Collaborate or Consolidate? A NLP Text-Mining Analysis of R&D Networks and M&A

Chih-Sheng Hsieh*, Michael D. König†, Gordon M. Phillips‡ and Jakob Rauch§

March 11, 2025

Abstract

We build a novel dataset of R&D collaborations between firms by text mining news articles. Our data set covers several decades and represents the largest currently existing database on R&D collaborations. The comprehensiveness of our data allows us to document differences in R&D collaboration activities across firm characteristics, industries, countries, or regions and how these differences change over time. We then estimate a structural model capturing the joint endogeneity of both R&D collaborations and M&A activities to identify technology spillovers and market competition effects in firm performance and how these change over time with the evolution of technologies and markets. Our results indicate that the changing pattern in technology spillovers is due to firms substituting R&D collaborations with M&A, emphasizing the close connection between R&D collaborations and M&A activities.

1 Introduction

Inter-firm collaborations take a large variety of forms, such as co-marketing, co-production, or joint product development. They have been shown to have a key role in the diffusion of

*National Taiwan University. Email: cshsieh@ntu.edu.tw

†CEPR, ETH-KOF, Tinbergen Institute and Department of Spatial Economics, Vrije Universiteit Amsterdam. Email: m.d.konig@vu.nl

‡Tuck School of Business, Dartmouth College and NBER. Email: Gordon.M.Phillips@tuck.dartmouth.edu

§Department of Spatial Economics, Vrije Universiteit Amsterdam. Email: j.m.rauch@vu.nl.

knowledge and technology and have attracted attention from economists, policy researchers, and strategic analysts. Understanding the role of collaboration on firm performance is a critical challenge that has been complicated by the lack of comprehensive data on existing collaborations.

To solve this problem, we propose an NLP text-mining algorithm for R&D alliance extraction. The algorithm can be applied to news articles (e.g., those mentioning specific firms or industries of interest) and text data from firm websites. It automatically detects mentions of collaborative agreements and extracts the names of the participants. It also classifies the agreement as either a Joint Venture or Strategic Alliance, detects whether the agreement is completed/signed or just planned, and flags the agreement according to purpose, e.g. manufacturing, marketing, licensing, and/or R&D.

To further capitalize on the comprehensiveness of our extracted data, we develop a structural R&D network formation model in which firms experience positive externalities from technology spillovers in R&D collaborations and negative externalities from market competition. Our approach uniquely incorporates M&As into the model, capturing the joint endogeneity of both R&D collaborations and M&A activities - a novel feature that has not been explored in the existing literature.

We estimate the structural model with the above-mentioned R&D collaboration data from news articles, combined with the text-based network industry classification (TNIC) developed by Hoberg and Phillips (2016), and information on M&A activities from the Thomson SDC database. Our estimates show that the technology spillover effect is positive and significant

across all sample periods from 2000 to 2020 and all specifications but exhibits a cyclical pattern over time. We further observe an upward bias correction for the technology spillover effect after controlling for R&D network formation and the endogeneity of M&As, likely due to unobserved firm-specific effects influencing jointly output, R&D collaborations and M&As. Furthermore, product similarity, past collaborations, or shared partners decrease R&D collaboration costs.

We find that market concentration increases M&A costs, likely due to antitrust regulations, while technological similarity and a higher patent stock of the acquirer reduce M&A costs. Similarly, past R&D collaborations and shared R&D collaboration partners further decrease M&A costs. The significance of the cost-reducing effects of prior R&D collaborations on M&As, and the bias in the estimated spillover effects, emphasize the importance of jointly controlling for R&D collaborations and M&As when estimating technology spillover effects. Moreover, our results suggest that firms substitute R&D collaborations with M&As, emphasizing the close and previously underexplored connection between R&D collaborations and M&A activities and their impact on technology spillover dynamics. We are extending the model to allow for mergers to also benefit firms by reducing duplicative R&D costs and by allowing firms to reduce ex post holdup costs as in Williamson (1998) through integrating.

Related literature. An extensive literature has documented that knowledge spillovers are important for firm performance and productivity (Audretsch and Belitski, 2020; Ugur et al., 2020). This is particularly relevant for spillovers created through R&D collaborations (Ahuja, 2000; Powell et al., 1999; Gulati, 2007; Baum et al., 2010; Rosenkopf and Padula, 2008). For example, Belderbos et al. (2004) provide evidence for a positive effect on firm performance

(measured by sales growth or productivity improvements) through R&D collaborations with various types of partners (such as competitors, suppliers, customers, and universities and research institutes). Kumar et al. (2022) find that 11% of the variance in firm profitability is explained by the network of a firm's alliances. Gulati et al. (2012) study the macro dynamics of R&D networks and document an evolutionary pattern where an increase in the small-worldliness (high clustering and small average path length between nodes) of the network is followed by its later decline. This indicates the importance of accounting for time-varying coefficients for measuring the positive and negative externalities in R&D networks as we do here. Complementarily, there exists a large literature studying the formation of networks such as Powell et al. (2005), Liben-Nowell and Kleinberg (2003), Goyal and Joshi (2003); Goyal and Moraga-Gonzalez (2001), Fafchamps et al. (2010) and Petrakis and Tsakas (2018). In this paper we combine the above mentioned separate strands of the literature by analyzing the two-way influence of the formation of R&D collaborations on firm performance and the effect of firm performance on the formation of R&D collaborations (cf. Koza and Lewin, 1998), and how these have changed over time with the evolution of technologies and markets (cf. Lucking et al., 2019).

Innovation dynamics have significantly changed in recent years (Bloom et al., 2020), raising the question of whether technology spillovers have experienced similar shifts. Lucking et al. (2019) study US firms over three decades and find no evidence of changes in spillovers, but their analysis is based on a restricted sample of publicly listed companies using patent data, and does not focus on spillovers within R&D collaborations as we do. Previous studies,

such as Bloom et al. (2013) and Lucking et al. (2019), measure spillovers through patent portfolio similarities, which may be less precise than R&D collaborations that are typically formed to facilitate technology spillovers. While König et al. (2019) and Hsieh et al. (2022, 2024) analyze R&D spillovers, they do not estimate temporal variation or coevolution between R&D investment and alliance formation. In contrast, we analyze a novel and comprehensive R&D collaboration dataset, allowing for more accurate measurement of spillover variations and the intensity of competition between R&D firms, as well as changes in competition over time (Hoberg and Phillips, 2016, 2024).

Moreover, an extensive body of literature examined the welfare effects of M&As when accounting for technology spillovers. Grossman and Hart (1986) argue that when a firm acquires its supplier, removing the supplier's manager's residual rights of control, the manager's incentives may be distorted to the point where common ownership becomes detrimental to firm performance. In contrast, López and Vives (2019) suggest that overlapping ownership can enhance welfare and consumer surplus, provided the technology spillovers are sufficiently large. Cunningham et al. (2021) highlights the risk that incumbent firms may acquire innovative targets with the sole intention of discontinuing the target's innovation projects, thus preempting future competition. Meanwhile, Antón et al. (2024) propose that the impact of common ownership on innovation is nuanced: it weakens when product market overlap is high but strengthens when technological proximity increases. We contribute to this literature by jointly modeling technology spillovers in R&D collaborations and M&A formation, and find that M&A deals lead to lower firm output. In comparison, the formation of R&D alliances

results in higher firm output.

The paper is organized as follows. The R&D and M&A data are described in Section 2. The structural model is introduced in Section 3. Section 4 presents the results from estimating the model and Section 5 discusses various counterfactuals and discusses policy implications. Finally, Section 6 concludes. A more detailed description on how the data set has been created (and how the text mining algorithms have been implemented) can be found in the Appendix.

2 R&D Collaboration Data with M&As

Our network panel dataset of R&D collaborations is generated by natural language processing (NLP) text mining a large number of news articles and training our algorithm with an LLM model to extract R&D alliances. We give a short overview in the following of how the algorithm was trained and the collection of news articles that it was applied to. We then complement the R&D alliance data with information on M&A activities from the Thomson SDC Platinum database and information on firm characteristics and performance measures from the S&P Compustat and Bureau van Dijk's Orbis databases. A more detailed explanation and discussion of how our text mining algorithm and how the data has been extracted and our LLM model has been trained can be found in Appendices A and B.[1]

[1]Supplementary information incl. all codes for reducibility of the results presented in this paper are available at `https://github.com/jakob-ra/financial_news`.

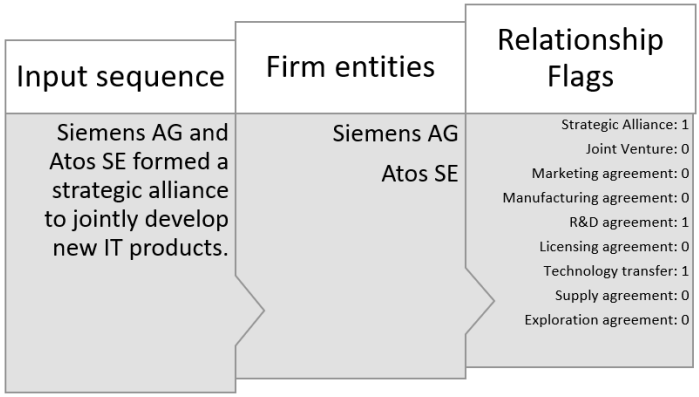


Figure 1. Joint entity and relationship extraction.

2.1 Automated Extraction of Collaborative Agreements

To generate our data set, we trained a text mining algorithm to extract mentions of collaborative agreements between two or more firms. An example is given in Figure 1. Given a news article with an alliance announcement, we extract the names of the involved firms and flag the type of agreement made. This is done via the pre-trained language model LUKE[2] (Yamada et al., 2020), which was fine-tuned to our joint entity and relation extraction task using manually annotated data from the Refinitiv (formerly Thomson Financial) SDC Platinum Joint Venture/Strategic Alliance database. The training data includes around 186,000 positive examples - that is, news articles containing an alliance announcement between firms - which we augment with the same number of negative examples - mentions of two or more firms that *do not* contain a relationship that we are interested in. The trained algorithm reaches an F1-score of 83% on the extraction of R&D collaborations.

[2]LUKE improves on the more widely used RoBERTa model (Liu et al., 2019) by adding entity embeddings and entity-aware self-attention, which makes it a good choice for relation extraction.

2.2 R&D Collaborations from the LexisNexis News Database

We applied the text mining algorithm to articles from LexisNexis data lab, which provides millions of business news articles per day, going back to 1980. We select 1.5 million candidate articles by filtering articles on a list of keywords. In particular, to obtain articles that could contain alliance announcements, we applied a set of filters to LexisNexis' collection of news articles. We selected English language articles tagged as business news from news wires and press releases, newspapers, web-based publications, or industry trade press. Additional information and details regarding the text mining algorithm and the news data can be found in Appendix A.

Our resulting R&D alliance database is much more comprehensive than other currently available databases. For R&D collaborations, SDC Platinum, the de facto only comprehensive data set, contains only around 12,000 such relations, while our data set contains around 52,000 and is less biased towards North American firms. We complement the network data with information about firms' balance sheet information, locations, industries and patent activities. Appendix B provides further details including various descriptive statistics of the R&D collaboration network. An illustration can be seen in Figure 2.

2.3 Text-Based Industry Classification and Accounting Data

We identify the competition matrix $\mathbf{B} = (b_{ij,t})_{1 \leq i,j \leq n}$ using the matrix of Text-based Network Industry Classifications (TNIC) constructed by Hoberg and Phillips (2016) from 10k filings of

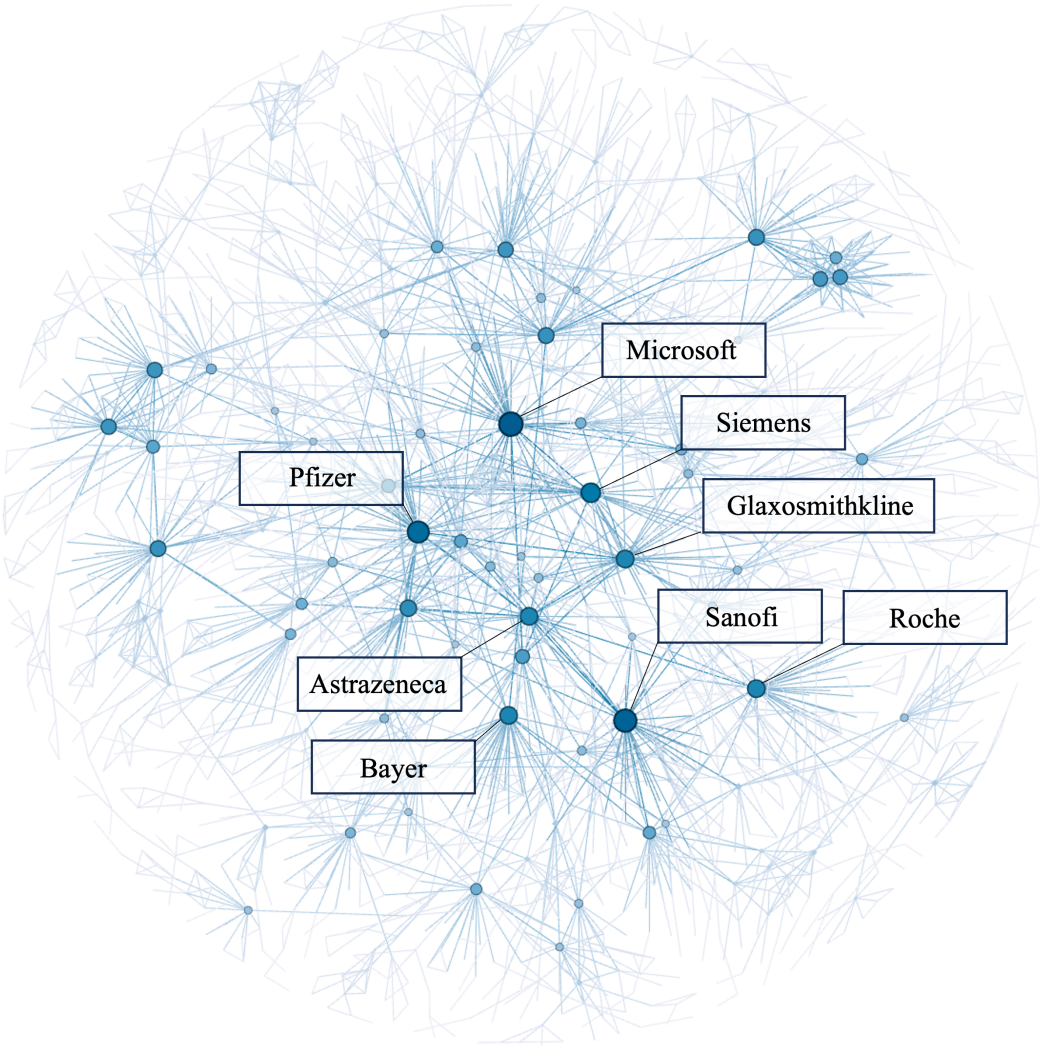


Figure 2. The largest connected component in the network of R&D collaborations in the year 2020. The color of a node and its size scale with the number of R&D collaborations, with the names of the 8 largest firms mentioned in the graph. The number of links is $3,307$, and the number of firms in the largest connected component is $1,826$.

publicly listed firms.[3] Since the bottom half of the TNIC pairwise similarity scores are small (less than 0.05), we define $b_{ij,t} = 1$ if the corresponding TNIC similarity score is higher than the median, and $b_{ij,t} = 0$ otherwise. Firms' sales and R&D expenditures were obtained from the Compustat North America and Global Fundamentals databases as well as Bureau van Dijk's Orbis database (Kalemlİ-Özcan et al., 2024). We compute a firm's output by dividing sales with the producer price index from the Bureau of Labor Statistics (at the NAICS 6-digit level where available and aggregated at the 4-digit level in the remaining cases). To measure a firm's productivity, we use the one-period lag of the firm's log R&D capital stock. The R&D capital stock is computed using a perpetual inventory method with 15% depreciation rate (Bloom et al., 2013).[4]

2.4 Sample Selection and Summary Statistics

To study how technology spillovers change over time, we divide the years from 2001 to 2020 into four segments, each spanning five years. As a result, we will estimate the proposed model separately for four periods: 2001-2005, 2006-2010, 2011-2015, and 2016-2020. To construct the sample for estimation, we employ the following sample selection procedure: We begin with a total of 34,102 U.S. firms in the Compustat data. Then, for the samples in each period, we drop firms that contain missing values on either output or R&D stock.[5] Output has been

[3]See also: `https://hobergphillips.tuck.dartmouth.edu/`.

[4]Following Hall et al. (2005) and Bloom et al. (2013), R&D is used to create R&D capital stocks calculated using a perpetual inventory method with a 15% depreciation rate where the R&D stock in year t is $G_t = R_t + (1 - \delta)G_{t-1}$, and R_t is the R&D flow expenditure in year t with $\delta = 0.15$.

[5]To mitigate the loss of firm observations due to missing values, we use linear interpolation to fill in missing values at certain years for the same firm.

computed using the Producer Price Index (PPI) of the US Bureau of Labor Statistics[6] matched to the year and NAICS code reported in the Compsutat sales item (in USD). Basic summary statistics can be found in Table 1. The M&A deals considered in our sample are summarized in Table 3.

Table 1. Summary statistics (Compustat sample).

	Output			Log R&D Stock			R&D Collab.		# of Firms
	Mean	s.d.	Max	Mean	s.d.	Max	Mean	Max	
2001-2005	6.43	35.53	934.08	2.23	2.37	9.85	0.10	8.00	1533
2006-2010	8.10	34.49	776.09	2.37	2.52	9.62	0.07	9.00	1394
2011-2015	10.86	44.01	999.73	2.51	2.63	9.44	0.04	4.00	1192
2016-2020	11.19	51.75	1119.10	2.47	2.65	9.55	0.04	4.00	1215

Notes: Output is computed using the Producer Price Index (PPI) of the US Bureau of Labor Statistics matched to the year and the 6-digit level NAICS code reported in the Compsutat sales item (in USD), and is measured in millions. R&D stock is measured in million USD.

Table 2. Number of M&A deals.

	Year 1	Year 2	Year 3	Year 4	Year 5
2001-2005	33	15	21	16	21
2006-2010	24	29	24	19	23
2011-2015	20	18	15	15	21
2016-2020	19	13	13	11	12

Table 3. Number of M&A deals.

Period	Year 1	Year 2	Year 3	Year 4	Year 5
2001-2005	33	15	21	16	21
2006-2010	24	29	24	19	23
2011-2015	20	18	15	15	21
2016-2020	19	13	13	11	12

[6]See: `https://www.bls.gov/ppi/`.

3 Model

In Section 3.1 we introduce the static environment of the firm. In Section 3.2 we generalize this to a dynamic panel context modeling the formation of R&D collaborations. In Section 3.3 we extend this framework to account for the formation of M&As.

3.1 Profits from R&D Collaborations

Consider a set $\mathcal{N} = \{1, \ldots, n\}$ of firms and a collaboration network $G \in \mathcal{G}^n$, where $\mathcal{G}^n$ denotes the set of all graphs/networks with n nodes. Let the firms' output levels be given by $\mathbf{y} = (y_1, \ldots, y_n)^\top \in \mathcal{Y}^n$. Firm $i \in \mathcal{N}$ sets its output $y_i \in \mathcal{Y}$ and earns a profit $\pi_i \colon \mathcal{Y}^n \times \mathcal{G}^n \to \mathbb{R}$ given by the following linear-quadratic function:[7]

$$\pi_i = \xi \left(\eta_i y_i - \frac{1}{2} y_i^2 - \lambda y_i \sum_{j \neq i}^{n} b_{ij} y_j + \rho y_i \sum_{j=1}^{n} a_{ij} y_j \right) - \sum_{j=1}^{n} a_{ij} \zeta_{ij}. \tag{1}$$

The indicator variables $a_{ij} \in \{0, 1\}$ in Equation (1) indicate whether firms i and j are collaborating (or not), and can be represented by the symmetric adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ (with a zero diagonal, $a_{ii} = 0$). Equation (1) is similar to the linear-quadratic profit/payoff function analyzed in König et al. (2019), Hsieh et al. (2024) and Ballester et al. (2006). This function is characterized by an own concavity term ($\eta_i y_i - \frac{1}{2} y_i^2$) reflecting decreasing returns in expanding a firm's output level, with η_i capturing firm heterogeneity in productivity, a

[7] A more detailed micro-foundation for the profit function in Equation (1) derived from a Cournot or, alternatively, a Bertrand competition model with cost-reducing R&D collaborations building on the seminal model by Goyal and Moraga-Gonzalez (2001) can be found in Appendix C.1. Moreover, note that the Cournot competition model can also be interpreted as a model of quantity pre-commitment followed by price competition (cf. Kreps and Scheinkman, 1983).

global substitutability term ($\lambda y_i \sum_{j \neq i}^{n} b_{ij} y_j$) reflecting market stealing effects from competition, where $b_{ij} \in \{0, 1\}$ is an indicator variable for whether firms i and j are competitors (or not) which can be represented by the symmetric competition matrix $\mathbf{B} = (b_{ij})_{1 \leq i,j \leq n}$, and a local complementarity term ($\rho y_i \sum_{j=1}^{n} a_{ij} y_j$) reflecting technology spillover effects from R&D collaborations,[8] capturing the fact that R&D is correlated with firm size (Cohen and Klepper, 1996a,b). ξ can be viewed as a weighting parameter to balance the measurement (scale) of output (y) and linking cost (ζ). To ensure $\xi > 0$, we reparameterize it by $\xi = \exp(\delta)$.

