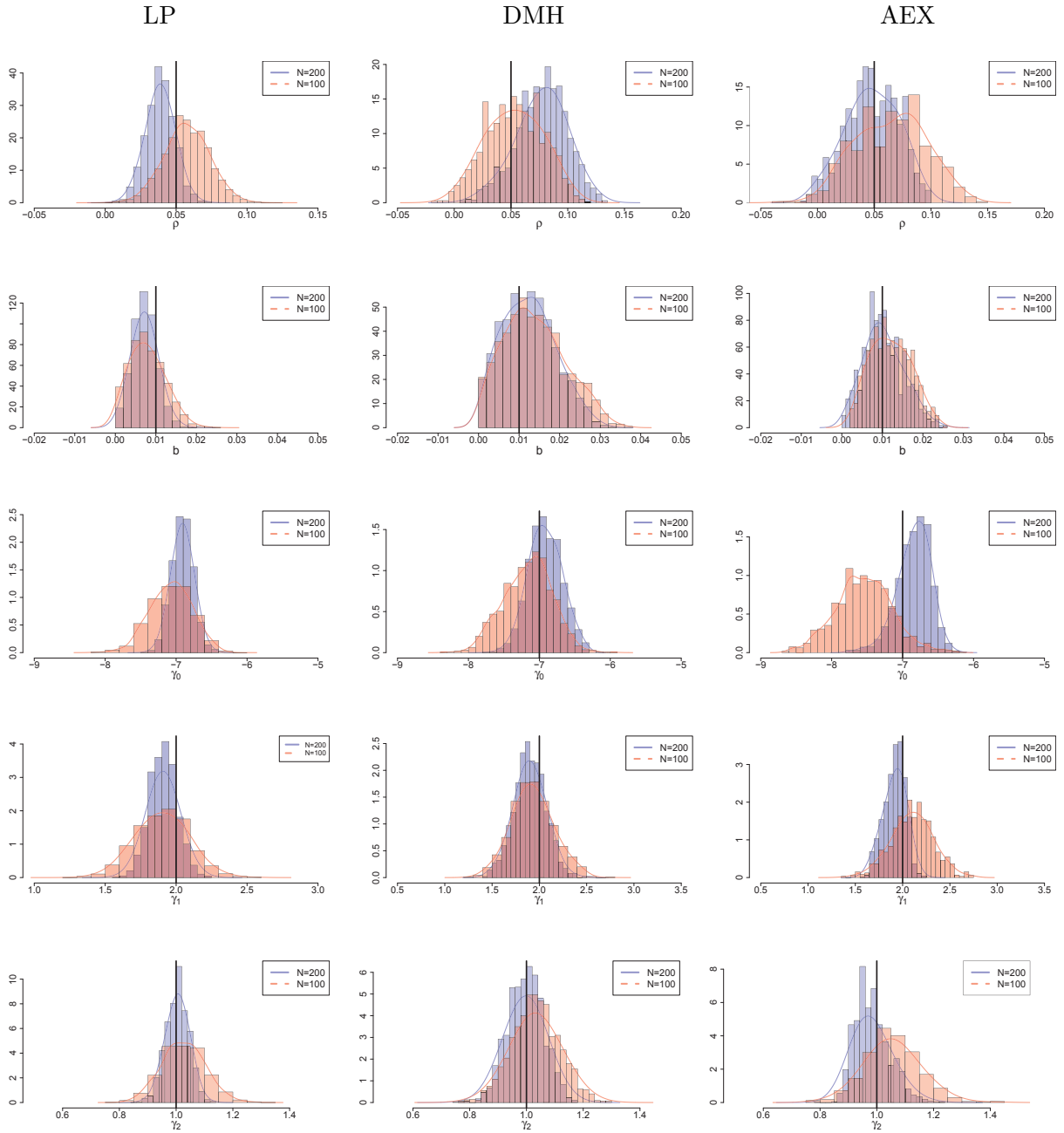


Figure 2: The distribution of the parameters  $\rho$ ,  $b$ , and  $\gamma$  across a total of 50,000 MCMC draws from the conditional posterior distributions using the likelihood partition (LP) approach (left column), the double Metropolis-Hastings (DMH) algorithm (middle column) and the AEX algorithm (right column) discussed in Section 3.3.

Table 1: Descriptive statistics.