The first order condition of Equation (1) with respect to output y_i is given by

$$\frac{\partial \pi_i}{\partial y_i} = \xi \left(\eta_i - y_i - \lambda \sum_{j \neq i}^{n} b_{ij} y_j + \rho \sum_{j=1}^{n} a_{ij} y_j \right) = 0, \tag{2}$$

from which we obtain the best response function

$$y_i = \underbrace{\eta_i}_{\substack{\text{productivity} \\ \text{effect}}} - \underbrace{\lambda \sum_{j \neq i}^{n} b_{ij} y_j}_{\substack{\text{competition} \\ \text{effect}}} + \underbrace{\rho \sum_{j=1}^{n} a_{ij} y_j}_{\substack{\text{spillover} \\ \text{effect}}}. \tag{3}$$

A firm's output (measured by y_i) is thus increasing with a firm's productivity (η_i), decreasing due to a market competition effect ($\lambda \sum_{j \neq i}^{n} b_{ij} y_j$), and increasing due to an R&D spillover effect ($\rho \sum_{j=1}^{n} a_{ij} y_j$). Inserting the best response function from Equation (3) into the profit

[8] Such R&D collaborations often involve competing firms. An illustrative example is the strategic alliance between *Pfizer* and *Bayer*, both operating in the pharmaceutical sector (with primary standard industry classification code 2834) developing treatments for obesity, type 2 diabetes and other related disorders.

function in Equation (1) gives

$$\pi_i = \frac{\xi}{2} y_i^2 - \sum_{j=1}^{n} a_{ij}\zeta_{ij}. \tag{4}$$

Hence, firms' profits (π_i) net of linking costs are an increasing (quadratic) function of output (y_i). Moreover, from Equation (3) we find that

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\eta} - \lambda \mathbf{B}\mathbf{y} + \rho \mathbf{A}\mathbf{y} \\ &= (\mathbf{I}_n + \lambda \mathbf{B} - \rho \mathbf{A})^{-1} \boldsymbol{\eta} \\ &= \boldsymbol{\eta} - \lambda \mathbf{B}\boldsymbol{\eta} + \rho \mathbf{A}\boldsymbol{\eta} + O(\lambda\rho), \end{aligned} \tag{5}$$

where we assume that the matrix $\mathbf{I}_n + \lambda \mathbf{B} - \rho \mathbf{A}$ is invertible (cf. Horn and Johnson, 1990). For the output of firm i this implies that

$$y_i = \underbrace{\eta_i}_{\substack{\text{own} \\ \text{productivity} \\ \text{effect}}} - \underbrace{\lambda \sum_{j \neq i}^{n} b_{ij}\eta_j}_{\substack{\text{direct} \\ \text{competition} \\ \text{effect}}} + \underbrace{\rho \sum_{j=1}^{n} a_{ij}\eta_j}_{\substack{\text{direct} \\ \text{spillover} \\ \text{effect}}} + \underbrace{O(\lambda\rho).}_{\substack{\text{higher order} \\ \text{effects}}} \tag{6}$$

This shows that firm output (y_i) is, up to first order effects, increasing with firm productivity (η_i), decreasing with competitor's productivities ($\sum_{j=1}^{n} b_{ij}\eta_j$), and increasing with collaborators' productivities ($\sum_{j=1}^{n} a_{ij}\eta_j$). This illustrates the positive and negative externalities affecting firm performance (measured by firm size) in R&D networks (cf. e.g. Bloom et al.,

2013; König et al., 2019).

Social welfare, $W(\mathbf{y})$, is given by the sum of consumer surplus, $U(\mathbf{y})$, and producer surplus, $\Pi(\mathbf{y})$. Consumer surplus is given by $U(\mathbf{y}) = \frac{\psi}{2}\sum_{i=1}^{n} y_i^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} b_{ij}y_iy_j$ with $\psi > \lambda > 0$ (Singh and Vives, 1984; Vives, 1999).[9] Producer surplus is given by firms' aggregate profits $\Pi(\mathbf{y}) = \sum_{i=1}^{n} \pi_i(\mathbf{y})$, with profits given by Eq. (4). The Herfindahl-Hirschman industry concentration index is defined as $H_k(\mathbf{y}) = \sum_{i\in\mathcal{M}_k} s_i^2(\mathbf{y})$, where the market share of firm i in market $\mathcal{M}_k$ is given by $s_i(\mathbf{y}) = \frac{y_i}{\sum_{j\in\mathcal{M}_k}^{n} y_j}$ (cf. e.g. Tirole, 1988; Hirschman, 1964).

3.2 R&D Network Formation

We assume that, in each time period $t = 1, \ldots, T$, firms undergo a two-stage process (Hsieh et al., 2020):

- In the first stage, each pair of firms independently determines whether to

  (i) establish an R&D collaboration,

  (ii) or to remain idle,

  depending on which yields higher expected profits in the second stage, irrespective of the choices made by other pairs.

- In the second stage, each firm takes the R&D collaboration network formed in the first stage as given and chooses output to maximize profits by interacting with other firms

[9] The assumption of $\psi > \lambda > 0$ ensures concavity of the consumer's utility function. Moreover, the assumption of $\lambda > 0$ ensures that products are (imperfect) substitutes. The higher λ, the more alike products are and the more intense is competition. See Supplementary Appendix C.1 for further details and explanation.

non-cooperatively.

Between the two stages, firms have complete information, i.e., they know the productivities (characteristics) of all other firms and which R&D collaboration links are formed in the first stage. We also assume that firms are not myopic in the sense that they will take the potential benefit of interactions in the second stage into account when forming links in the first stage.

First stage. In the following we consider the first stage of the process where the network G_t is formed. We take into account various network externalities that affect the formation of R&D collaborations in the pairwise linking costs, ζ_{ij} in Equation (1), between firms (cf. Powell et al., 2005; Gulati, 2007; Gulati et al., 2012). For example, Liben-Nowell and Kleinberg (2003) show that two nodes and are more likely to form a link if they have a large overlap in their characteristics. We incorporate this in terms of observable characteristics (geographical, spatial and technological) and unobservable characteristics ("latent variables"). Moreover, Fafchamps et al. (2010) show that a new collaboration emerges faster among two nodes if they are "closer" (in terms of network distance) in the existing network. We incorporate this feature by network dependent variables affecting the formation of collaborations or mergers via a reduction of the linking cost if nodes have a common neighbor in the past. More specifically,

we assume that[10]

$$\zeta_{ij,t} = \underbrace{\gamma_0 + \mathbf{c}_{ij,t}\gamma_1 + \gamma_2|y_{i,t-1} - y_{j,t-1}| + \gamma_3|y_{i,t-1} - y_{j,t-1}|^2}_{\text{R\&D network independent}} \underbrace{+\gamma_4 a_{ij,t-1} + \gamma_5 \sum_{k=1}^{n} a_{ik,t-1}a_{jk,t-1}}_{\text{R\&D network dependent}} . \tag{7}$$

We further distinguish between network-dependent and network-independent explanatory variables to examine the explanatory power of network variables over and above knowledge of firm characteristics only (Ductor et al., 2014). More specifically, the r-dimensional (row) vector of dyad-specific variables, $\mathbf{c}_{ij,t}$, in Equation (7) captures the similarity between firms i and j regarding observable characteristics such as sector, location, productivity, etc., that might affect the collaboration costs (see, e.g., Lychagin et al., 2016). We include $|y_{i,t-1} - y_{j,t-1}|$ and its square to capture the potential nonlinear homophily (heterophily) effect from past R&D efforts. The term $a_{ij,t-1}$ in Equation (7) captures the effect of repeated collaborations of firms i and j (Fafchamps et al., 2010; Goerzen, 2007). Further, the term $\sum_{k=1}^{n} a_{jk,t-1}a_{ik,t-1}$ captures the effect of common collaborators of firms i and j in the past (clustering or cyclic triangle effect),[11] and γ_0 is a constant.

The profit of firm i from forming an R&D collaboration with firm j at time t, given the

[10]Note that it is straightforward to introduce higher order network dependencies in Equation (7) as in Fafchamps et al. (2010).

[11]See also Hanaki et al. (2010), Mele (2017) and Fafchamps et al. (2010).

network G_{t-1} at time $t-1$, is given by

$$\Pi_{ijt}^{\mathrm{R\&D}}(G_{t-1}) = \frac{\delta}{2} y_i(G_{t-1}, a_{ij,t}=1)^2 - \zeta_{ij,t},$$

where, using Equation (5), we have denoted $\mathbf{y}(G) = (\mathbf{I}_n + \lambda\mathbf{B} - \rho\mathbf{A})^{-1}\boldsymbol{\eta}$, and the collaboration cost $\zeta_{ij,t}$ is given by Equation (7). The profit for firm i from remaining idle is

$$\Pi_{it}^{\mathrm{ID}}(G_{t-1}) = \frac{\delta}{2} y_i(G_{t-1})^2.$$

The probability that i wants to collaborate with j is given by

$$\begin{aligned}\mathbb{P}(i \xrightarrow{co} j | G_{t-1}, \mathbf{y}_{t-1}, \gamma) &= \mathbb{P}\left(\Pi_{ijt}^{\mathrm{R\&D}}(G_{t-1}) + \varepsilon_{ij,t}^{\mathrm{R\&D}} > \Pi_{it}^{\mathrm{ID}}(G_{t-1}) + \varepsilon_{i,t}^{\mathrm{ID}}\right) \\ &= \frac{e^{\Pi_{ijt}^{\mathrm{R\&D}}(G_{t-1})}}{e^{\Pi_{ijt}^{\mathrm{R\&D}}(G_{t-1})} + e^{\Pi_{it}^{\mathrm{ID}}(G_{t-1})}}. \end{aligned} \tag{8}$$

We assume that a collaboration between i and j requires mutual consent (cf. e.g. Hanaki et al., 2010). The conditional likelihood taking into account the collaboration decisions of all pairs of firms at time t is then given by

$$\begin{aligned}\mathbb{P}(G_t | G_{t-1}, \mathbf{y}_{t-1}, \gamma) = \prod_{i=1}^{n} \prod_{j=i+1}^{n} &[\mathbb{P}(i \xrightarrow{co} j | G_{t-1}, \mathbf{y}_{t-1}, \gamma)\mathbb{P}(j \xrightarrow{co} i | G_{t-1}, \mathbf{y}_{t-1}, \gamma)]^{a_{ij,t}} \\ &\times [1 - \mathbb{P}(i \xrightarrow{co} j | G_{t-1}, \mathbf{y}_{t-1}, \gamma)\mathbb{P}(j \xrightarrow{co} i | G_{t-1}, \mathbf{y}_{t-1}, \gamma)]^{1-a_{ij,t}}, \end{aligned} \tag{9}$$

with the probability that i wants to collaborate with j given by Equation (8). An R&D

collaboration is observed whenever it is beneficial to both parties involved, and not if it is detrimental to at least one of them. This corresponds to the "double-hurdle" model discussed, for example, in De Paula (2020) and the "partially observable" model introduced by Poirier (1980). Comola (2012) analyzes partial observability in network formation games. Partial observability occurs when a positive outcome for one response variable is only observed if the other response variable is also positive. In our context, when $a_{ij,t} = 1$ we know that i wants to collaborate with j and j wants to collaborate with i, which happens when $\Pi_{ijt}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{ijt}^{\text{R\&D}} > \Pi_{it}^{\text{ID}}(G_{t-1}) + \varepsilon_{i,t}^{\text{ID}}$ and $\Pi_{jit}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{jit}^{\text{R\&D}} > \Pi_{jt}^{\text{ID}}(G_{t-1}) + \varepsilon_{j,t}^{\text{ID}}$. Viceversa, when $a_{ij,t} = 0$ we cannot distinguish between the three following cases: (i) i does not want to collaborate with j but j wants to collaborate with i, which happens when $\Pi_{ijt}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{ijt}^{\text{R\&D}} < \Pi_{it}^{\text{ID}}(G_{t-1}) + \varepsilon_{i,t}^{\text{ID}}$ and $\Pi_{jit}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{jit}^{\text{R\&D}} > \Pi_{jt}^{\text{ID}}(G_{t-1}) + \varepsilon_{j,t}^{\text{ID}}$. (ii) i wants to collaborate with j but j does not want to collaborate with i, which happens when $\Pi_{ijt}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{ijt}^{\text{R\&D}} > \Pi_{it}^{\text{ID}}(G_{t-1}) + \varepsilon_{it}^{\text{ID}}$ and $\Pi_{jit}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{jit}^{\text{R\&D}} < \Pi_{jt}^{\text{ID}}(G_{t-1}) + \varepsilon_{jt}^{\text{ID}}$. (iii) Both, i and j do not want to collaborate with each other, which happens when $\Pi_{ijt}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{ijt}^{\text{R\&D}} < \Pi_{it}^{\text{ID}}(G_{t-1}) + \varepsilon_{it}^{\text{ID}}$ and $\Pi_{jit}^{\text{R\&D}}(G_{t-1}) + \varepsilon_{jit}^{\text{R\&D}} < \Pi_{jt}^{\text{ID}}(G_{t-1}) + \varepsilon_{jt}^{\text{ID}}$.

Second stage. We next consider the second stage of the process where firms choose their output levels. After the network G_t is formed, each firm i takes G_t, G_{t-1} and $\mathbf{y}_{t-1}$ as given, and chooses $y_{i,t}$ to maximize profits in Equation (1). Recall that profits in Equation (1) depend on firms' productivities ($\boldsymbol{\eta}_t$). Here we assume that a firm i's productivity, η_{it}, evolves according

to (cf. Peters et al., 2017):[12]

$$\eta_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \mu_i + \tau_t + v_{i,t}. \tag{10}$$

In Equation (10), $\mathbf{x}_{i,t}$ is a vector of observable characteristics, such as lagged R&D capital stock of firm i, μ_i is time-invariant firm-specific effect, τ_t is time-specific effect, and v_{it} is an i.i.d. (across firms and time) error term.

From Equation (3), it then follows that a firm i's best response function is given by

$$y_{i,t} = \rho \sum_{j=1}^{n} a_{ij,t} y_{j,t} - \lambda \sum_{j=1}^{n} b_{ij,t} y_{j,t} + \mathbf{x}_{it}\boldsymbol{\beta} + \mu_i + \tau_t + v_{it}.$$

In vector-matrix notation, this can be written as

$$\mathbf{y}_t = \rho \mathbf{A}_t \mathbf{y}_t - \lambda \mathbf{B}_t \mathbf{y}_t + \mathbf{x}_t\boldsymbol{\beta} + \mu + \tau_t \ell + \mathbf{v}_t, \tag{11}$$

where ℓ is a vector of ones. In the following, we denote by

$$\mathbf{S}_t = \mathbf{I}_n - \rho \mathbf{A}_t + \lambda \mathbf{B}_t.$$

If the inverse $\mathbf{S}_t^{-1}$ exists (see e.g. König et al., 2019, for sufficient conditions), we can write the reduced form as

$$\mathbf{y}_t = \mathbf{S}_t^{-1}(\mathbf{x}_t\boldsymbol{\beta} + \mu + \tau_t \ell + \mathbf{v}_t). \tag{12}$$

[12]See also the stochastic productivity growth models discussed in Olley and Pakes (1996), Levinsohn and Petrin (2003), and the R&D capital stock introduced in Hall et al. (2005, 2010) that is computed using a perpetual inventory method based on the firms' past R&D expenditures discounted with a given depreciation rate (see also Bloom et al., 2013; König et al., 2019).

Assume that $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I}_n)$. Then, $\mathbf{y}_t$ is an affine transformation of a multivariate normal distribution $\mathbf{y}_t \sim \mathcal{N}(\mathbf{S}_t^{-1}(\mathbf{x}_t\boldsymbol{\beta} + \mu + \tau_t\ell), \sigma_v^2\mathbf{S}_t^{-2})$. The conditional likelihood of the effort levels $\mathbf{y}_t$ at time t is thus given by

$$\mathbb{P}(\mathbf{y}_t|G_{t-1}, \mathbf{x}_t, \theta, \boldsymbol{\beta}, \mu, \tau_t, \sigma_v^2) = (2\pi\sigma_v^2)^{-\frac{n}{2}}|\mathbf{S}_t| \exp\left(-\frac{1}{2\sigma_v^2}\mathbf{v}_t^\top \mathbf{v}_t\right),$$

where from Equation (12) we have that $\mathbf{v}_t = \mathbf{S}_t\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\beta} - \mu - \tau_t\ell$, and $\theta = (\rho, \lambda)$.

Joint likelihood. We next construct the full likelihood of the model taking into account both, the formation of the network as well as the endogenous output choices. Assume the initial period $\mathbf{y}_0$ and G_0 are exogenously given. We denote unknown parameters by $\Theta = (\theta, \beta, \delta, \gamma, \nu, \sigma_v^2)$, the joint likelihood function of $\{G_t\}$ and $\{\mathbf{y}_t\}$ can be decomposed as follows:

$$\begin{aligned}\mathbb{P}(\{\mathbf{y}_t\}, \{G_t\}|\{\mathbf{y}_{t-1}\}, \{G_{t-1}\}, \Theta) &= \prod_{t=1}^{T} \mathbb{P}(\mathbf{y}_t|G_t, \Theta)\mathbb{P}(G_t|G_{t-1}, \mathbf{y}_{t-1}, \Theta) \\ &= \prod_{t=1}^{T} |\mathbf{S}_t(G_t)| f(\mathbf{v}_t|G_t, \Theta)\mathbb{P}(G_t|G_{t-1}, \mathbf{y}_{t-1}, \Theta)\end{aligned} \tag{13}$$

where $\mathbf{S}_t(G_t) = \mathbf{I}_n - \rho\mathbf{A}_t + \lambda\mathbf{B}_t$ and $\mathbf{v}_t = \mathbf{S}_t\mathbf{y}_t - \mathbf{x}_t\boldsymbol{\beta} - \mu - \tau_t\ell$, and $\mathbb{P}(G_t|\mathbf{v}_t, G_{t-1}, \mathbf{y}_{t-1}, \Theta)$ is given by Equation (9). Finally, note that we can test for time-varying spillover effects/externalities by incorporating time-varying coefficients, $\{\rho_t\}$ and $\{\lambda_t\}$, in the likelihood function of Equation (13) (cf. Gulati et al., 2012; Lucking et al., 2019).

3.3 R&D Network and M&A Formation

We extend the previous model by introducing M&As in the first stage of the model. In particular, similar to the previous section, we assume that in each time period $t = 1, \ldots, T$ firms undergo a two-stage process (Phillips and Zhdanov, 2013; Hsieh et al., 2020):

- In the first stage, each pair of firms independently determines whether to form

  (i) an R&D collaboration,

  (ii) a merger or

  (iii) to remain idle,

  depending on which yields higher expected profits in the second stage, irrespective of the choices made by other pairs.

- In the second stage, each firm takes the R&D collaboration network and mergers formed in the first stage as given and chooses output to maximize profits by interacting with other firms non-cooperatively.

First stage. Firm profits from the formation of R&D collaborations are identical to the previous section. For the profits from M&As we assume that a merger between firms i and j creates a single firm k, i.e., a single node in the network (Phillips and Zhdanov, 2013; Bimpikis et al., 2019). Firm k inherits all the collaborations of firms i and j in the case of a merger.

We assume that the cost for firm i to merge firm j is given by

$$\nu_{ij,t} = \kappa_0 + \mathbf{c}_{ij,t}\kappa_1 + \kappa_2(y_{i,t-1} - y_{j,t-1}) + \kappa_3(y_{i,t-1} - y_{j,t-1})^2 + \kappa_4 a_{ij,t-1} + \kappa_5 \sum_{k=1}^{n} a_{ik,t-1}a_{jk,t-1}. \quad (14)$$

Note that the existence of a collaboration between i and j in the past (indicated by $a_{ij,s} = 1$ for $s = 1, \ldots, t-1$) reduces the cost of a merger in Equation (14). After the merger, we assume

$$\eta_{kt} = \max\{\eta_{i,t-1} + \mathbf{x}_{jt}\boldsymbol{\beta}, \eta_{j,t-1} + \mathbf{x}_{it}\boldsymbol{\beta}\}, \quad (15)$$

where i acquires j and becomes k by incorporating the R&D stock of firm j.[13] Moreover, we assume that firm i proposes a merger to firm j by making a take-it-or-leave-it ("TIOLI") offer. The TIOLI bargaining protocol follows Igami and Uetake (2020) and implies that the acquisition price equals the target firm j's outside option (i.e. staying independent). The profit of firm i from acquiring firm j forming firm k at time t, given the network G_{t-1} at time $t-1$, is given by

$$\Pi_{ijt}^{\mathrm{M\&A}}(G_{t-1}) = \frac{\xi}{2} y_k(G_{t-1}, i \text{ acquires } j)^2 - \frac{\xi}{2} y_j(G_{t-1})^2 - \nu_{ij,t}, \quad (16)$$

where, using Equation (5), we have $\mathbf{y}(G) = (\mathbf{I}_n + \lambda\mathbf{B} - \rho\mathbf{A})^{-1}\boldsymbol{\eta}$, and the acquisition cost $\kappa_{ij,t}$ is given by Equation (14).

[13] Typically acquirers have a higher productivity.

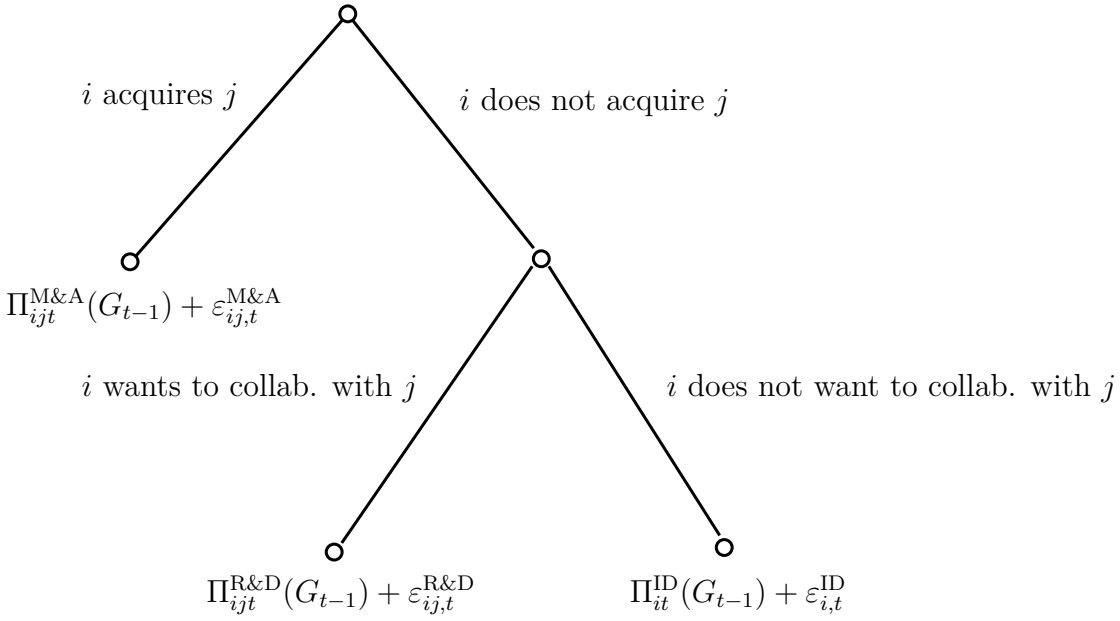


Figure 3. Decision tree of firm i acquiring or collaborating with firm j.

When firm i decides about collaborating, merging or remaining idle, we assume that the corresponding profits $\Pi_{ijt}^{\text{R\&D}}(G_{t-1})$, $\Pi_{ijt}^{\text{M\&A}}(G_{t-1})$ and $\Pi_{it}^{\text{ID}}(G_{t-1})$ experience additive shocks, $\varepsilon_{ij,t}^{\text{R\&D}}$, $\varepsilon_{ij,t}^{\text{M\&A}}$ and $\varepsilon_{i,t}^{\text{ID}}$, which are identically and independently type I extreme value distributed (McFadden, 1974; Train, 2009). These shocks are observed by the firm but not by the econometrician. Given these shocks, firm i decides whether it wants to acquire firm j. If firm i does not want to acquire firm j, it can form an R&D collaboration with firm j. This is illustrated in the decision tree in Figure 3. Conditional on not wanting to acquire firm j (corresponding to the lower level in the decision tree in Figure 3), the probability that i wants to collaborate

with j is given

$$\begin{aligned}\mathbb{P}(i \xrightarrow{co} j|G_{t-1}, \mathbf{y}_{t-1}, \gamma) &= \mathbb{P}\left(\Pi^{\text{R\&D}}_{ijt}(G_{t-1}) + \varepsilon^{\text{R\&D}}_{ij,t} > \Pi^{\text{ID}}_{it}(G_{t-1}) + \varepsilon^{\text{ID}}_{i,t}\right) \\ &= \frac{\exp\left(\Pi^{\text{R\&D}}_{ij,t}(G_{t-1})\right)}{\exp\left(\Pi^{\text{R\&D}}_{ij,t}(G_{t-1})\right) + \exp\left(\Pi^{\text{ID}}_{ij,t}(G_{t-1})\right)}. \end{aligned} \tag{17}$$

Further, the probability that firm i acquires firm j (corresponding to the upper level in the decision tree in Figure 3) is given by

$$\begin{aligned}&\mathbb{P}(i \xrightarrow{ac} j|G_{t-1}, \mathbf{y}_{t-1}, \kappa) \\ &= \mathbb{P}\left(\left\{\Pi^{\text{M\&A}}_{ijt}(G_{t-1}) + \varepsilon^{\text{M\&A}}_{ij,t} > \Pi^{\text{R\&D}}_{ijt}(G_{t-1}) + \varepsilon^{\text{R\&D}}_{ij,t}\right\} \text{and } \left\{\Pi^{\text{M\&A}}_{ijt}(G_{t-1}) + \varepsilon^{\text{M\&A}}_{ij,t} > \Pi^{\text{ID}}_{it}(G_{t-1}) + \varepsilon^{\text{ID}}_{i,t}\right\}\right) \\ &= \frac{\exp\left(\Pi^{\text{M\&A}}_{ij,t}(G_{t-1})\right)}{\exp\left(\Pi^{\text{M\&A}}_{ij,t}(G_{t-1})\right) + \exp\left(\Pi^{\text{R\&D}}_{ij,t}(G_{t-1})\right) + \exp\left(\Pi^{\text{ID}}_{ij,t}(G_{t-1})\right)}. \end{aligned} \tag{18}$$

With Equations (17) and (18) the likelihood of the network G_t, conditional on G_{t-1}, $\mathbf{y}_{t-1}$, $\gamma = (\gamma_0, \gamma_1', \gamma_2, \gamma_3, \gamma_4, \gamma_5)$ and $\kappa = (\kappa_0, \kappa_1', \kappa_2, \kappa_3, \kappa_4, \kappa_5)$, can then be written as follows

$$\begin{aligned}&\mathbb{P}(G_t, M_t|G_{t-1}, \mathbf{y}_{t-1}, \gamma, \kappa) \\ &= \prod_{i=1}^{n}\prod_{j\neq i}^{n} \mathbb{P}(i \xrightarrow{ac} j|G_{t-1}, \mathbf{y}_{t-1}, \kappa)^{I(m_{ij,t}=1)} \Big(1 - \mathbb{P}(i \xrightarrow{ac} j|G_{t-1}, \mathbf{y}_{t-1}, \kappa)\Big)^{I(m_{ij,t}=0)} \\ &\quad \times \prod_{i=1}^{n} \prod_{j=i+1, m_{ij,t}=m_{ji,t}=0}^{n} \Big(\mathbb{P}(i \xrightarrow{co} j|G_{t-1}, \mathbf{y}_{t-1}, \gamma)\mathbb{P}(j \xrightarrow{co} i|G_{t-1}, \mathbf{y}_{t-1}, \gamma)\Big)^{I(a_{ij,t}=1)} \\ &\quad \times \Big(1 - \mathbb{P}(i \xrightarrow{co} j|G_{t-1}, \mathbf{y}_{t-1}, \gamma)\mathbb{P}(j \xrightarrow{co} i|G_{t-1}, \mathbf{y}_{t-1}, \gamma)\Big)^{I(a_{ij,t}=0)} \end{aligned} \tag{19}$$

The second stage is identical to the previous section, and the joint likelihood of $\mathbf{y}_t$, G_t,

and M_t for $t = 1, \cdots, T$, is given by

$$\begin{aligned}\mathbb{P}(\{\mathbf{y}_t\}, \{G_t\}, \{M_t\} | \{\mathbf{y}_{t-1}\}, \{G_{t-1}\}, \Theta) &= \prod_{t=1}^{T} \mathbb{P}(\mathbf{y}_t | G_t, \Theta) \mathbb{P}(G_t, M_t | G_{t-1}, \mathbf{y}_{t-1}, \Theta) \\ &= \prod_{t=1}^{T} |\mathbf{S}_t(G_t)| f(\mathbf{v}_t | G_t, \Theta) \mathbb{P}(G_t, M_t | G_{t-1}, \mathbf{y}_{t-1}, \Theta).\end{aligned} \tag{20}$$

We then apply a Bayesian MCMC approach for estimating the unknown parameter vector $\Theta = (\theta,\ \beta,\ \delta,\ \gamma, \delta, \sigma_v^2)$. After dividing Θ into blocks, we adopt the Metropolis-Hastings-within-Gibbs sampling procedure to simulate draws with the conditional posterior distribution derived in Equation (20).

4 Estimation Results

Based on our theoretical model derived in Section 3.2 we estimate the following equation:

$$y_{i,t} = \rho \sum_{j=1}^{n} a_{ij,t} y_{j,t} - \lambda \sum_{j=1}^{n} b_{ij,t} y_{j,t} + \beta x_{i,t-1} + \mu_i + \tau_t + v_{i,t}, \tag{21}$$

where $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})^\top$ denotes output (measured in a million units), $\mathbf{x}_t = (x_{1t}, \ldots, x_{nt})$ denotes the R&D capital stock, $\mathbf{A}_t = (a_{ij,t})_{1 \le i,j \le n}$ is the adjacency matrix of R&D collaborations, and $\mathbf{B} = (b_{ij,t})_{1 \le i,j \le n}$ is the competition matrix based on TNIC with granularity at the 3-digit SIC code level.

The estimation results for the R&D network formation model outlined in Section 3.2 can

be found in Table 4. Model (1) corresponds to the estimation of Equation (11). Model (2) corresponds to the joint estimation of the output and R&D collaboration decisions in Equation (13). We find that the spillover effect, ρ, is significant with the expected signs across all sample periods and specifications considered. Moreover, comparing the estimates for Models (1) and (2), there is an upward bias correction on the estimate of ρ after controlling for network formation. This bias typically stems from unobserved firm-specific correlated effects that affect both the output of a firm and the R&D collaborations it forms. Comparing the different sample periods we further find a cyclical pattern in the estimated spillover coefficient ρ. Moreover, product or market similarity (TNIC) tends to reduce the R&D collaboration cost. Besides, past R&D collaborations as well as having common R&D collaboration partners in the past both reduce the cost of R&D collaborations.

Next, Table 5 presents the estimation results for the model outlined in Section 3.3 where we control for both, the formation of R&D collaborations and M&As. Model (1) corresponds to the estimation Equation (11). Model (2) corresponds to the joint estimation of the output and R&D collaboration decisions in Equation (20). Similar to the estimates for the spillover coefficient, ρ, in Table 4, we observe a cyclical pattern across the observation periods and an upward bias when failing to control for the endogeneity of R&D collaborations and M&As. Moreover, we find that market concentration (TNIC Similarity$*$TNIC HHI) leads to higher M&A costs, possibly due to barriers from antitrust policy regulations. Further, technological similarity and a higher patent stock of the acquire lead to lower M&A costs. Notably, past R&D collaborations and common R&D collaboration partners in the past reduce the M&A

Table 4. Estimation results for US publicly listed firms in the Compustat sample.

		2001-2005		2006-2010		2011-2015		2016-2020	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Output									
R&D Spillover	(ρ)	0.3569	0.3120	0.1886	0.1841	0.3257	0.3168	0.2860	0.2799
		(0.0064)	(0.0080)	(0.0058)	(0.0052)	(0.0107)	(0.0080)	(0.0095)	(0.0062)
Competition	(λ)	0.0077	0.0083	-0.0004	-0.0005	0.0058	0.0059	0.0061	0.0060
		(0.0018)	(0.0017)	(0.0018)	(0.0018)	(0.0030)	(0.0031)	(0.0031)	(0.0030)
Productivity	(β)	2.7701	2.8332	4.2044	4.2204	5.3203	5.3357	5.7674	5.7893
		(0.0997)	(0.0934)	(0.1121)	(0.1086)	(0.1523)	(0.1516)	(0.2021)	(0.1964)
Error Variance	(σ_v^2)	34.4858	31.3157	49.6113	49.5285	399.2705	397.7822	1108.2000	1109.0559
		(0.8347)	(0.6235)	(1.1706)	(1.1626)	(22.3486)	(23.2114)	(30.1000)	(30.4574)
R&D Collaboration Cost									
Constant	(γ_0)		6.1761		6.8030		6.7468		6.8541
			(0.0970)		(0.1541)		(0.2022)		(0.2233)
TNIC Similarity	(γ_1)		-10.6231		-10.9140		-10.7475		-8.3180
			(1.1403)		(1.0507)		(1.4031)		(1.3134)
Diff. Production	(γ_2)		-0.0150		-0.0256		-0.0083		-0.0097
			(0.0019)		(0.0048)		(0.0028)		(0.0019)
Diff. Production^2	(γ_3)		2.94E-05		9.38E-05		1.13E-05		9.79E-06
			(7.00E-06)		(2.57E-05)		(6.41E-06)		(3.24E-06)
Past R&D Collab.	(γ_4)		-7.9829		-13.4429		-15.3902		-14.1500
			(0.1871)		(1.7307)		(2.7478)		(2.4338)
R&D Triangle Effect	(γ_5)		-1.4348		-2.3288		-4.5298		-2.2428
			(0.6488)		(0.6123)		(0.7163)		(1.1674)
# of Firms	(n)		1533		1394		1192		1215

Notes: Model (1) corresponds to the estimation of Equation (11). Model (2) corresponds to the joint estimation of the output and R&D collaboration decisions in Equation (13). Output is computed using NAICS 6-digit price indices from the BLS. The competition matrix $\mathbf{B}$ is defined based on the TNIC data calibrated to be as granular as the three-digit SIC codes. The R&D collaboration cost is specified as $\zeta_{ij,t} = \gamma_0 + \mathbf{c}_{ij,t}\gamma_1 + \gamma_2|y_{i,t-1} - y_{j,t-1}| + \gamma_3|y_{i,t-1} - y_{j,t-1}|^2 + \gamma_4 a_{ij,t-1} + \gamma_5 \sum_{k=1}^{n} a_{ik,t-1}a_{jk,t-1}$.

cost. The significance and magnitudes of the cost reducing effect of R&D collaborations on M&As, as well as the bias in the coefficient ρ in Model (1), show the importance of jointly controlling for R&D collaboration and M&As in estimating technology spillover effects.

Table 5. Estimation results for US publicly listed firms in the Compustat sample.

		2001-2005		2006-2010		2011-2015		2016-2020	
		(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Output									
R&D Spillover	(ρ)	0.3569	0.3139	0.1886	0.1888	0.3257	0.3087	0.2860	0.2775
		(0.0064)	(0.0050)	(0.0058)	(0.0029)	(0.0107)	(0.0020)	(0.0095)	(0.0046)
Competition	(λ)	0.0077	0.0084	-0.0004	0.0003	0.0058	0.0059	0.0061	0.0078
		(0.0018)	(0.0010)	(0.0018)	(0.0017)	(0.0030)	(0.0029)	(0.0031)	(0.0029)
Productivity	(β)	2.7701	2.9105	4.2044	4.2178	5.3203	5.2635	5.7674	5.6955
		(0.0997)	(0.0534)	(0.1121)	(0.1080)	(0.1523)	(0.0452)	(0.2021)	(0.0540)
Error Variance	(σ_v^2)	34.4858	31.4933	49.6113	49.7531	399.2705	391.6413	1108.2000	1108.3400
		(0.8347)	(0.6219)	(1.1706)	(1.0254)	(22.3486)	(22.5784)	(30.1000)	(30.6060)
R&D Collaboration Cost									
Constant	(γ_0)		6.3625		6.7895		6.7661		6.8744
			(0.1048)		(0.1503)		(0.1673)		(0.2047)
TNIC Similarity	(γ_1)		-4.8884		-8.1889		-4.1039		-4.9340
			(0.5962)		(1.5166)		(1.2886)		(0.9331)
Patent Tech Similarity	(γ_2)		-2.0930		-1.6979		-1.2356		-1.7724
			(0.1909)		(0.3153)		(0.2263)		(0.2929)
Diff. Production	(γ_3)		-0.0161		-0.0168		-0.0070		-0.0083
			(0.0024)		(0.0156)		(0.0025)		(0.0014)
Diff. Production^2	(γ_4)		3.71E-05		3.53E-04		8.78E-06		8.50E-06
			(1.16E-05)		(2.27E-04)		(5.20E-06)		(2.56E-06)
Past R&D Collab.	(γ_5)		-9.7717		-12.5376		-7.1158		-14.1050
			(0.3126)		(0.9090)		(0.2290)		(1.9414)
R&D Triangle Effect	(γ_6)		-0.3811		-1.8251		-1.4537		-1.5866
			(0.2201)		(0.4567)		(0.4304)		(0.8387)
M&A Cost									
Constant	(κ_0)		12.1493		12.6970		13.6491		13.7673
			(0.0648)		(0.1606)		(0.1949)		(0.2636)
TNIC Similarity	(κ_1)		2.3924		1.0196		0.0485		-1.0029
			(0.2425)		(0.5533)		(0.1937)		(0.2924)
TNIC Similarity$*$TNIC HHI	(κ_2)		0.4378		0.6702		1.0300		0.4419
			(0.0541)		(0.2254)		(0.1054)		(0.2264)
Patent Tech Similarity	(κ_3)		-3.9583		-3.8461		-4.2779		-3.7835
			(0.2402)		(0.2117)		(0.2478)		(0.1489)
Patent Stock (Acquirer)	(κ_4)		-0.1013		-0.2085		-0.3071		-0.3720
			(0.0442)		(0.0506)		(0.0418)		(0.0760)
Patent Stock (Target)	(κ_5)		0.1235		0.0471		-0.0686		-0.1251
			(0.0532)		(0.0509)		(0.0421)		(0.0630)
Diff. Production	(κ_6)		-0.0167		-0.0017		-0.0505		-0.0393
			(0.0037)		(0.0019)		(0.0097)		(0.0096)
Diff. Production^2	(κ_7)		3.24E-05		-1.32E-06		3.98E-04		2.92E-04
			(1.14E-05)		(3.27E-06)		(9.51E-05)		(8.95E-05)
Past R&D Collab.	(κ_8)		-4.3513		-4.8520		-2.9136		-3.2576
			(0.0746)		(0.4833)		(0.3299)		(0.1833)
R&D Triangle Effect	(κ_9)		-2.3413		-2.5135		-1.1348		-1.3936
			(0.0599)		(0.3750)		(0.3405)		(0.1609)
# of Firms	(n)	1533		1394		1192		1215	

Notes: Model (1) corresponds to the estimation Equation (11). Model (2) corresponds to the joint estimation of the output and R&D collaboration decisions in Equation (20). Output is computed using NAICS 6-digit price indices from the BLS. The competition matrix $\mathbf{B}$ is defined based on the TNIC data calibrated to be as granular as the three-digit SIC codes. The R&D collaboration cost is specified as $\zeta_{ij,t} = \gamma_0 + \mathbf{c}_{ij,t}\gamma_1 + \mathbf{p}_{ij,t}\gamma_2 + \gamma_3|y_{i,t-1} - y_{j,t-1}| + \gamma_4|y_{i,t-1} - y_{j,t-1}|^2 + \gamma_5 a_{ij,t-1} + \gamma_6 \sum_{k=1}^{n} a_{ik,t-1}a_{jk,t-1}$. The M&A cost is specified as $\nu_{ij,t} = \kappa_0 + \mathbf{c}_{ij,t}\kappa_1 + \mathbf{c}_{ij,t} * \text{HHI}_{i,t}\kappa_2 + \mathbf{p}_{ij,t}\kappa_3 + \mathbf{s}_{i,t}\kappa_4 + \mathbf{s}_{j,t}\kappa_5 + \kappa_6(y_{i,t-1} - y_{j,t-1}) + \kappa_7(y_{i,t-1} - y_{j,t-1})^2 + \kappa_8 a_{ij,t-1} + \kappa_9 \sum_{k=1}^{n} a_{ik,t-1}a_{jk,t-1}$.

5 Counterfactuals

Based on the estimation results from Table 5 we can conduct simulations to analyze various counterfactual policy scenarios. Specifically, we examine the effects of bans on (or the absence of) forming R&D alliances and merger-acquisitions in separate counterfactual scenarios.

For illustration, we consider total firm output as the policy target, using the existing firm output as the benchmark. We analyze three counterfactual scenarios:

1. not allowing R&D alliances,

2. not allowing M&As, or

3. not allowing both R&D alliances and M&As.

In the first scenario, we eliminate all R&D alliance links in the R&D collaboration network. In the second scenario, we remove all M&A deals and restore the merged firms to their original acquirers and targets. In the third scenario, we remove both, R&D collaborations and M&A deals.

To obtain the missing productivities of acquirers and targets under the counterfactual scenario where M&A deals are prevented, we use their estimated productivity values (cf. Eq. (10)) from the previous sample period and apply linear extrapolation for the current period. We need the estimated productivity from the previous sample period for running extrapolation. We only obtain the counterfactual result from the second (2005-2010) to the fourth sample period (2015-2020).

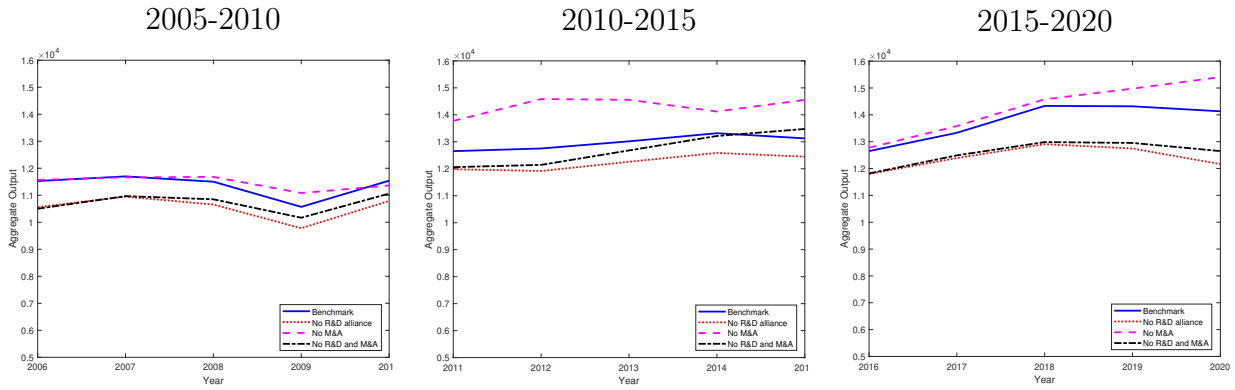


Figure 4. Counterfactual simulation results for the periods 2005 to 2020.

The simulation results shown in Figure 4 across different sample periods consistently indicate that banning M&A deals leads to higher firm output, while banning R&D alliances results in lower firm output. We are extending the model to allow for mergers to also benefit firms by reducing duplicative R&D costs and by allowing firms to reduce ex post holdup costs as in Williamson (1998) by integrating.

6 Conclusion

We estimate an R&D network formation model using a novel dataset on R&D collaborations extracted using natural language processing (NLP) from news articles over several decades. We find that technology spillovers in R&D collaborations have changed over time. Our results indicate that the changing pattern in technology spillovers is due to firms substituting R&D collaborations with M&A, emphasizing the close connection between R&D collaborations and M&A activities. Moreover, a counterfactual analysis indicates that M&As unanimously reduce combined firm output, while R&D alliances increase combined firm output.