Sample	# of firms	Log R&D Expenditure			Productivity			Log # of Patents		
		mean	min	max	mean	min	max	mean	min	max
SIC-36	198	9.8114	4.5667	14.8850	1.5414	0.0019	9.6015	5.1430	0.0000	11.8726
SIC-361	16	8.8642	7.1316	12.9694	0.7264	0.2081	1.3740	4.8608	3.0445	6.1137
SIC-362	11	9.6755	6.2545	11.9781	0.9454	0.0177	1.9536	5.9453	3.1781	9.5412
SIC-364	5	9.6953	6.6545	11.3892	1.0237	0.3052	1.3553	5.9174	1.3863	8.3134
SIC-365	9	10.6156	8.3330	13.6045	1.0857	0.1055	2.7267	6.9759	4.3694	10.7241
SIC-366	45	9.5828	4.8083	14.7500	1.8621	0.1967	4.6009	4.4755	0	11.8726
SIC-367	100	9.8987	4.5667	14.8850	1.7027	0.0019	9.6015	5.0129	0	10.4502
SIC-369	12	8.6530	6.0535	10.5134	0.8442	0.0068	3.7974	4.0802	0.6931	7.5358

Note: The logarithm of a firm's R&D expenditures (by thousand dollars) measures its R&D effort. A firm's productivity is measured by the ratio of sales to employment. The log of the number of patents is used as a control variable in the linking cost function [Hanaki et al., 2010].

Table 2: The number of R&D collaborations within and across the subsectors of the SIC-36 sector.

	364	361	365	362	369	366	367
364	0	1	0	0	0	0	0
361	1	2	5	0	0	2	4
365	0	5	0	0	0	0	0
362	0	0	0	0	0	0	0
369	0	0	0	0	0	0	0
366	0	2	0	0	0	3	5
367	0	4	0	0	0	5	7

sectors within SIC-36, ranging from “electronic transmission and distribution equipment” (SIC-361) to “miscellaneous electrical machinery, equipment, and supplies” (SIC-369). The summary statistics of these sectors are provided in Table 1 and the number of R&D collaborations within the SIC-36 sector is reported in Table 2. Among the sub-sectors within SIC-36, the electronic components and accessories sector (SIC-367) is the largest one. It consists of 100 firms and 7 within sectoral R&D collaborations.

The estimation results for the SIC-36 sector are reported in Table 3. We obtain three estimates based on, respectively, the LP, the DMH, and the AEX algorithms. The estimates for the technology spillover effect  $\rho$  and the competition effect  $b$  across the three different estimation algorithms are similar and significant, which confirms the robustness of our estimates. Compared to the results of the SIC-28 sector in Table 2, the technology spillover effects in the SIC-36 sector is 1.7 times larger (based on DMH estimates), while the size of competition effect is similar in the two sectors. However, for the R&D linking cost, due to a smaller sample size and spare sample variation on the network, we only found a robust significant effect for the sum of log patent numbers on the linking cost.

Table 3: Estimation results based on the SIC-36 sector.

		LP	DMH	AEX
R&D Spillover	$(\rho)$	0.0612*** (0.0043)	0.0700*** (0.0058)	0.0696*** (0.0042)
Substitutability	$(b)$	0.0001* (0.0001)	0.0002* (0.0001)	0.0001* (0.0001)
Prod.	$(\delta_1)$	0.0634** (0.0287)	0.0762 (0.0756)	0.0681 (0.0589)
Sector FE	$(\delta_2)$	Yes	Yes	Yes

---

**Linking Cost**

Constant	$(\gamma_0)$	16.3972*** (0.3329)	17.7430*** (1.3463)	17.9334*** (1.7124)
Same Sector	$(\gamma_1)$	-0.7698* (0.4309)	-1.2118* (0.7709)	-0.1174 (0.9089)
Same Country	$(\gamma_2)$	-0.9180** (0.4273)	-0.5550 (0.5693)	-0.5854 (0.6309)
Diff-in-Prod.	$(\gamma_3)$	-0.5287 (0.4550)	-0.4750 (0.5576)	-0.2123 (0.5637)
Diff-in-Prod. Sq.	$(\gamma_4)$	0.0964 (0.0988)	0.1313 (0.1384)	0.0702 (0.1023)
Patents	$(\gamma_5)$	-0.0597* (0.0312)	-0.1469** (0.0715)	-0.1680*** (0.0674)

---

Sample size 198

Note: The dependent variable is log R&D expenditures. The parameters  $\theta = (\rho, b, \delta^\top, \gamma^\top, \varkappa)$  correspond to Equation (24), where  $\psi_{ij} = \gamma^\top \mathbf{c}_{ij}$  and  $\eta_i = \mathbf{X}_i \delta$  (see Section 3.2). We make 50,000 MCMC draws where we drop the first 2,000 draws during a burn-in phase and keep every 20th of the remaining draws 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 its 99% (95%, 90%) highest posterior density range does not cover zero.