We acknowledge that mergers may have positive benefits by reducing ex post hold-up costs. We are extending the model to allow for mergers to also benefit firms by reducing duplicative R&D costs and by allowing firms to reduce ex post holdup costs as in Williamson (1998) through integrating.

References

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3):425–455.

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

Alves, L. G., Mangioni, G., Rodrigues, F. A., Panzarasa, P., and Moreno, Y. (2022). The rise and fall of countries in the global value chains. *Scientific Reports*, 12(1):9086.

Antón, M., Ederer, F., Giné, M., and Schmalz, M. (2024). Innovation: The bright side of common ownership? *Management Science*.

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

Audretsch, D. B. and Belitski, M. (2020). The role of R&D and knowledge spillovers in innovation and productivity. *European economic review*, 123:103391.

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.

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. A., Cowan, R., and Jonard, N. (2010). Network-independent partner selection and the evolution of innovation networks. *Management Science*, 56(11):2094–2110.

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

Bimpikis, K., Ehsani, S., and Ilkılıç, R. (2019). Cournot competition in networked markets. *Management Science*, 65(6):2467–2481.

Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44.

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.

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.
Comola, M. (2012). Estimating local network externalities. *SSRN*, 946093.
Cunningham, C., Ederer, F., and Ma, S. (2021). Killer acquisitions. *Journal of political economy*, 129(3):649–702.
Dai, R. (2012). International accounting databases on WRDS: Comparative analysis. *Working paper, Wharton Research Data Services, University of Pennsylvania*.
De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, 12:775–799.
Dell, M. (2009). GIS analysis for applied economists. *Mimeo, MIT Department of Economics*.
Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
Ding, X., Zhang, Y., Liu, T., and Duan, J. (2014). Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1415–1425.
Ductor, L., Fafchamps, M., Goyal, S., and Van der Leij, M. J. (2014). Social networks and research output. *Review of Economics and Statistics*, 96(5):936–948.
Fafchamps, M., Van der Leij, M. J., and Goyal, S. (2010). Matching and network effects. *Journal of the European Economic Association*, 8(1):203–231.
Goerzen, A. (2007). Alliance networks and firm performance: The impact of repeated partnerships. *Strategic Management Journal*, 28(5):487–509.
Goyal, S. and Joshi, S. (2003). Networks of collaboration in oligopoly. *Games and Economic Behavior*, 43(1):57–85.
Goyal, S. and Moraga-Gonzalez, J. L. (2001). R&D networks. *RAND Journal of Economics*, 32(4):686–707.
Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4):691–719.
Gulati, R. (2007). *Managing Network Resources: Alliances, Affiliations, and Other Relational Assets*. Oxford University Press, USA.
Gulati, R., Sytch, M., and Tatarynowicz, A. (2012). The rise and fall of small worlds: Exploring the dynamics of social structure. *Organization Science*, 23(2):449–471.
Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2005). Market value and patent citations: A first look. *RAND Journal of Economics*, 36:16–38.
Hall, B. H., Mairesse, J., and Pierre, M. (2010). Mesuring the returns to R&D. *Handbook in Economics*, 2.
Hanaki, N., Nakajima, R., and Ogura, Y. (2010). The dynamics of R&D network in the IT industry. *Research Policy*, 39(3):386–399.
Hirschman, A. O. (1964). The paternity of an index. *The American Economic Review*, pages 761–762.
Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465.
Hoberg, G. and Phillips, G. M. (2024). Scope, scale, and concentration: The 21st-century firm. *The Journal of Finance*.
Honnibal, M. and Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.
Horn, R. A. and Johnson, C. R. (1990). *Matrix Analysis*. Cambridge University Press.
Hsie, C.-S., König, M. D., and Liu, X. (2018). Network formation with local complements and global substitutes: The case of R&D networks. *CEPR Discussion Paper No. DP13161*.

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

Hsieh, C.-S., König, M. D., and Liu, X. (2024). Endogenous technology spillovers in dynamic r&d networks. *Forthcoming in the RAND Journal of Economics*.

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

Igami, M. and Uetake, K. (2020). Mergers, innovation, and entry-exit dynamics: Consolidation of the hard disk drive industry, 1996–2016. *The Review of Economic Studies*, 87(6):2672–2702.

İmrohoroğlu, A. and Tüzel, Ş. (2014). Firm-level productivity, risk, and return. *Management Science*, 60(8):2073–2090.

Kalemlİ-Özcan, Ş., Sørensen, B. E., Villegas-Sanchez, C., Volosovych, V., and Yeşiltaş, S. (2024). How to construct nationally representative firm-level data from the Orbis global database: New facts on SMEs and aggregate implications for industry concentration. *American Economic Journal: Macroeconomics*, 16(2):353–374.

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

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

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. (2014). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics*, 9:695–752.

König, M. D., Battiston, S., Napoletano, M., and Schweitzer, F. (2012). The efficiency and stability of R&D networks. *Games and Economic Behavior*, 75:694—713.

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

König, M. D. and Rogers, T. (2023). Endogenous technology cycles in dynamic r&d networks. *European Economic Review*, 158:104531.

Koza, M. P. and Lewin, A. Y. (1998). The co-evolution of strategic alliances. *Organization science*, 9(3):255–264.

Kreps, D. M. and Scheinkman, J. A. (1983). Quantity precommitment and Bertrand competition yield Cournot outcomes. *The Bell Journal of Economics*, pages 326–337.

Kumar, P., Liu, X., and Zaheer, A. (2022). How much does the firm's alliance network matter? *Strategic Management Journal*, 43(8):1433–1468.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341.

Liben-Nowell, D. and Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559.

Liu, Y. (2018). The upward trend in the volatility of firm productivity shocks. *Economics Letters*, 163:68–71.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

López, Á. L. and Vives, X. (2019). Overlapping ownership, R&D spillovers, and antitrust policy. *Journal of Political Economy*, 127(5):2394–2437.

Lucking, B., Bloom, N., and Van Reenen, J. (2019). Have R&D spillovers declined in the 21st century? *Fiscal Studies*, 40(4):561–590.

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.

Martin, T., Zhang, X., and Newman, M. E. (2014). Localization and centrality in networks. *Physical review E*, 90(5):052808.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

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.

Meyer, C. D. (2000). *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics.

Nagel, S. (2016). Common crawl news. https://commoncrawl.org/2016/10/news-dataset-available/.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

Olley, G. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, pages 1263–1297.

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.

Peters, B., Roberts, M. J., Vuong, V. A., and Fryges, H. (2017). Estimating dynamic R&D choice: An analysis of costs and long-run benefits. *The RAND Journal of Economics*, 48(2):409–437.

Petrakis, E. and Tsakas, N. (2018). The effect of entry on R&D networks. *The RAND Journal of Economics*, 49(3):706–750.

Phillips, G. M. and Zhdanov, A. (2013). R&D and the incentives from merger and acquisition activity. *The Review of Financial Studies*, 26(1):34–78.

Poirier, D. (1980). Partial observability in bivariate probit models. *Journal of Econometrics*, 12(2):209–217.

Powell, W. W., Koput, K. W., Smith-Doerr, L., and Owen-Smith, J. (1999). Network position and firm performance: Organizational returns to collaboration in the biotechnology industry. *Research in the Sociology of Organizations*, 16(1):129–159.

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.

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.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, 4(1):87–94.

Schwenkler, G. and Zheng, H. (2019). The network of firms implied by the news. *Available at SSRN 3320859*.

Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words

with subword units. *arXiv preprint arXiv:1508.07909*.
Singh, N. and Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics*, 15(4):546–554.
Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*.
Tirole, J. (1988). *The Theory of Industrial Orgnization*. MIT Press.
Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
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.
Ugur, M., Churchill, S. A., and Luong, H. M. (2020). What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power. *Research Policy*, 49(1):103866.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
Vig, J. (2019). A multiscale visualization of attention in the transformer model. *arXiv preprint arXiv:1906.05714*.
Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review*, 23(176):88–93.
Vives, X. (1999). *Oligopoly Pricing: Old Ideas and New Tools*. MIT Press.
Williamson, O. E. (1998). The institutions of governance. *The American economic review*, 88(2):75–79.
Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.
Yamada, I., Asai, A., Shindo, H., Takeda, H., and Matsumoto, Y. (2020). Luke: deep contextualized entity representations with entity-aware self-attention. *arXiv preprint arXiv:2010.01057*.

Online Appendix

Contents

A Text Mining Algorithm: Implementation and performance **39**
A.1 LexisNexis document search . . . 39
A.2 Relationship extraction pipeline . . . 39
A.2.1 Problem statement . . . 39
A.2.2 Firm name recognition . . . 39
A.2.3 Representations from deep neural networks and transfer learning . . . 40
A.2.4 The transformer architecture and BERT . . . 41
A.2.5 Tokenization and input embeddings . . . 42
A.2.6 Details, output layer, and loss . . . 43
A.3 Knowledge Base and Training Data . . . 45
A.3.1 Knowledge base . . . 45
A.3.2 Corpus . . . 47
A.3.3 Constructing a training set . . . 48
A.4 Model performance . . . 49
A.4.1 Firm name recognition . . . 49
A.4.2 Relationship classification . . . 50
A.4.3 Inference on the news corpus . . . 51

B Additional Data Sources and Descriptives **52**
B.1 Network Characteristics . . . 52
B.2 Sectoral Composition . . . 59
B.3 Balance Sheet Statements, R&D and Productivity . . . 59
B.4 Geographic Location and Distance . . . 62
B.5 Nestedness and Modularity . . . 68
B.6 Orbis Company Information . . . 69
B.7 Patent Data . . . 69
B.8 Mergers and Acquisitions . . . 69
B.9 Compustat Data, Interpolation and Imputation . . . 75

C Cost Reducing R&D Collaborations **79**
C.1 Cournot Competition . . . 79
C.2 Herfindahl Index . . . 82
C.3 Bertrand Competition . . . 83

D Matrix Perturbation **87**

A Text Mining Algorithm: Implementation and performance

A.1 LexisNexis document search

To obtain articles that could contain alliance announcements, we applied a set of filters to LexisNexis' collection of news articles. We selected English language articles tagged as business news from newswires & press releases, newspapers, web-based publications, or industry trade press. Articles also have to contain at least one of the following list of keywords (which is based on the most frequent terms in the training data):

> alliance, venture, partnership, collaboration, agreement, form joint, formed joint, form a joint, formed a joint, plan to form, planned to form, marketing rights, marketing services, agreed to manufacture, development services, research and development, R&D, granted license, licensing rights, exclusive rights, exclusive license, license to manufacture, exploration services, mining exploration, gas exploration, marketing services, joint marketing, jointly market, joint research, jointly research, jointly manufacture

A.2 Relationship extraction pipeline

Extracting alliances from text data is a multi-step challenge. We will first give a problem statement, then discuss the firm name recognition solution, and finally discuss the relation classification model, including its theoretical underpinnings.

A.2.1 Problem statement

Our goal is to extract mentions of collaborative agreements between two or more firms from news articles and populate a database with the extracted relations. An example is given in figure 1. Given the input sequence from a news article, we want to realize that a collaborative agreement is mentioned, extract the names of the involved firms, and finally flag the type of agreement made. The inception date of the alliance is simply set to the date of the news article announcing it.

The relation extraction problem can be broken down into two sub-tasks: Extracting firm names is a Named Entity Recognition (NER) task. Flagging the type of relationship that is mentioned is a sequence classification task. We employ an existing solution for the NER task and train our own model for the sequence classification task.

A.2.2 Firm name recognition

In order to detect firm names, we use the Spacy transformer NER pipeline (Honnibal and Montani, 2017). The pipeline has state of the art performance and is much less computa-

tionally burdensome than other systems. The pipeline was trained on a manually annotated dataset to distinguish various types of entities, including organization names.[14] This is a supercategory to firms/corporation, however, we can simply filter out organizations that cannot be matched to a firm database. This also allows the model to extract university-industry collaborations (these are also present in the knowledge base), which might be of interest to some researchers. The more difficult problem is that of classifying the input sequence - identifying relevant sequences and then giving them the appropriate flags. We will discuss solutions to this problem below.

A.2.3 Representations from deep neural networks and transfer learning

There are a number of ways in which we could set up a model to extract and flag relations: For instance, we could come up with a set of rules (e.g. keywords occurring in a certain order) that would determine how a sequence is classified. This approach (called symbolic Natural Language Processing) usually falls short because of the high degree of variation in natural language and quickly becomes very complex, especially in a multi-label setting like ours.

A more modern approach is to construct representations of words in a vector space that try to capture its semantic properties in the given context. Given these vectors, we can then train a classifier (such as logistic regression or support vector machines). Naturally, the question is how to construct representations that are useful for downstream tasks. In the past, manually engineered vectors (e.g. binary indicators for whether a word co-occurs with a certain other word or is negated) or statistical embeddings (e.g. TF-IDF) have been used.

Since the 2010s, representations learned from deep neural networks have become the dominant methodology as they tend to perform substantially better. Commonly used language models now have hundreds of millions of parameters, and large neural networks require large amounts of training data. One important paradigm that has emerged because of that *transfer learning*. Language models like BERT, GPT, or XLM are pre-trained on massive amounts of raw text data (such as the Common Crawl web crawl corpus) on a task like next word prediction. The resulting model can then be used for supervised downstream tasks like sequence classification or question answering, either by taking the hidden states of the neural network as a direct input for another classifier or by adding a task-specific output layer and fine-tuning the neural network using the given labels. The pre-trained model therefore serves as very useful initialization of the network, that appears to contain a lot of knowledge about the semantics of a language. Compared to pre-training, which requires substantial computational resources (and produces emissions comparable to those of a transatlantic flight, see Strubell et al. (2019)), fine-tuning is relatively inexpensive and can be done on a single GPU.

[14] According to the definition at `https://schema.org/Organization`.

A.2.4 The transformer architecture and BERT

Another important innovation has been the transformer architecture proposed by Vaswani et al. (2017). Recurrent architectures such as in Long Short Term Memory (LSTM) networks have the disadvantage of being computationally costly (their recurrent nature precludes parallelization). The transformer architecture solves this problem by feeding entire input sequences simultaneously into multiple layers of encoder and decoder blocks. Given our classification task, we are only concerned with the encoder blocks (the decoder serves to produce an output sequence, e.g. for translation).

First, since the words of input sequences are not fed to the model in order, we need to add positional encodings to input tokens. This is simply a number corresponding to the relative position of an input token in the sequence. Now, what happens in each encoder block? The inputs first go through a self-attention mechanism and then into a fully connected feedforward neural network.

Similar to how humans gain understanding by focusing on particular combinations of words when reading a sentence (e.g. subject, verb, and object), the attention mechanism lets the network focus on a subset (in a fuzzy sense) of other input tokens when encoding each token. Using the attention patterns as input helps the neural network produce better context-dependent encodings. The transformer architecture uses scaled dot-product attention[15]:

$$A(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where the query Q is the current word that is operated on and (K, V) is a set of learned key-value pairs. First, a similarity score of query and key is computed as $softmax(QK^T)$. Then it is multiplied with the stored values V. The division by $\sqrt{d_k}$ is simply a normalization to avoid vanishing gradients in the softmax (d_k is the dimension of keys). The transformer architecture in fact employs multi-head attention, meaning that attention is computed multiple times in parallel with different learned matrices K and V. The independent outputs from attention heads are then concatenated and projected into the desired dimension. This method of ensembling creates more useful and varied attention outputs. Figure A.1 shows the output of three attention heads (visualized with different colors) when queried with the word 'alliance' within the example sequence. The pink attention head makes the connection to the verb, while the orange head recognizes a connection to the word 'to', potentially allowing the model to make a connection to the purpose of the alliance. The blue head produces a less peaked output, connecting to the subjects, the verb, and the purpose of the alliance. The set of keys and values is learned like other parameters and varies between attention heads and layers.

[15]It is worth noting that the time complexity of this type of self-attention scales quadratically with the sequence length T, so shorter input sequence are preferable.

The outputs of the self-attention layer for each position (each input token) are now each fed independently to a fully connected feed-forward neural network (FNN). The FNN applies two learned linear transformations with a rectified linear (ReLU) activation function in between:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2$$

where the matrices (W_1, W_2) and the vectors (b_1, b_2) represent the learned weights and biases, respectively. These learned parameters vary from layer to layer. It is worth noting that each encoder block also employs residual connections between the self-attention and FNN sublayers, and finally applies layer normalization (Ba et al., 2016), though we will not discuss these techniques in detail here.

The transformer was initially developed for sequence to sequence tasks (such as translation) but it turns out we can use the encoder part of the model to get useful representations for downstream tasks. This is exactly what BERT does: it builds **B**idirectional **E**ncoder **R**epresentations from **T**ransformers (Devlin et al., 2018).

$\text{BERT}_{\text{large}}$ uses the transformer architecture with 24 encoder layers, each with 16 attention heads and 1024 neurons in the FFN. This makes for a total of 340 million parameters. The model is trained on two tasks, masked language modeling (randomly replacing a certain input token with [MASK] and letting the model predict the original input), and next sentence prediction (making the binary prediction whether a sentence follows another one). The pre-training corpora are the Books Corpus and English Wikipedia.

A.2.5 Tokenization and input embeddings

So far we have not discussed how words are fed into the encoder - after all, it takes as inputs vectors of real numbers and not a string sequence. A demonstration of how BERT generates input embeddings is given in figure A.2. BERT uses WordPiece tokenization (Wu et al., 2016) with a 30,000-word vocabulary. Essentially, it finds the 30,000 most common (sub-)words in the pre-training corpus and associates them with unique identifiers, which are then used as an input to the encoder. Out-of-vocabulary words are broken down greedily into sub-words that can be found in the vocabulary. As can be seen in the figure, the firm name 'Atos' is broken down into 'At', and '##os', where the '##' characters signify the continuation of a word. The BERT tokenizer also adds a [CLS] token to the beginning of each sequence, and a [SEP] token to separate sentences. The hidden states associated with the [CLS] token can be used for sequence classification tasks (they are representative for the whole sequence). The input layer of the neural network has a fixed size, so all inputs need to have the same length T. We, therefore, add a [PAD] token to bring shorter sequences to the desired length (these tokens are given 0 weight by an attention mask so they do not affect results). Longer sequences will simply be truncated at the maximum input length.

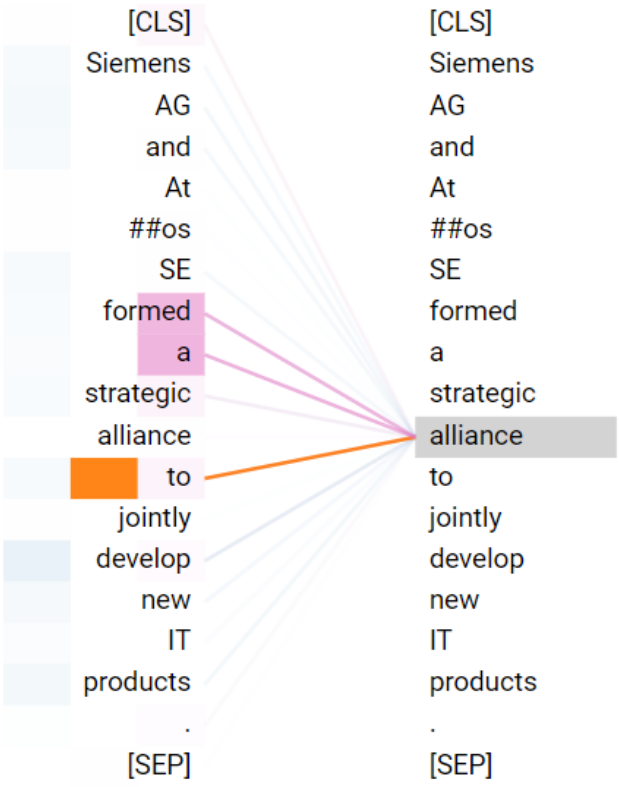


Figure A.1. Visualization of the attention patterns produced by three attention heads in a BERT model. The patterns are at the 5th layer for the word 'alliance'. Visualization produced using Vig (2019).

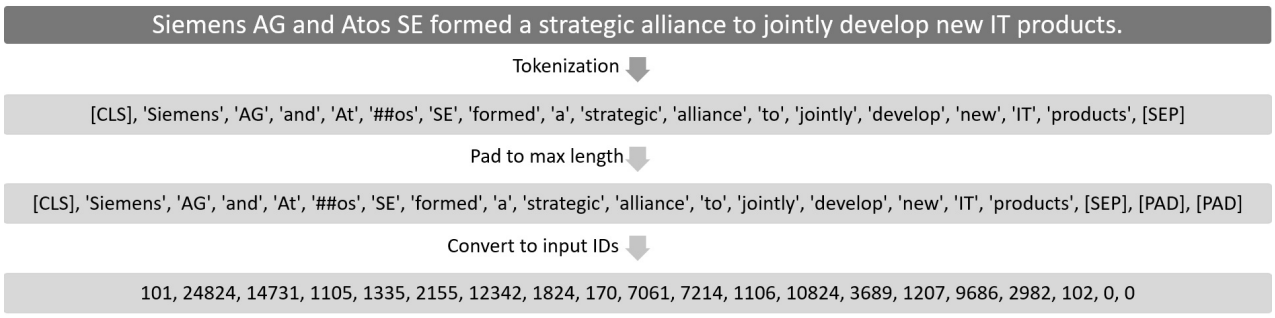


Figure A.2. How BERT creates input representations.

A.2.6 Details, output layer, and loss

In practice, we will make use of Facebook's RoBERTa model (Liu et al., 2019). RoBERTa is architecturally the same as $\text{BERT}_{\text{large}}$, but is shown to improve on BERT in almost all down-

stream tasks by using longer pre-training with bigger batches, simplified training strategies (only masked language modeling), and a better set of hyperparameters. It is also trained on a larger set of corpora, importantly for us including the Common Crawl News dataset (Nagel, 2016). The fact that the model has been pre-trained on a large news dataset likely leads to better performance in our downstream task in the same domain, with similar documents. Note that RoBERTa uses the BPE tokenizer (Sennrich et al., 2015) with a larger vocabulary that we will not discuss in detail here.

Let us now discuss the output layer we need to add to the RoBERTa model for our multi-label classification task. We decide to include 10 labels in the analysis: whether an agreement is a Joint Venture or Strategic Alliance, whether the agreement is completed/signed or planned, and 6 flags for the type/purpose of the agreement: marketing, manufacturing, R&D, licensing, supply[16], and exploration (these flags are not mutually exclusive). We also try to predict whether a technology transfer takes place (which is also not mutually exclusive with any other labels). The final label is simply whether a given sequence is relevant - in the sense of containing an alliance announcement. Note that in this setup negative examples, which are encoded as a vector of zeros with length 10, are strictly speaking incorrect for the two label-encoded dichotomous classifications (JV/SA and completed/signed). Since negative examples do not contain an alliance announcement it does not make sense to assign them to either of these classes. However, we have not found this to affect results.

Since we have 10 labels, we need 10 neurons in the output layer, each taking inputs of size 1024 (the size of the last hidden layer of RoBERTa associated with classifier token [CLS]), and producing output of size 1. Since we want non-exclusive outputs, we take a sigmoid activation function for each neuron:

$$\hat{p}(y_i) = \sigma(z_i) = \frac{1}{1 + e^{-z_i}},$$

where $\hat{p}(y_i)$ is the predicted probability for class i and z_i is the output of neuron i. This setup means we can output probabilities for all classes independently from each other. We also need an appropriate loss function. Since we already take the sigmoid activation function in the output layer, we use a weighted binary cross-entropy loss, which for a given example is:

$$L(\hat{p}(y), y) = -\sum_{i=1}^{C} w_i \big(y_i \log(\hat{p}(y_i)) + (1 - y_i) \log(1 - \hat{p}(y_i)) \big)$$

where y is the vector of ground-truth labels for that example, w_i is the weight of label i, and C is the number of labels. This function computes the loss for each element in y (each class) independently. With the label weights, we can adjust how much emphasis is put on getting correct predictions for each label. In practice, we will use them to make up for the class

[16]Note that the supply flag in the SDC data does not identify the direction of a buyer-supplier relationship.

imbalance in the dataset. We want to predict all labels equally well, so we calculate label weights as

$$w_i = \frac{1}{2\text{SP}_i},$$

where SP_i is the share of examples in the training set that is labeled positively for label i. This means that a label with an equal number of positive and negative examples receives weight 1, while a label with only one-quarter of examples being positive would receive weight 2.

For training the model, we take a batch size of 64 and fine-tune over 2 epochs over the data. We set the maximum sequence length to $T = 64$ tokens. This is enough to accommodate almost all examples in full length (and in the cases where the sequence is cut off the tokens at the end would likely not be important for classification). We use the Adam optimizer (Kingma and Ba, 2014) with a learning rate of $5e^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, as Devlin et al. (2018) used for fine-tuning BERT. The model was trained on a single Tesla T4 in 6 hours using 16-bit floating-point precision.

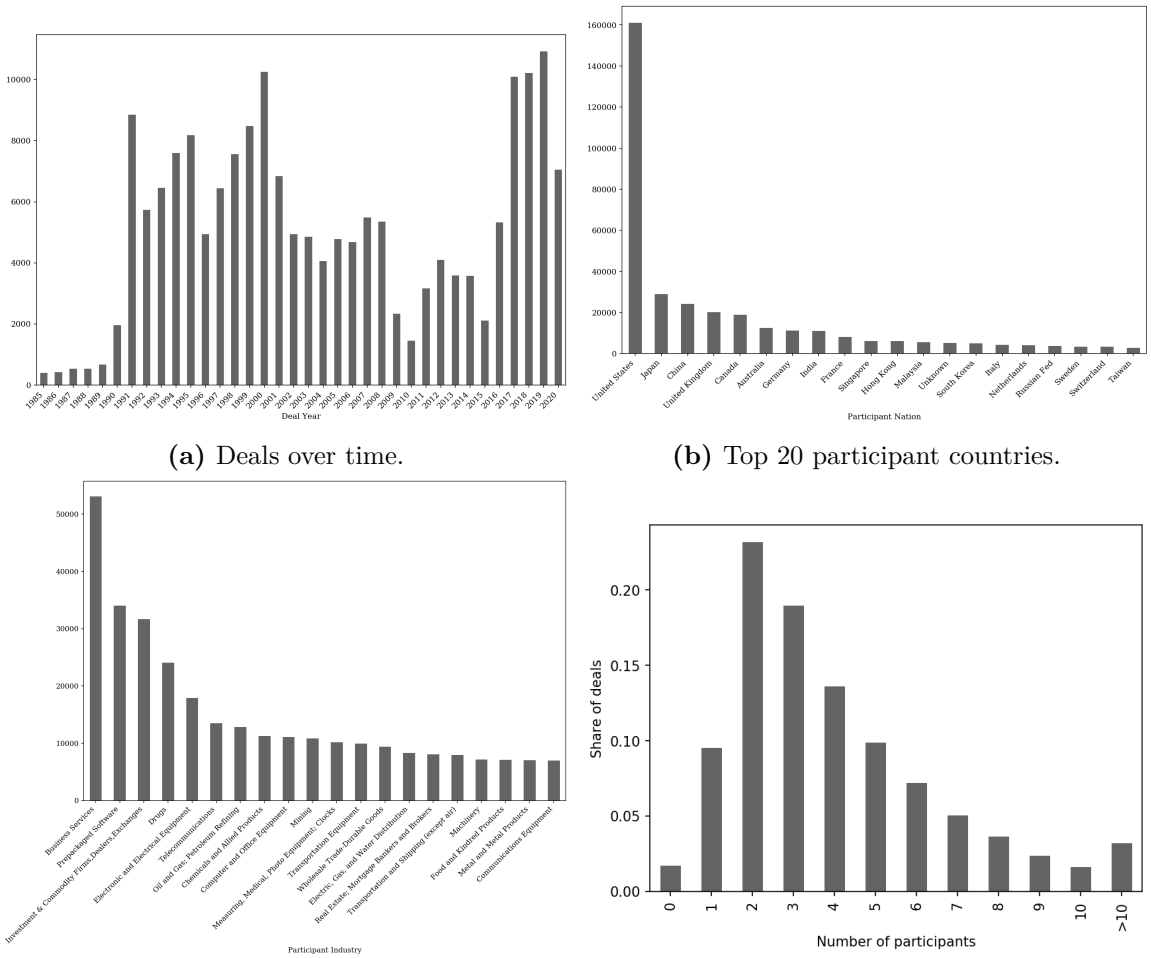
A.3 Knowledge Base and Training Data

We use a knowledge base of alliances, which provides positive examples to train and evaluate our model on. We also use a corpus of news articles, which is used to generate negative examples for the training set and finally run inference on.

A.3.1 Knowledge base

We use the Refinitiv (formerly Thomson Financial) SDC Platinum Joint Venture (JV)/Strategic Alliance (SA) database as our knowledge base. The database has a wide coverage of worldwide inter-firm relations: It collects JV/SA transactions from newspapers, trade journals, government filings, and newswires. It features a little over 196,000 entries in the period from 1985 to the present. The distribution of these entries along time, countries, and industries is given in Figure A.3, along with the number of participants per deal. The number of deals per year varies quite significantly, with a noticeable drop in entries between the years 2009 and 2016. Given that the data is incomplete, we cannot evaluate whether this is a real-world trend or is due to a change in methodology. The overwhelming majority of participants is headquartered in the United States. The data covers a wide range of industries and does not appear to have the strong bias towards particular industries that other alliance databases have. The number of participants is typically between two and four. The data does have some missing participant names, with zero or only one recorded participants (clearly an alliance takes at least two participants). Our firm name recognition system can be used to fill in these missing names.

According to Refinitiv, the data is created by content analysts manually evaluating each source document. Since the database also features the original 'deal text' that an entry is



(a) Deals over time.

(b) Top 20 participant countries.

(c) Top 20 participant industries.

(d) Number of participants per deal.

Figure A.3. Distribution of deals in the SDC Platinum JV/SA database.

based on, it represents a rich source of manually labeled examples that can be used to teach information extraction algorithms.

We narrow down our sample according to several criteria. First, we drop observations with missing deal text, missing flags for the type of deal, or clearly erroneous entries. We then focus on deals with a ?completed/signed?, "pending", or "terminated" status. There are only a few entries that do not fall into these categories; we remove them to populate our test set only with examples of 'positive' developments, i.e. an agreement being reached rather than reviewed. We use the terminated agreements as negative examples. As to why we do not remove "pending" observations, a typical deal text for entries with "pending" status is given below:

"Redcargo Logistics and GD Express Carrier Bhd planned to form a strategic alliance.The purpose of the strategic alliance is to provide GDEX customers with access to AirAsias extensive network, meaning goods are able to be transported and delivered efficiently on more than 5,000 weekly flights across Asia Pacific."

The original source only mentions the agreement is "planned". Unfortunately, these entries are not typically updated after an agreement is actually signed. In many cases, it is easily verified using a search engine that these agreements came into effect at one point. Not considering these entries would therefore mean missing a lot of developments (about 44% of entries have "pending" as status). We simply take the pending/completed distinction as another variable to be predicted. This also means that our system can improve upon the existing data by updating agreements from planned to signed once this is confirmed in a news article.

There are 16 flags in the dataset that describe the kind of relation stated in the source document. Every entry is first categorized as a strategic alliance or joint venture. Then there are a number of flags that describe the purpose of the agreement, e.g. manufacturing or marketing. We are only interested in the most common types of relations and exclude several categories that are empirically rare from our analysis. We combine the flags "Licensing Agreement", "Exclusive Licensing Agreement", and "Cross-Licensing Agreement" into a common licensing category, as well as combining "Technology Transfer" and "Cross-Technology Transfer" (it is often unclear from the deal text whether there is technology transfer in both directions; the distinction between the two labels seems to be assigned relatively arbitrary). The number of occurrences and co-occurrences for each flag is given in Figure A.4. The SA/JC and pending/completed classification are mutually exclusive, all other labels have some overlap. The latter is empirically plausible, as a deal might for instance include both a manufacturing and R&D component. This makes this a multi-label classification problem. To account for the fact that SA/JC and pending/completed are mutually exclusive categories, we choose label encoding for these variables and one-hot encoding for the other flags. This transforms the problem into a multi-label binary classification problem. The corresponding neurons in the output layer are therefore tasked to classify a deal text as JV or SA, pending or completed, licensing agreement or not, manufacturing agreement or not, etc.

A.3.2 Corpus

Our corpus is a collection of financial news articles from Bloomberg and Reuters between October 2006 to November 2013. The collection was released by Ding et al. (2014). After removing articles with missing dates, and empty or scrambled bodies of text, we are left with a total of 421,387 news articles. Most articles date to the years 2010 to 2013 and three-quarters of articles are sourced from Bloomberg (the remaining articles are from Reuters). The average length of each article is 525 words (including the headline). Few articles only have a headline; most fall somewhere between 100 and 1000 words. Since we have to choose

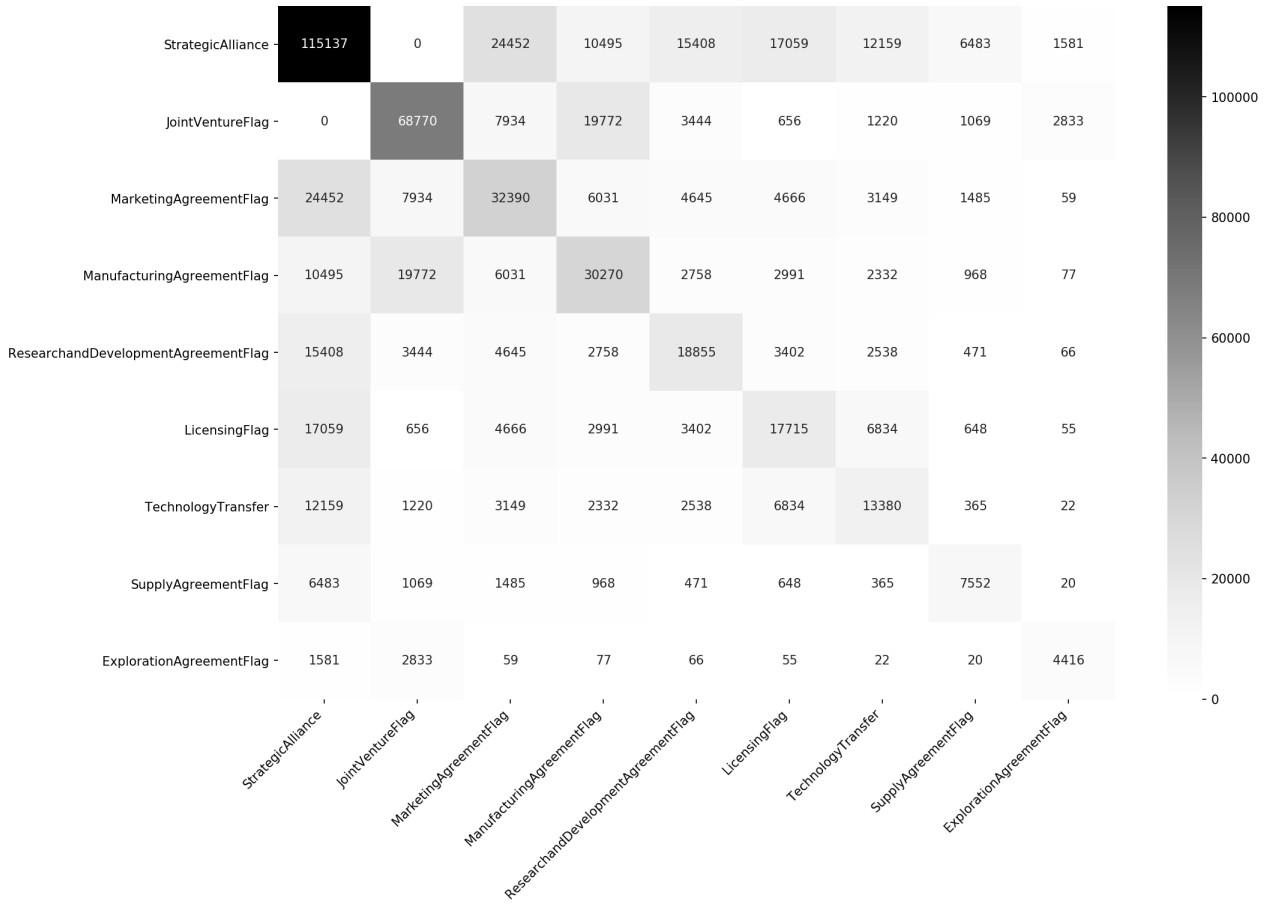


Figure A.4. Ocurrences and co-occurences of relationship labels in the SDC Platinum JV/SA database.

a maximum sequence length to feed into our model (e.g. 128 tokens), we would have to split longer documents into smaller chunks of that length with some overlap as not to lose the context. We choose not to do this for inference and instead just parse the beginning of each document. This is because we want to focus on articles where a given deal is the main piece of news reported, not articles that mention an old deal as background information somewhere deep in the body of the text. First, this makes it much more plausible to take the date of the article as the date of the inception of the deal. Second, such articles will provide a better and more succinct description of the deal.

A.3.3 Constructing a training set

We want the model to be able to first detect if a document/sequence is relevant and then give it the appropriate flags. To detect if a document is relevant (in the sense of mentioning

a collaborative deal between two or more firms) we add another outcome variable, where all examples from the knowledge base are labeled positive. We construct negative examples by randomly sampling sequences from the beginning of documents in the corpus. Note that this will result in some incorrect labels - a share of the randomly sampled sequences will contain mentions of collaborative agreements (these are the mentions we are trying to extract after all). We, therefore, exclude sequences with the most common keywords ('Strategic Alliance', 'Joint Venture', 'Research and Development'). To keep negative examples from the corpus as close to the examples from the knowledge base, we only use full sentences. We also try to fit the length of negative texts as close as possible to the examples in the knowledge base. We, therefore, copy the distribution of deal text lengths (in sentences) in the knowledge base to generate examples with variable lengths from this distribution. The variation in the randomly sampled negative examples should help the model learn to differentiate what makes a sequence relevant, and the similar structure in positive and negative examples should prevent overfitting.

We also add negative examples from the knowledge base: alliance entries with status 'terminated'. Without these examples, the model would likely also extract mentions of deal terminations, which is not what we want (we do not have enough terminated deals in the knowledge base to teach the model to recognize terminations as a separate class). We get 2476 examples of terminated agreements from the knowledge base. We construct the rest of the negative examples by randomly sampling from the corpus as explained above until we reach an equal number of negative and positive examples.

We have 359,622 total examples, of which half are negative and half are positive. We use a train-test split of 2:1 since we have plenty of training data and a large test set allows us to get a more accurate estimate of model performance. We do not bother with making a stratified train-test split. Considering the number of examples in relation to the class distribution (compare Figure A.4) we get a very even split by the law of large numbers (we verified this to be true for our particular split).

A.4 Model performance

In the following, we evaluate the pipeline on firm name recognition, relation classification, and finally the quality of inference on previously unseen articles from the corpus.

A.4.1 Firm name recognition

First, we want to check how accurately our NER system detects the firm names mentioned in deal texts. For each deal in the knowledge base, we compare the set of recorded participants A with the set of organizations recognized by the algorithm B. We compute the share of recognized participants as

$$S(A, B) = \frac{|A \cap B|}{|A|},$$

meaning that we give partial credit for recognizing some of the participants. We then simply average this score over all deals in the knowledge base. The Spacy model was trained on completely unrelated data so we can use the full knowledge base as a test set, to check whether the model works well for this application. To check whether individual recognized firm names are the same, we strip the names from both A and B of all legal business type identifiers (such as "Co.", Inc.", and "Ltd"). We take the full list of legal identifiers from Schwenkler and Zheng (2019), who also use it to solve the problem of firm name matching.

The average share of recognized participants in the full knowledge base is an excellent 89.2 percent. Note that about 10 percent of entries in the knowledge base have zero or only one recorded participant. These entries were excluded from calculating the score because they are clearly incomplete (for these cases the firm name recognition algorithm can also help to fill in the missing participant names).

A.4.2 Relationship classification

Next, we need to evaluate the performance of our model in the relation classification task. We will look at precision (true positives as a share of true positives plus false positives), recall (true positives as a share of true positives plus false negatives), the F1-measure (the harmonic mean of precision and recall), as well as accuracy (the share of correct predictions). We do this for each label individually, but we also look at the Macro score subset accuracy, which is the share of examples where the predictions exactly match the set of *all* labels. Note that the alliance flags only make sense if an alliance is mentioned. We, therefore, evaluate the accuracy for these labels only on knowledge base examples, as not to inflate the accuracy scores (it does not make a difference for the other scores).

Macro average gives greather emphasis to rare labels The performance metrics are given in Table 6. The model achieves great precision and recall on finding relevant sequences with alliance agreements, distinguishing pending from completed agreements and joint ventures from strategic alliances. It also does fairly well on assigning the manufacturing and licensing agreement flags, and a little bit worse on the R&D, marketing, and exploration agreement flags. The model is not able to predict the supply agreement and technology transfer agreement flags well. This can be caused both by the limited number of positive examples for these labels, as well as labels that do not follow a clear decision rule. When inspecting the original deal text along with the supply agreement and technology transfer labels it becomes clear that these are relatively subjective codings with reliability issues. The supply agreement flag seems to have been applied The system also achieves a subset accuracy of 89.1 percent. The relatively poor performance on the technology transfer and supply agreement labels does not affect this score much because these examples are relatively uncommon.

	Precision	Recall	F1-score	Accuracy
Relevant	0.989	0.971	0.980	0.981
Pending	0.938	0.964	0.951	0.970
Joint Venture	0.983	0.990	0.987	0.990
Marketing agreement	0.864	0.739	0.797	0.933
Manufacturing agreement	0.878	0.830	0.853	0.953
R&D agreement	0.793	0.749	0.777	0.954
Licensing agreement	0.913	0.800	0.853	0.973
Supply agreement	0.674	0.461	0.548	0.969
Exploration agreement	0.776	0.764	0.770	0.990
Technology transfer	0.694	0.540	0.607	0.950

Table 6. Model performance on the relation classification task.

A.4.3 Inference on the news corpus

Finally, we also want to get an idea of how well the model does in terms of inference on our news corpus. First off, the model is highly scalable: on one Tesla T4 GPU, we can do inference on roughly 150 documents per second. This means we can process millions of documents within a couple of hours.

An interesting question to ask is, for a given period of time, how many of the deals recorded in the knowledge base we can extract from our corpus. Naturally, this number is limited by the number of relevant documents in the corpus. Our corpus does not feature many of the primary sources used in the SDC data, such as business wires. Therefore we cannot expect to be able to extract a large share of deals from the knowledge base, but it is nevertheless good to see the amount of overlap between extracted alliances and alliances in the SDC data. We will take the year 2012 as a test for this since this is the year with most articles in our corpus. We take both knowledge base entries and negative examples from the corpus from 2012 out of the training set so this data has not been seen before by the model.

Out of 159,207 news articles in the year 2012, the model identified an alliance in 605 articles. On provisional inspection of the extracted alliances. only very few of them seem to be incorrect and the labels seem accurate, too.[17] 112, or 18.5 percent of the extracted alliances can be found in the SDC data (we match entries if at least two of the participants are identical). Conversely, 2.7 percent of the 4156 alliances in the SDC data for this year are extracted from the corpus. As mentioned before, we would expect this share to rise with the quality and size of the analyzed corpus.

[17]Ideally, one would manually label the extracted alliances based on the news text to systematically evaluate the precision of the model on this data.