## References

- Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3):425–455.
- Anderson, S., Goeree, J., and Holt, C. (2004). Noisy directional learning and the logit equilibrium. *Scandinavian Journal of Economics*, 106(3):581–602.
- Anderson, S. P., De Palma, A., and Thisse, J. F. (1992). *Discrete choice theory of product differentiation*. MIT Press.
- Anderson, S. P., Goeree, J. K., and Holt, C. A. (2001). Minimum-effort coordination games: Stochastic potential and logit equilibrium. *Games and Economic Behavior*, 34(2):177–199.
- Atkinson, K. E. (1989). *An introduction to numerical analysis*. John Wiley & Sons.
- Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who’s who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417.
- Banerjee, A. and Duflo, E. (2005). Growth theory through the lens of development economics. *Handbook of Economic Growth*, 1:473–552.
- 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*.
- Belderbos, R., Carree, M., and Lokshin, B. (2004). Cooperative r&d and firm performance. *Research policy*, 33(10):1477–1492.
- Ben-Akiva, M. and Watanatada, T. (1981). Application of a continuous spatial choice logit model. *Structural analysis of discrete data with econometric applications*, pages 320–343.
- Bena, J., Fons-Rosen, C., and Ondko, P. (2008). Zephyr: Ownership changes database. *London School of Economics working paper*.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media.
- 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.
- Bernstein, J. I. (1988). Costs of production, intra-and interindustry R&D spillovers: Canadian evidence. *Canadian Journal of Economics*, 21(2):324–347.
- 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’Institut Mathématique*, 53:45–54.
- Byong-Hun, A. (1983). Iterative methods for linear complementarity problems with upper bounds on primary variables. *Mathematical Programming*, 26(3):295–315.
- Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435.
- Chib, S. and Greenberg, E. (1996). Markov chain Monte Carlo simulation methods in econometrics. *Econometric theory*, 12(03):409–431.
- Cohen, W. and Klepper, S. (1996a). Firm size and the nature of innovation within industries: the case of process and product R&D. *The 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.
- Copeland, A. and Fixler, D. (2012). Measuring the price of research and development output. *Review of Income and Wealth*, 58(1):166–182.
- Cvetkovic, D. and Rowlinson, P. (1990). The largest eigenvalue of a graph: A survey. *Linear and Multilinear Algebra*, 28(1):3–33.
- Dai, R. (2012). International accounting databases on WRDS: Comparative analysis. *Working paper, Wharton Research Data Services, University of Pennsylvania*.
- D’Aspremont, C. and Jacquemin, A. (1988). Cooperative and noncooperative R&D in duopoly with spillovers. *The American Economic Review*, 78(5):1133–1137.
- 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*.
- D’Haultfoeuille, X. and Iaria, A. (2016). A convenient method for the estimation of the multinomial logit model with fixed effects. *Economics Letters*, 141:77–79.
- 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.
- Fort, G., Moulines, E., and Priouret, P. (2011). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *The Annals of Statistics*, 36(4):3262–3289.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, pages 169–193.
- Graham, B. S. (2014). An empirical model of network formation: detecting homophily when agents are heterogeneous. *National Bureau of Economic Research Working Paper No. w20341*.
- Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1):465–485.
- 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. 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.
- Jin, I. H., Yuan, Y., and Liang, F. (2013). Bayesian analysis for exponential random graph models using the adaptive exchange sampler. *Statistics and its interface*, 6(4):559.
- Kamien, M. I., Muller, E., and Zang, I. (1992). Research joint ventures and R&D cartels. *The American Economic Review*, 82(5):1293–1306.
- König, M., 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. (2014b). R&D networks: Theory, empirics and policy implications. *CEPR Discussion Paper No. 9872*.
- 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.
- Liang, F., Jin, I. H., Song, Q., and Liu, J. S. (2015). An adaptive exchange algorithm for sampling from distributions with intractable normalizing constants. *Journal of the American Statistical Association*, 111(513):377–393.
- 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.
- McFadden, D. (1976). The mathematical theory of demand models. *Behavioral Travel Demand Models*, pages 305–314.
- McManus, O., Blatz, A., and Magleby, K. (1987). Sampling, log binning, fitting, and plotting durations of open and shut intervals from single channels and the effects of noise. *Pflügers Archiv*, 410(4-5):530–553.
- Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3):825–850.
- 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 Maggione, 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.
- 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*, volume 2012, pages 1–1. Academy of Management.
- 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., Koput, K. W., and Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1):116–145.
- 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 Raftery, A. E. (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, page 4.
- Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Verlag.
- Roberts, G. O. and Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of applied probability*, 44(2):458–475.
- Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110.
- Roijackers, N. and Hagedoorn, J. (2006). Inter-firm R&D partnering in pharmaceutical biotechnology since 1975: Trends, patterns, and networks. *Research Policy*, 35(3):431–446.
- Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal*, 30(3):233–260.
- Singh, N. and Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics*, 15(4):546–554.
- Snijders, T. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.
- Söderberg, B. (2002). General formalism for inhomogeneous random graphs. *Physical Review E*, 66(6):066121.
- Spence, M. (1984). Cost reduction, competition, and industry performance. *Econometrica*, 52(1):101–122.
- 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*.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, 22(4):1701–1728.
- 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*, volume 34. Society for Industrial Mathematics.