B Additional Data Sources and Descriptives

B.1 Network Characteristics

The total number of firms in our sample is 15,103. Tables 7, 8, 9 and 10 show the 25 highest ranked firms according to their degree (i.e. the number of collaborations) in the year 1990, 2000, 2010 and 2020, respectively. From these tables we observe that the rankings are highly unstable, with the highest ranked firms changing across the years considered. This could be due to the changing importance of different sectors in terms of the R&D collaboration intensity as illustrated in Figure B.3.

Figure 2 shows the largest connected component in the observed network of R&D collaborations in the year 2020. The size of the connected component is increasing, in particular, reaching a peak in the year 2000, consistent with the time evolution of the largest component size shown in the panel in the third row and last column in Figure B.2. Moreover, the identities of the firms with the largest number of collaborations are changing across the years considered (see Tables 7, 8, 9 and 10, respectively, for a more complete list). As mentioned above, this could be due to the changing importance of different sectors in terms of the R&D collaboration intensity as illustrated in Figure B.3.

Figure B.1 shows the degree distribution, $P(d)$, the average nearest neighbor connectivity (the average degree of the neighbors of a node with degree d), $k_{\text{nn}}(d)$, the clustering degree distribution (the average clustering coefficient of a node with degree d), $C(d)$,[18] and the component size distribution, $P(s)$. All distributions are highly skewed, indicating a large degree of inequality (indicated with a power-law decay in the upper tails) among firms in the network.

Figure B.2 shows the evolution of the number of nodes/firms, n, the number of links/collaborations, m, the average degree, $\bar{d}$, the degree variance, σ_d^2, the degree coefficient of variation (standard deviation divided by the mean), c_v, the assortativity coefficient, γ, the clustering coefficient, C,[18] the average path length, ℓ,[19] the size of the largest connected component, $\max_{H\subseteq G}|H|$, the largest real eigenvalue, λ_{PF}, and the eigenvector centralization, C_v.

We observe that the number of firms, n, the number of links, m, the average degree, $\bar{d}$, the degree variance, σ_d^2, the clustering coefficient, C, and the size of the largest connected component, $\max_{H\subseteq G}|H|$, are showing an oscillatory behavior similar to the oscillations documented in König and Rogers (2023) where an alternative data set on R&D collaborations has been analyzed. The assortativity coefficient (measuring degree correlations of connected nodes) shows a decreasing trend until the year 2000 indicating that over this time period a transition from an assortative (positive degree correlation) to a dissortative network (negative degree correlation) has taken place. The average path length, ℓ, is increasing until the year 1995 and stabilizing thereafter. The largest real eigenvalue, λ_{PF}, and the eigenvector

[18] The clustering coefficient is the fraction of the number of links between the neighbors of a node divided by the maximum number of links that could possibly exist between the neighbors.

[19] The average path length is defined as the average number of links along the shortest paths for all possible pairs of nodes in the network.

Table 7. The 25 highest ranked firms according to their degree (i.e. the number of collaborations) in the year 1990.

Name	Deg.	NACE	NACE Description	Rank
Eastman Kodak Co	15	26	Manufacture of computer, electronic and optical products	1
Du Pont Limited	14	20	Manufacture of chemicals and chemical products	2
General Electric Company	14	30	Manufacture of other transport equipment	3
Sun Limited	11	55	Accommodation	4
Ciba - Geigy	10	20	Manufacture of chemicals and chemical products	5
AT&T Inc.	10	61	Telecommunications	6
Texas Instruments Inc	10	26	Manufacture of computer, electronic and optical products	7
Sun Microsystems (south Africa) (pty) Ltd	10	58	Publishing activities	8
Philips Gmbh	9	64	Financial service activities, except insurance and pension funding	9
Kodak Limited	9	58	Publishing activities	10
Mitsubishi Corp.	9	46	Wholesale trade, except of motor vehicles and motorcycles	11
Toshiba Corp.	8	26	Manufacture of computer, electronic and optical products	12
Amoco	8	47	Retail trade, except of motor vehicles and motorcycles	13
Olin Corp	8	20	Manufacture of chemicals and chemical products	14
Tandem Computers Inc	8	46	Wholesale trade, except of motor vehicles and motorcycles	15
Combustion Engineering	8	35	Electricity, gas, steam and air conditioning supply	16
Monsanto Co	8	20	Manufacture of chemicals and chemical products	17
Phillips Petroleum	7	61	Telecommunications	18
Eos Holding Aktiengesellschaft	7	70	Activities of head offices	19
Fujitsu Limited	7	26	Manufacture of computer, electronic and optical products	20
Mobil	7	47	Retail trade, except of motor vehicles and motorcycles	21
BASF Se	6	20	Manufacture of chemicals and chemical products	22
Sumitomo Chemical Company Limited	6	20	Manufacture of chemicals and chemical products	23
Computer Systems (private) Limited	6	62	Computer programming, consultancy and related activities	24
Bass	6	22	Manufacture of rubber and plastic products	25

Note: The degree measures the number of R&D collaborations of a firm. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

Table 8. The 25 highest ranked firms according to their degree (i.e. the number of collaborations) in the year 2000.

Name	Deg.	Sales [%]	NACE	NACE Description	Rank
Microsoft Corporation	75	1.6413	58	Publishing activities	1
Siemens Ag	61	0.0003	28	Manufacture of machinery and equipment nec	2
Dupont	48	0.0593	45	Wholesale and retail trade and repair of motor vehicles and motorcycles	3
Hitachi Ltd	36	38.3966	27	Manufacture of electrical equipment	4
Fujitsu Limited	35	14.8351	26	Manufacture of computer, electronic and optical products	5
Ericsson Ab	31	5.7755	62	Computer programming, consultancy and related activities	6
Toshiba Corporation	29	16.2305	26	Manufacture of computer, electronic and optical products	7
NEC Corporation	29	14.0908	26	Manufacture of computer, electronic and optical products	8
Cisco Systems Inc	29	0.0534	26	Manufacture of computer, electronic and optical products	9
Mitsubishi Corporation	27	17.8074	46	Wholesale trade, except of motor vehicles and motorcycles	10
Johnson & Johnson	27	0.0006	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	11
Texas Instruments Inc	26	0.0335	26	Manufacture of computer, electronic and optical products	12
Oracle Corp	25	0.7315	58	Publishing activities	13
Intel Corp	25	0.0001	26	Manufacture of computer, electronic and optical products	14
Dow Chemical Company	24	0.0003	20	Manufacture of chemicals and chemical products	15
Lockheed Martin Corp	22	0.0002	30	Manufacture of other transport equipment	16
AT&T Inc.	21	0.3888	61	Telecommunications	17
Novartis Ag	20	0.7222	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	18
Monsanto Co	20	0.0001	20	Manufacture of chemicals and chemical products	19
3Com Corp	19	0.0357	61	Telecommunications	20

Note: The degree measures the number of R&D collaborations of a firm. Sales are given as relative percentages to the sum across all firms without missing information in the same sector. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

Table 9. The 25 highest ranked firms according to their degree (i.e. the number of collaborations) in the year 2010.

Name	Deg.	R&D Exp. [%]	Sales [%]	NACE	NACE Description	Rank
GlaxoSmithKline Plc	65	0.3072	0.0322	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	1
Roche Holding Ag	50	0.7393	0.0538	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	2
Merck & Co., Inc.	37	0.6077	0.0522	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	3
Sanofi	36	0.3245	0.0345	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	4
Pfizer Inc	33	0.6967	0.0739	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	5
Siemens Ag	31	0.4692	0.1806	28	Manufacture of machinery and equipment nec	6
NEC Corporation	29	12.6315	7.3834	26	Manufacture of computer, electronic and optical products	7
General Electric Company	27	0.3652	0.6312	30	Manufacture of other transport equipment	8
Microsoft Corporation	26	15.6695	2.3998	58	Publishing activities	9
Dupont	24		0.0009	45	Wholesale and retail trade and repair of motor vehicles and motorcycles	10
Johnson & Johnson	24	0.5151	0.0702	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	11
Bristol-Myers Squibb Company	21	0.2525	0.0221	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	12
Bayer Ag	19	0.2251	0.0398	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	13
Utek Srl	19		0.0000	28	Manufacture of machinery and equipment nec	14
Astellas Pharma Inc	19	14.4207	1.1057	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	15
Sony Group Corporation	19	19.7732	14.8650	26	Manufacture of computer, electronic and optical products	16
Galapagos Nv	17	0.0743	0.0015	72	Scientific research and development	17
Novartis Ag	17	0.6688	0.0574	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	18
Fujitsu Limited	17	10.2963	9.6425	26	Manufacture of computer, electronic and optical products	19
Mitsubishi Corporation	17	23.1795	9.7727	46	Wholesale trade, except of motor vehicles and motorcycles	20
Dow Chemical Company (the)	17	0.0007	0.0003	20	Manufacture of chemicals and chemical products	21
Boeing Company	17	0.3858	0.2714	30	Manufacture of other transport equipment	22
Chevron Corporation	16	15.6677	5.1036	19	Manufacture of coke and refined petroleum products	23
Takeda Gmbh	15		0.0000	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	24
BP Plc	15	10.6546	7.6506	19	Manufacture of coke and refined petroleum products	25

Note: The degree measures the number of R&D collaborations of a firm. R&D expenditures and sales are given as relative percentages to the sum across all firms without missing information in the same sector. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

Table 10. The 25 highest ranked firms according to their degree (i.e. the number of collaborations) in the year 2020.

Name	Deg.	R&D Exp. [%]	Sales [%]	NACE	NACE Description	Rank
Microsoft Corporation	42	40.0118	37.8335	58	Publishing activities	1
Sanofi	40	4.1969	4.2934	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	2
Pfizer Inc	38	5.7766	4.0434	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	3
Siemens Ag	34	24.6868	11.0049	28	Manufacture of machinery and equipment nec	4
Roche Holding Ag	31	8.5992	6.4252	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	5
Bayer Ag	31	5.4674	5.0688	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	6
GlaxoSmithKline Plc	31	3.8910	4.4424	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	7
Merck & Co., Inc.	29		0.0261	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	8
Hitachi Ltd	28	14.7669	19.7496	27	Manufacture of electrical equipment	9
Airbus Se	28	13.7443	8.3549	30	Manufacture of other transport equipment	10
Boeing Company	28	9.7036	8.0015	30	Manufacture of other transport equipment	11
Volkswagen Ag	27	11.4087	12.1360	29	Manufacture of motor vehicles, trailers and semi-trailers	12
Denso Corporation	27	5.5751	1.9788	29	Manufacture of motor vehicles, trailers and semi-trailers	13
Thales	26	1.0438	1.4423	26	Manufacture of computer, electronic and optical products	14
Novartis Ag	25	5.5539	4.8440	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	15
Fujitsu Limited	24	0.8535	2.2435	26	Manufacture of computer, electronic and optical products	16
Toyota Motor Corporation	23	12.3557	10.9087	29	Manufacture of motor vehicles, trailers and semi-trailers	17
Evotec Se	22	0.0485	0.0597	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	18
Lockheed Martin Corp	21	5.0948	8.9212	30	Manufacture of other transport equipment	19
Takeda Gmbh	20		0.1970	21	Manufacture of basic pharmaceutical products and pharmaceutical preparations	20
Intel Corp	20	11.2532	5.3872	26	Manufacture of computer, electronic and optical products	21
Mitsubishi Corporation	19	0.0000	10.4974	46	Wholesale trade, except of motor vehicles and motorcycles	22
Qualcomm Inc	19	4.9600	1.5034	26	Manufacture of computer, electronic and optical products	23
BASF Se	18	13.1553	8.6832	20	Manufacture of chemicals and chemical products	24
NEC Corporation	18	0.0000	1.8712	26	Manufacture of computer, electronic and optical products	25

Note: The degree measures the number of R&D collaborations of a firm. R&D expenditures and sales are given as relative percentages to the sum across all firms without missing information in the same sector. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

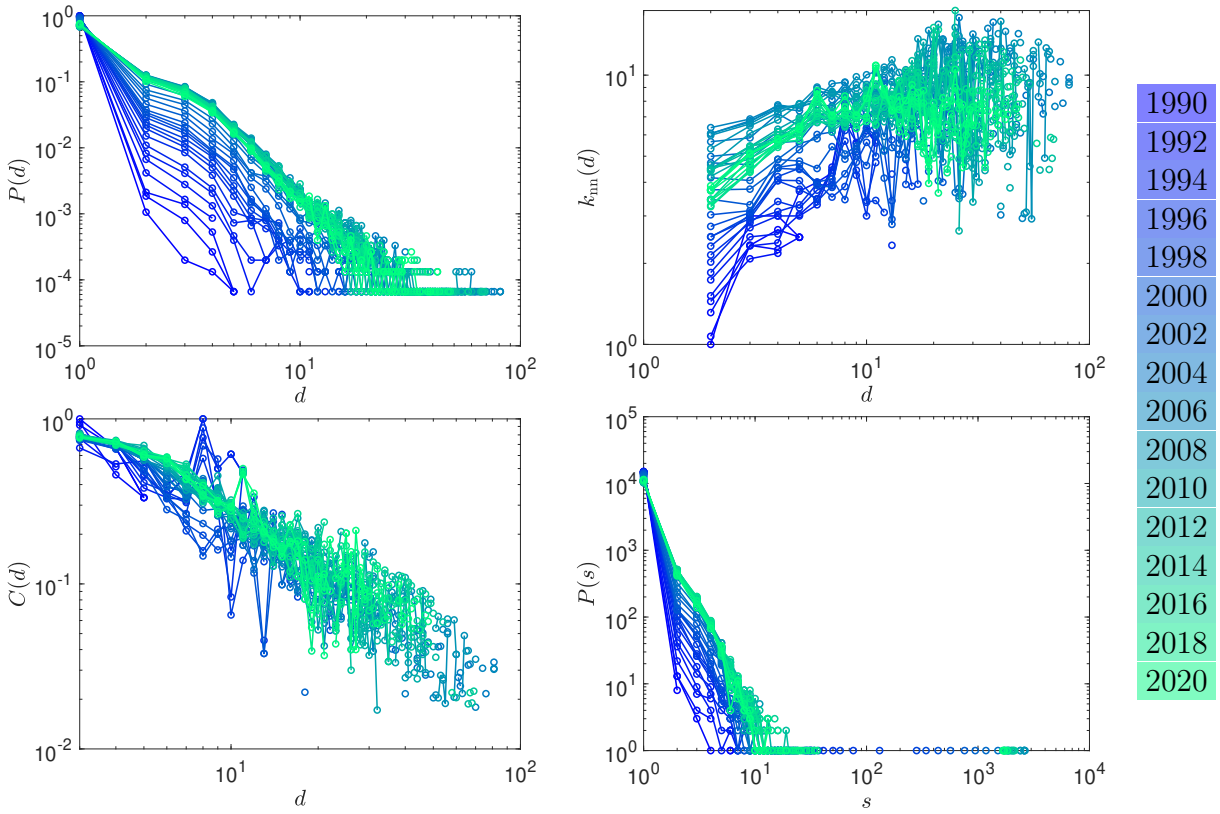


Figure B.1. The degree distribution, $P(d)$, the average nearest neighbor connectivity, $k_{\mathrm{nn}}(d)$, the clustering degree distribution, $C(d)$, and the component size distribution, $P(s)$.

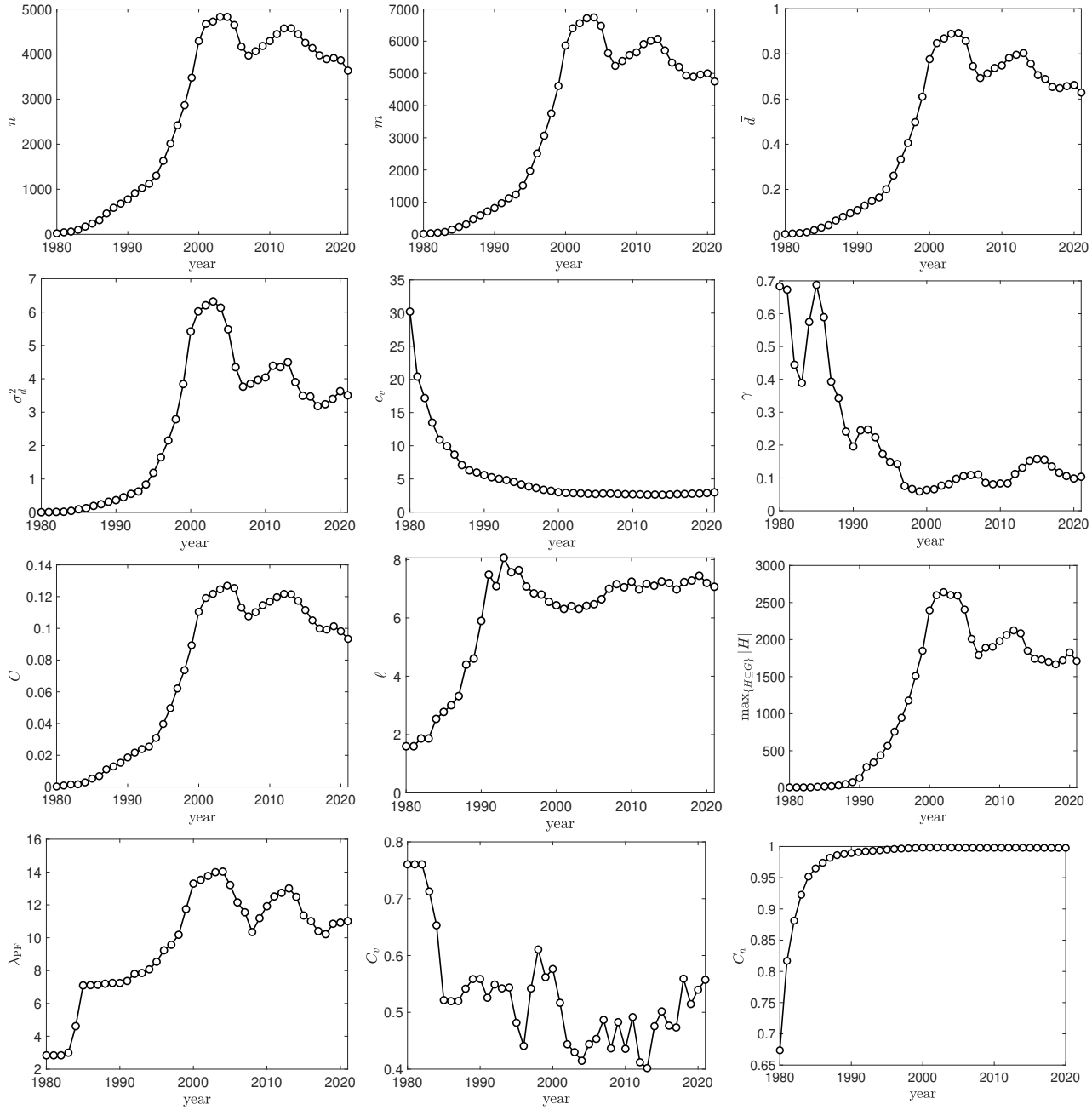


Figure B.2. The number of nodes, n, the number of links, m, the average degree, $\bar{d}$, the degree variance, σ_d^2, the degree coefficient of variation (standard deviation divided by the mean), c_v, the assortativity coefficient, γ, the clustering coefficient, C, the average path length, ℓ, the size of the largest connected component, $\max_{H\subseteq G}|H|$, the largest real eigenvalue, λ_{PF}, the eigenvector centralization, C_v, and the nestedness coefficient, C_n.

centralization, C_v, of the underlying adjacency matrix of the R&D network are related to the technology diffusion properties of the R&D network (as discussed in König et al. (2012)). The largest real eigenvalue, λ_{PF} is increasing up to the year 2005 and oscillating thereafter. The eigenvector centralization, C_v, is decreasing from an initially high value with a peak in 2000 and shows an increasing trend after 2012.

Hsie et al. (2018) have documented that R&D networks tend to have a hierarchical structure that can be measured by the "nestedness" of the underlying adjacency matrix (see also König et al. (2014)). Nested graphs have a core-periphery structure which has also been documented in empirical studies on R&D networks (Kitsak et al., 2010; Rosenkopf and Schilling, 2007). We compute the degree of nestedness of a matrix by calculating the *matrix temperature*, T_n (Atmar and Patterson, 1993). Typically, the lower the temperature T_n, the higher the degree of nestedness. More precisely, T_n is normalized in such a way that it ranges between 0 for a perfectly nested matrix and 100 for a maximally "unnested" matrix. From the nestedness temperature we can compute the nestedness coefficient, C_n, as $C_n = (100 - T_n)/100$, such that C_n ranges from zero for a totally random (non-nested) network to one for a totally nested network (Almeida-Neto et al., 2008). The bottom right panel in Figure B.2 shows the evolution over time of the nestedness coefficient, C_n, with a nestedness coefficient close to one in the years after 2000.

B.2 Sectoral Composition

Figures B.17 and B.5 together with Tables 11 and 12 show the largest sectors in our data at the 2-digit and 3-digit NACE Rev. 2 levels, respectively. The dominance of the pharmaceutical sector becomes apparent. However, the sectoral composition is changing over time, as illustrated in Figure B.3, with pharmaceuticals appearing as the dominant sector only after the year 2005.

B.3 Balance Sheet Statements, R&D and Productivity

We matched the firms in our R&D collaboration database with Bureau van Dijk (BvD)'s Orbis database to obtain information about their balance sheets and income statements. The Orbis database is a commercial dataset, which contains administrative data on 130 million firms worldwide. Orbis is an umbrella product that provides firm-level data covering over 120 countries, both developed and emerging, since 2005. The financial and balance-sheet information in Orbis comes from business registers collected by the local Chambers of Commerce to fullfill legal and administrative requirements and are relayed to BvD via over 40 different information providers. Orbis contains not only information about publicly listed firms, but provides also information about private firms.[20]

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

[20]For further discussion of the Orbis databases see Dai (2012), Bloom et al. (2013) and Papadopoulos (2012).

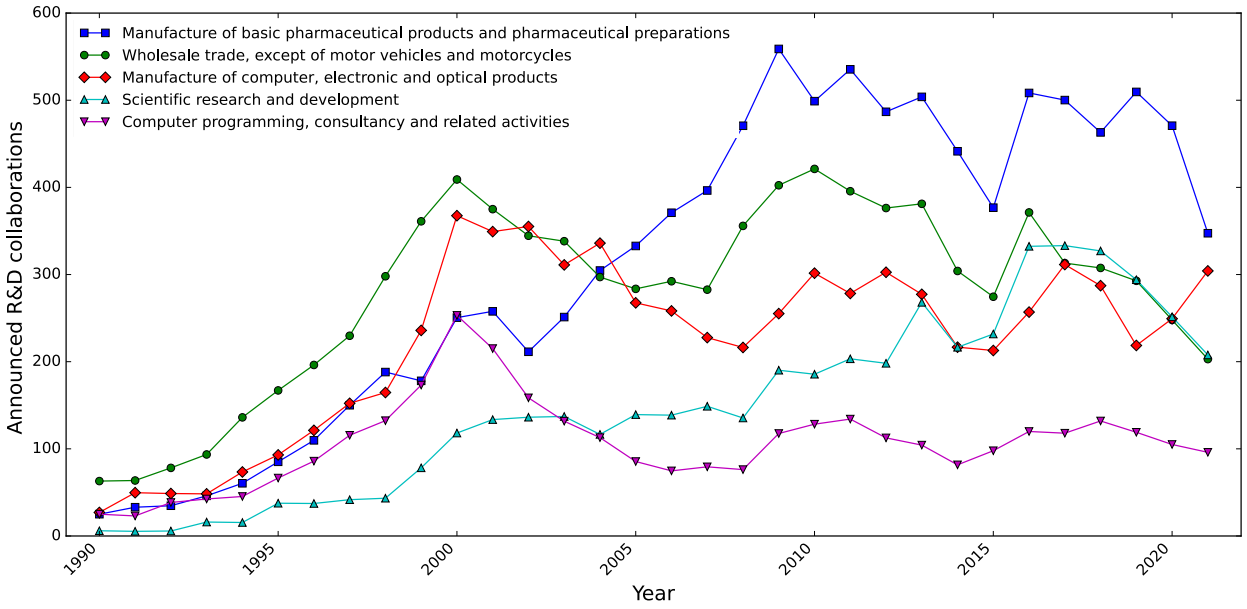


Figure B.3. R&D collaborations in the top 5 industries over time. Note that data for 2021 is only partially available (until August 2021).

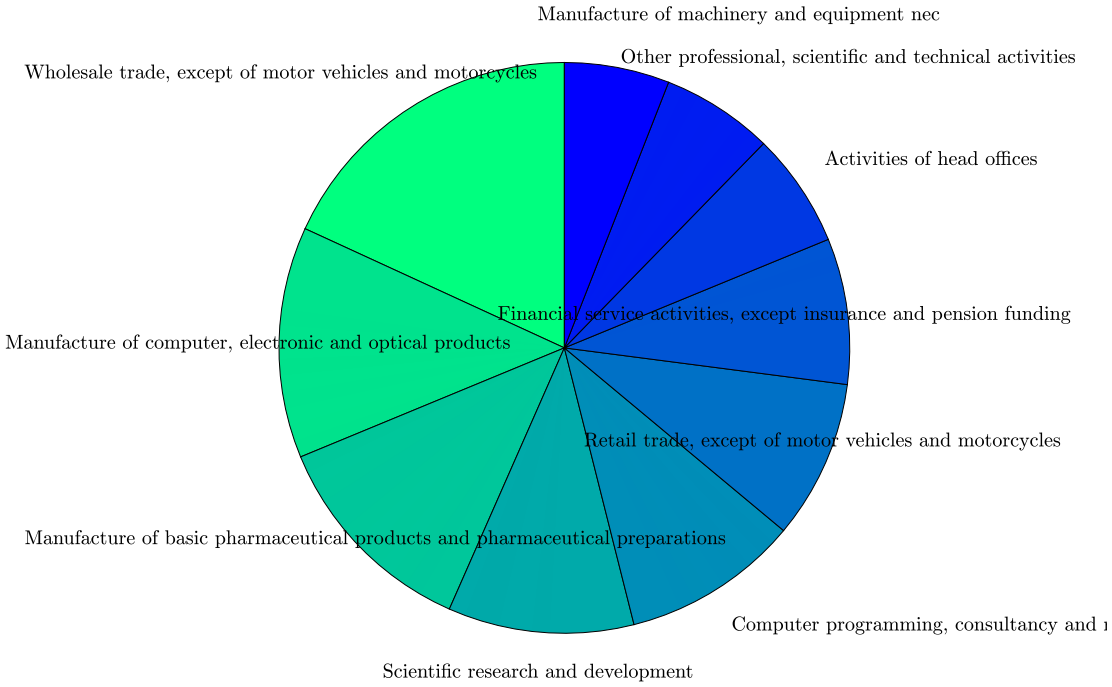


Figure B.4. The shares of the ten largest sectors at the 2-digit NACE Rev. 2 levels. See also Table 11, respectively. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

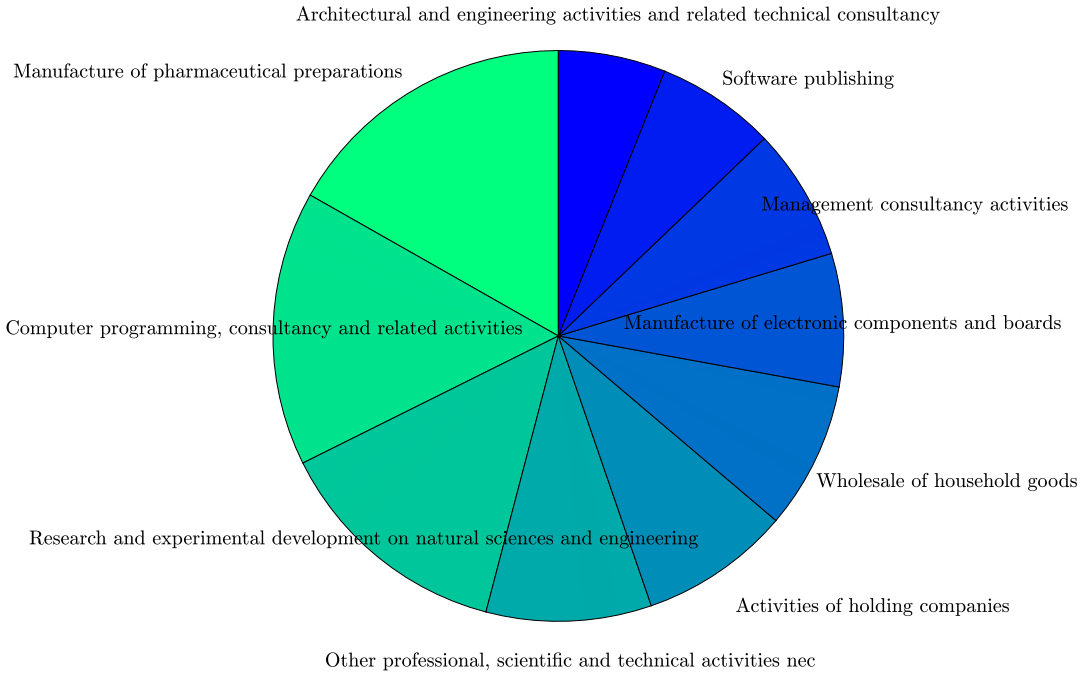


Figure B.5. The shares of the ten largest sectors at the 3-digit NACE Rev. 2 levels. See also Table 12, respectively. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

Table 11. The 20 largest sectors at the 2-digit NACE Rev. 2 level.

Sector	2-dig NACE	# firms	% of tot.	Rank
Wholesale trade, except of motor vehicles and motorcycles	46	1201	8.43	1
Manufacture of computer, electronic and optical products	26	867	6.09	2
Manufacture of basic pharmaceutical products and pharmaceutical preparations	21	805	5.65	3
Scientific research and development	72	697	4.89	4
Computer programming, consultancy and related activities	62	663	4.66	5
Retail trade, except of motor vehicles and motorcycles	47	596	4.19	6
Financial service activities, except insurance and pension funding	64	547	3.84	7
Activities of head offices	70	430	3.02	8
Other professional, scientific and technical activities	74	420	2.95	9
Manufacture of machinery and equipment nec	28	394	2.77	10
Human health activities	86	394	2.77	11
Publishing activities	58	361	2.54	12
Manufacture of chemicals and chemical products	20	360	2.53	13
Specialised construction activities	43	315	2.21	14
Telecommunications	61	310	2.18	15
Architectural and engineering activities	71	305	2.14	16
Other manufacturing	32	299	2.10	17
Activities auxiliary to financial services and insurance activities	66	265	1.86	18
Manufacture of fabricated metal products, except machinery and equipment	25	251	1.76	19
Office administrative, office support and other business support activities	82	247	1.73	20

Note: NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

Table 12. The 20 largest sectors at the 3-digit NACE Rev. 2 level.

Sector	3-dig NACE	# firms	% of tot.	Rank
Manufacture of pharmaceutical preparations	212	715	5.02	1
Computer programming, consultancy and related activities	620	663	4.66	2
Research and experimental development on natural sciences and engineering	721	578	4.06	3
Other professional, scientific and technical activities	749	399	2.80	4
Activities of holding companies	642	363	2.55	5
Wholesale of household goods	464	355	2.49	6
Manufacture of electronic components and boards	261	321	2.25	7
Management consultancy activities	702	319	2.24	8
Software publishing	582	288	2.02	9
Architectural and engineering activities and related technical consultancy	711	259	1.82	10
Other telecommunications activities	619	240	1.69	11
Retail sale of other goods in specialised stores	477	219	1.54	12
Other human health activities	869	212	1.49	13
Activities auxiliary to financial services, except insurance and pension funding	661	202	1.42	14
Other specialised wholesale	467	196	1.38	15
Manufacture of medical and dental instruments and supplies	325	191	1.34	16
Business support service activities nec	829	188	1.32	17
Construction of residential and non-residential buildings	412	181	1.27	18
Non-specialised wholesale trade	469	181	1.27	19
Manufacture of instruments and appliances for measuring, testing and navigation	265	173	1.21	20

Note: NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

92.5% of the firms with the Orbis database.

Figure B.6 shows the sales, output (deflated sales), R&D expenditures, and the productivity (value added per worker) distribution for different years ranging from 1990 to 2020. All distributions are highly skewed, indicating a large degree of inequality (indicated with a power-law decay in the upper tails) in firms' sizes, R&D expenditures and productivity.

Moreover, Figure B.7 shows a correlation scatter plot for sales, output, R&D expenditures and productivity. All are highly correlated except for employment and productivity.

B.4 Geographic Location and Distance

The number of firms in each country is shown in Figures B.10, B.11 and B.11, respectively, while Table 13 shows the 25 countries with the largest numbers of firms. The dominant role of the U.S. with 7,758 firms making up 52.61% of the total number of firms is clearly visible. The second largest country in terms of R&D collaborating firms is the United Kingdom with 967 firms, which comprises 6.56% of all collaborations, which is far below the corresponding number for the US. This unequal distribution across countries is a persistent phenomenon across time, as illustrated in Figure B.8.

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

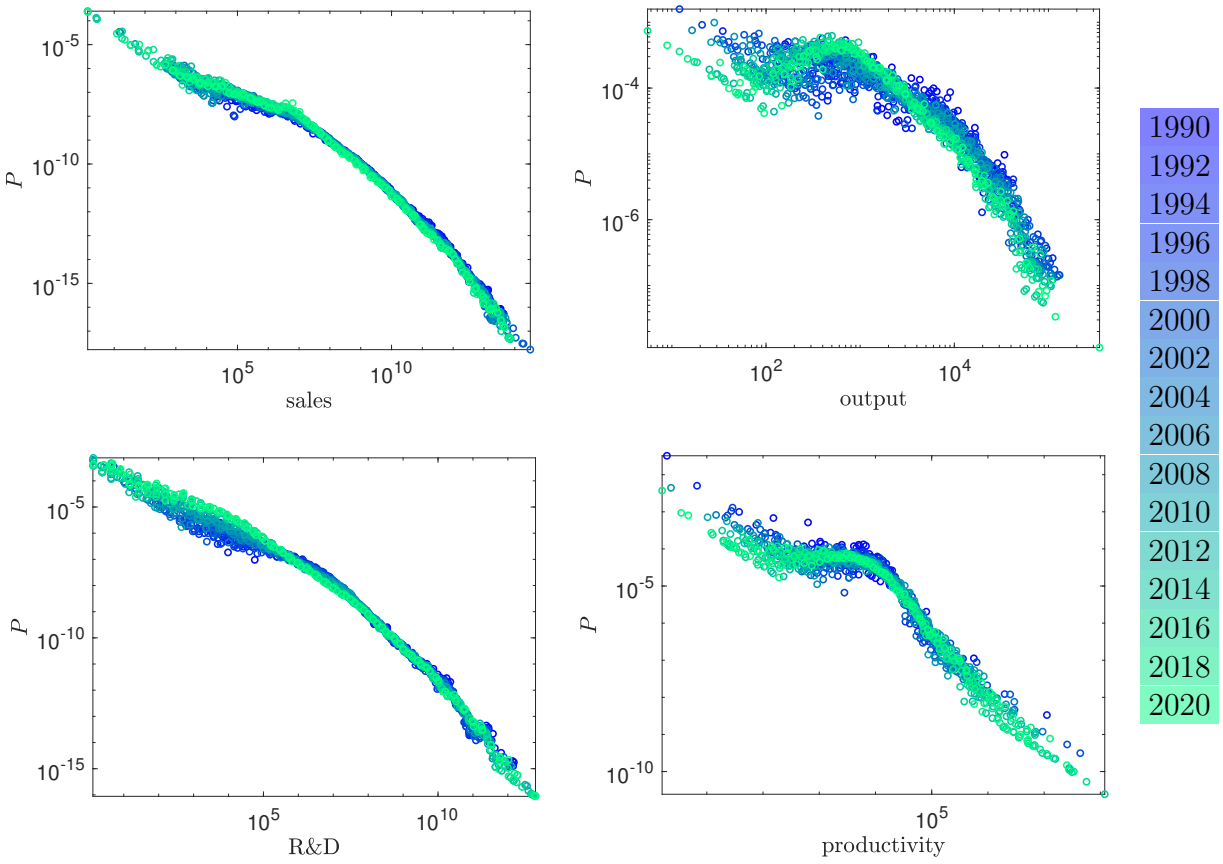


Figure B.6. The sales distribution, output distribution, R&D expenditures distribution, and the productivity distribution, across different years shown by using a logarithmic binning of the data (McManus et al., 1987).

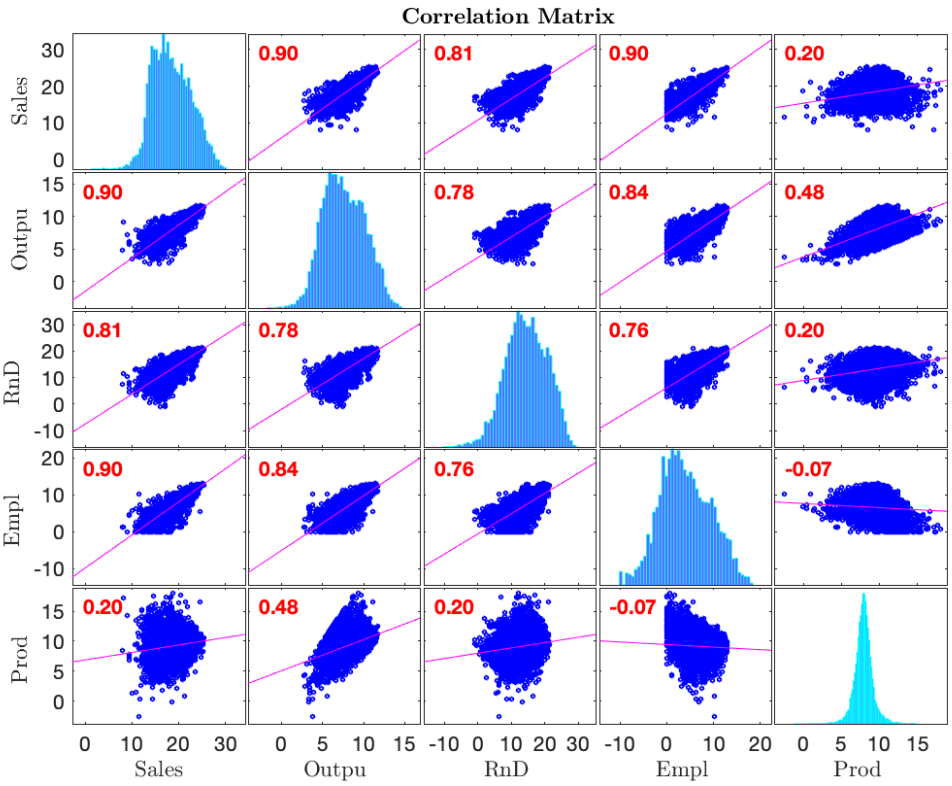


Figure B.7. Correlation scatter plot for sales, output, productivity, R&D expenditures, employment and productivity at the firm level for the stacked data cross the years 1990 to 2021.

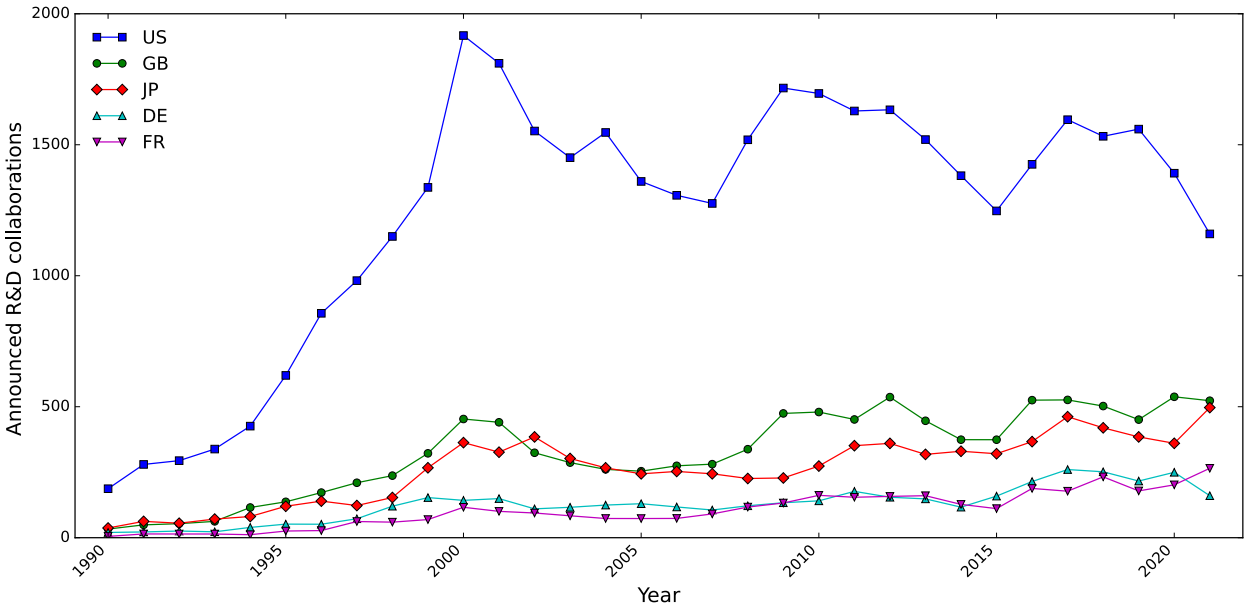


Figure B.8. R&D collaborations in the top 5 countries over time. Note that data for 2021 is only partially available (until August 2021).

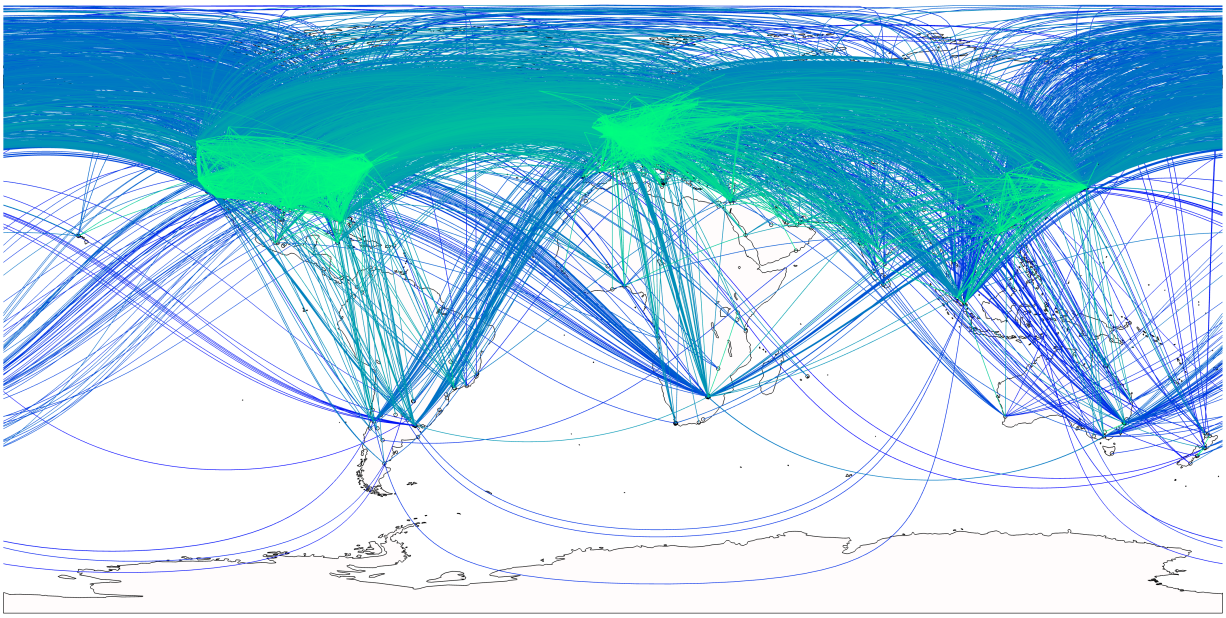


Figure B.9. The locations (at the city level) and collaborations of the firms accumulated over all years (1980 to 2021).

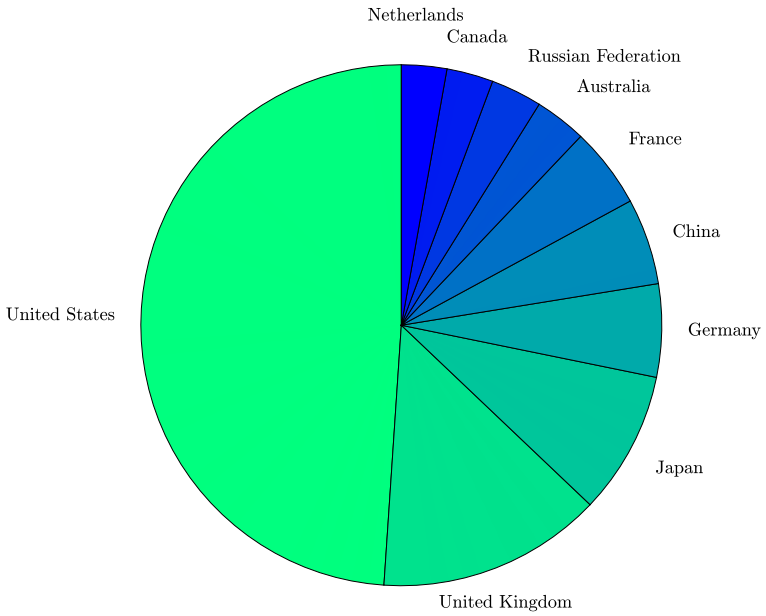


Figure B.10. Pie chart illustrating the relative number of firms across countries.

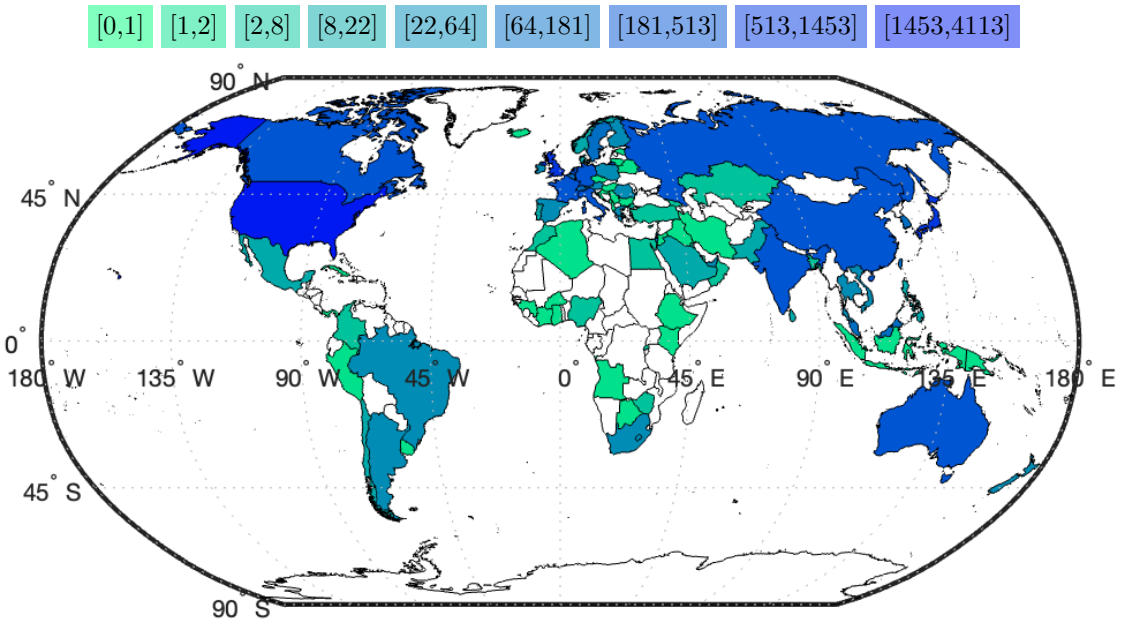


Figure B.11. The number of firms in each country in the data sample.

Table 13. The 25 countries with the largest numbers of firms.

Country Name	Code	# firms	% of tot.	Rank
United States	USA	4113	38.31	1
United Kingdom	GBR	1174	10.94	2
Japan	JPN	747	6.96	3
Germany	DEU	485	4.52	4
China	CHN	450	4.19	5
France	FRA	420	3.91	6
Australia	AUS	270	2.51	7
Russia	RUS	267	2.49	8
Canada	CAN	243	2.26	9
Netherlands	NLD	236	2.20	10
Italy	ITA	219	2.04	11
India	IND	193	1.80	12
Hong Kong	HKG	173	1.61	13
Korea	KOR	171	1.59	14
Sweden	SWE	129	1.20	15
Switzerland	CHE	117	1.09	16
Singapore	SGP	115	1.07	17
Belgium	BEL	90	0.84	18
Taiwan	TWN	90	0.84	19
Malaysia	MYS	80	0.75	20
Israel	ISR	79	0.74	21
Finland	FIN	63	0.59	22
Ireland	IRL	57	0.53	23
South Africa	ZAF	53	0.49	24
Poland	POL	48	0.45	25

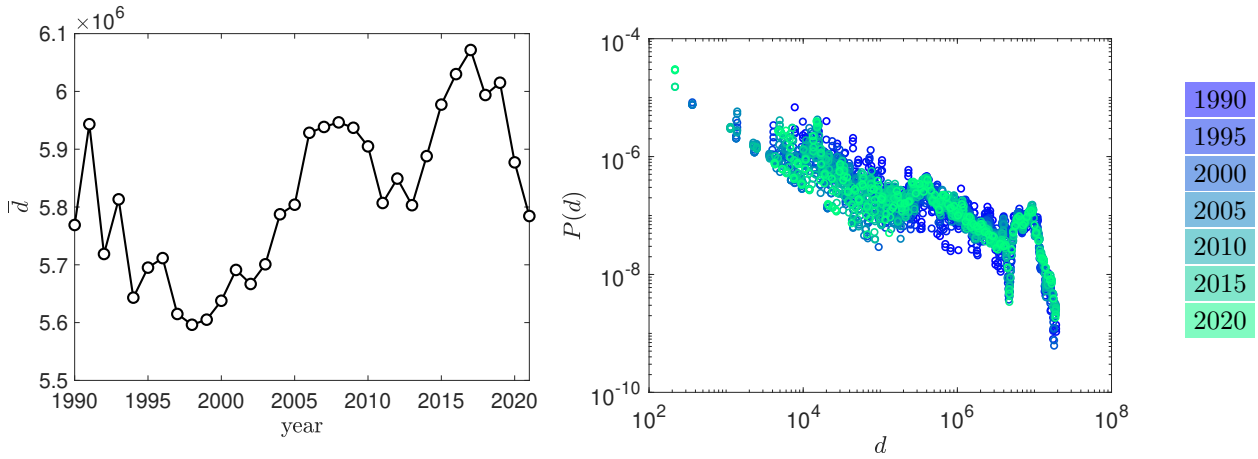


Figure B.12. (Left panel) The average distance between collaborating firms. (Right panel) The distance distribution, $P(d)$, across collaborating firms.

e.g., Dell, 2009) and the Google Maps Geocoding API. 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 B.12, while Figure B.9 shows the locations (at the city level) and collaborations of the firms in the database. The distance distribution, $P(d)$, is heavily skewed. We find that R&D collaborations tend to be more likely between firms that are close, showing that geography matters for R&D collaborations and this spillovers, in line with previous empirical studies (Lychagin et al., 2016). Moreover, Figure B.12 shows that the distance between R&D collaborating firms is fluctuating over time.

B.5 Nestedness and Modularity

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

After having identified these modules we can compute the nestedness of each module (cf. Figure B.2). The top left panel in Figure B.13 shows the adjacency matrix of the largest connected component of the R&D network in the year 1990 where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices of each module (sub-network) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Almeida-Neto

et al., 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n). Figures B.14, B.15 and B.16 show the same information but for the years 2000, 2010 and 2020. We observe that all modules are highly nested, and most modules considered are also significantly nested, in particular, in the years 2000, 2010 and 2020. This is consistent with the bottom right panel in Figure B.2 showing the evolution over time of the nestedness coefficient, C_n, with a nestedness coefficient close to one in the years after 2000.

B.6 Orbis Company Information

For company data, we use Bureau van Dijk's Orbis database. It contains firms' address and contact details, industry information (NACE codes), the number of employees; as well as financial information such as revenue, value added, assets, liabilities, and R&D expenditure. We focus on companies for which all this information is available in the years up to 2021. We match the firm names detected in news articles to Orbis, after which we are left with 165,929 firms globally. The distribution of companies across sectors from the Orbis data can be seen in Figure B.17.

B.7 Patent Data

The Worldwide Patent Statistical Database (PATSTAT) is provided by the European Patent Office (EPO). It covers patent applications to authorities in 90 countries. The data include: applicant and inventor names and addresses, filing date, titles, and abstracts of patent applications; technology classification and technically related patents, and bibliographical information such as number of citations and citation links. Our version of the database is from spring 2021.

Applicants in the database can be both natural and legal persons (while inventors may only be natural persons) and we focus on company applicants.[21] We are left with around 3.6 million distinct companies, which filed 31.9 million patents, some 17.9 million of which were granted. We match company name and address to Orbis company name and address. We find that of the 122,246 firms in our sample, 18,709 have applied for at least one patent.

B.8 Mergers and Acquisitions

To get the ownership history of the firms in our sample, we use the Thomson SDC Platinum M&A dataset. We focus on completed mergers, acquisitions, and acquisitions of majority interest. There are 302,151 such transactions between the years 1971 and 2022. We match

[21] The exact query used was: psn_sector in ('COMPANY', 'COMPANY GOV NON-PROFIT', 'COMPANY GOV NON-PROFIT UNIVERSITY', 'COMPANY HOSPITAL', 'COMPANY UNIVERSITY', 'GOV NON-PROFIT', 'GOV NON-PROFIT HOSPITAL', 'GOV NON-PROFIT UNIVERSITY'.

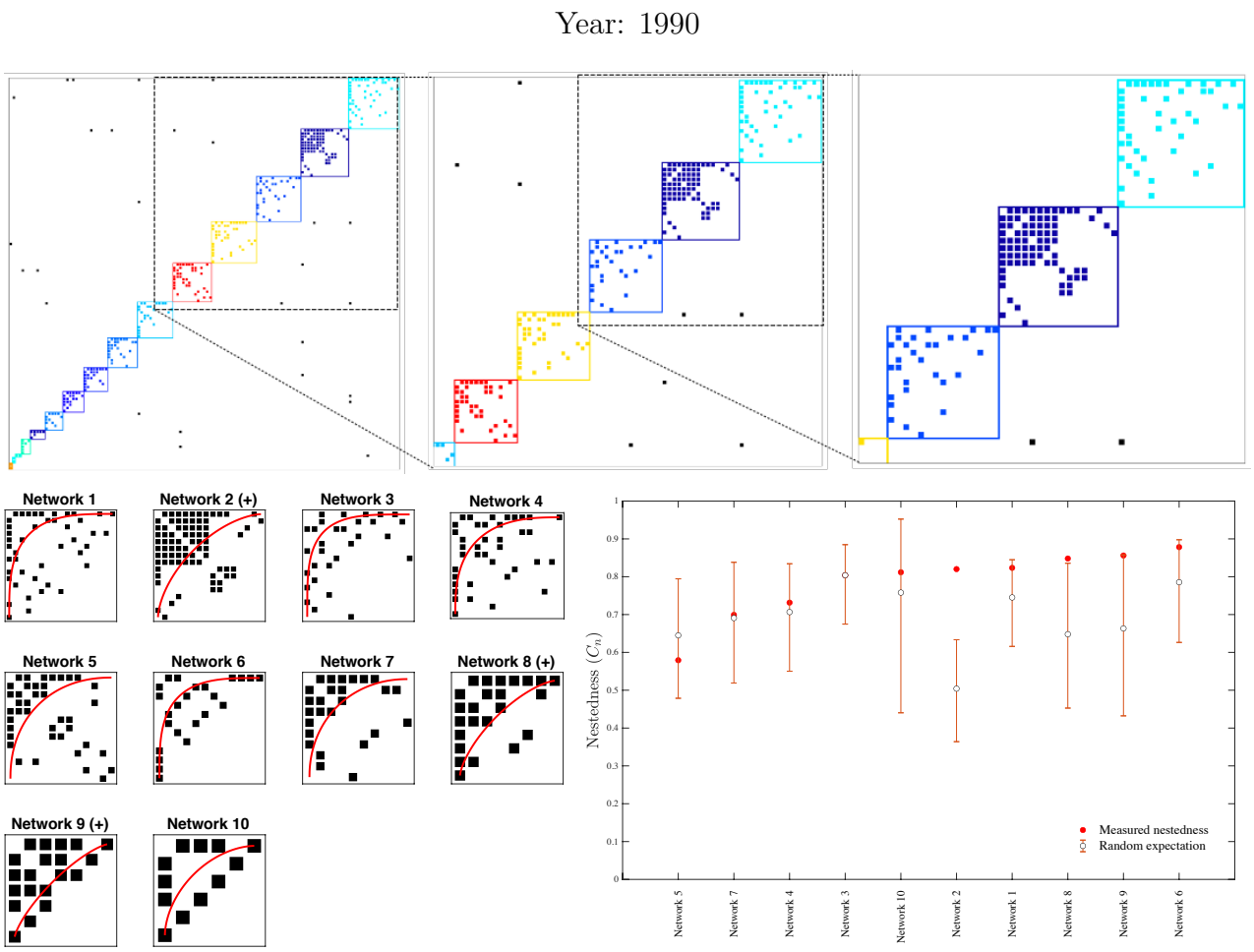


Figure B.13. The top left panel shows the adjacency matrix of the largest connected component of the R&D network in the year 1990 where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices of each module (subnetwork) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Almeida-Neto et al., 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n).

Year: 2000

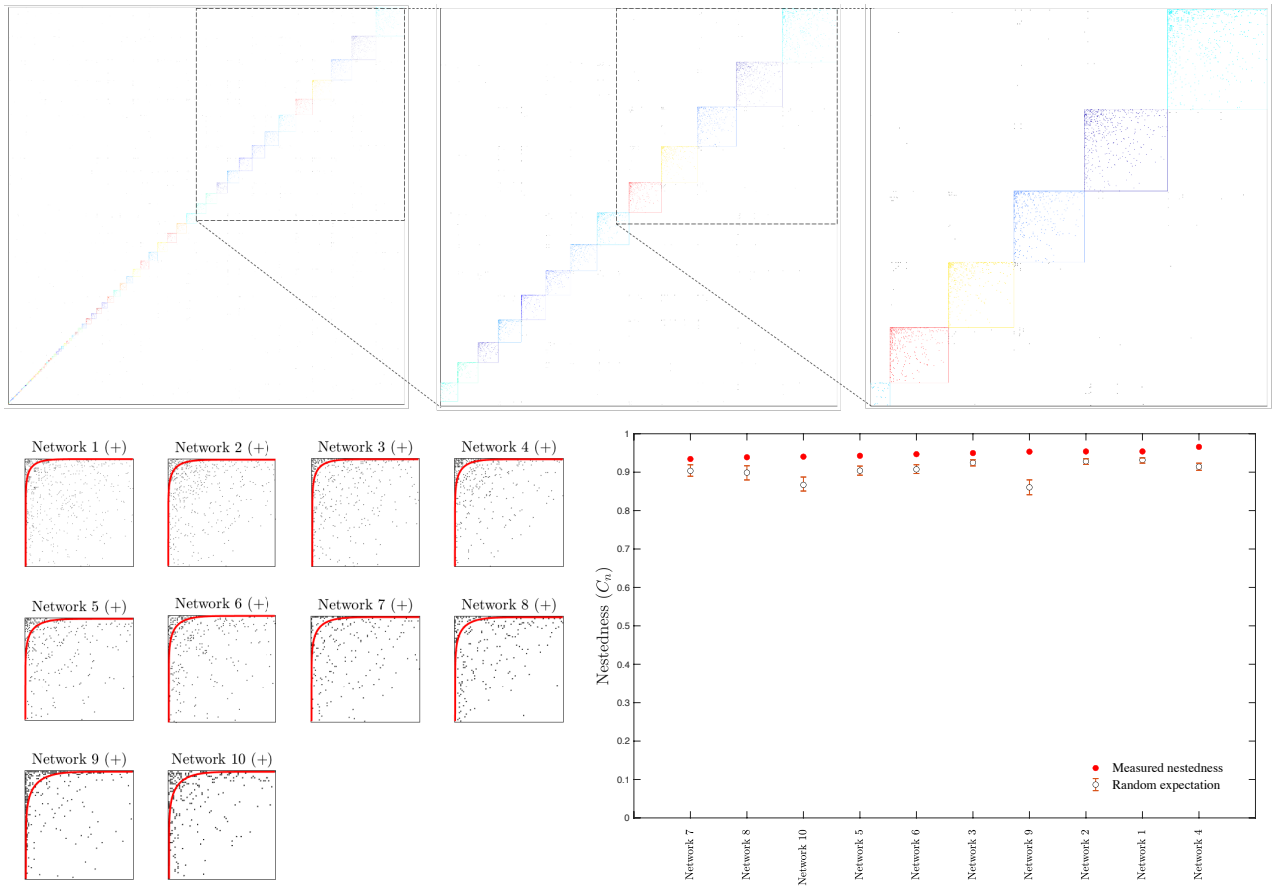


Figure B.14. The top left panel shows the adjacency matrix of the largest connected component of the R&D network in the year 2000 where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices of each module (subnetwork) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Almeida-Neto et al., 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n).

Year: 2010

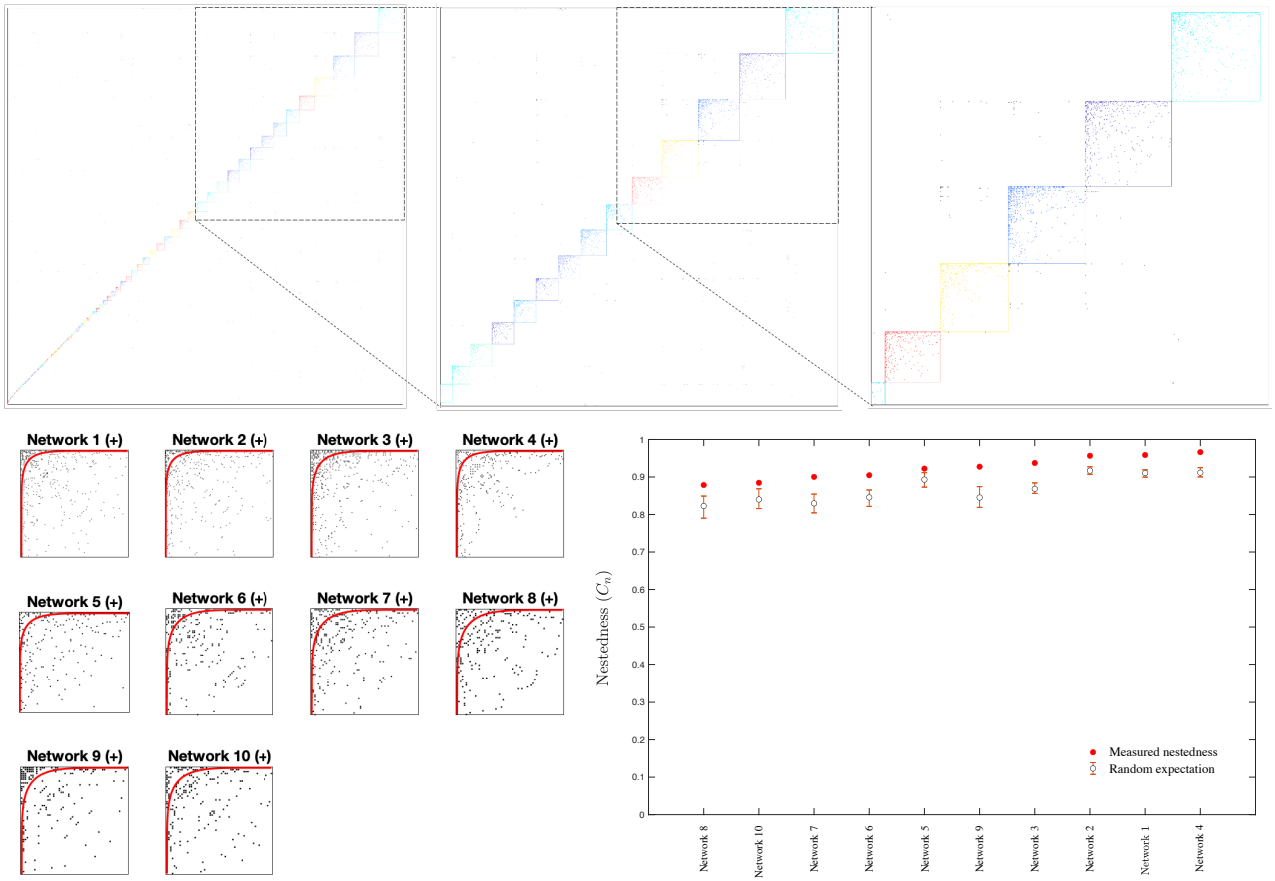


Figure B.15. The top left panel shows the adjacency matrix of the largest connected component of the R&D network in the year 2010 where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices of each module (subnetwork) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Almeida-Neto et al., 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n).

Year: 2020

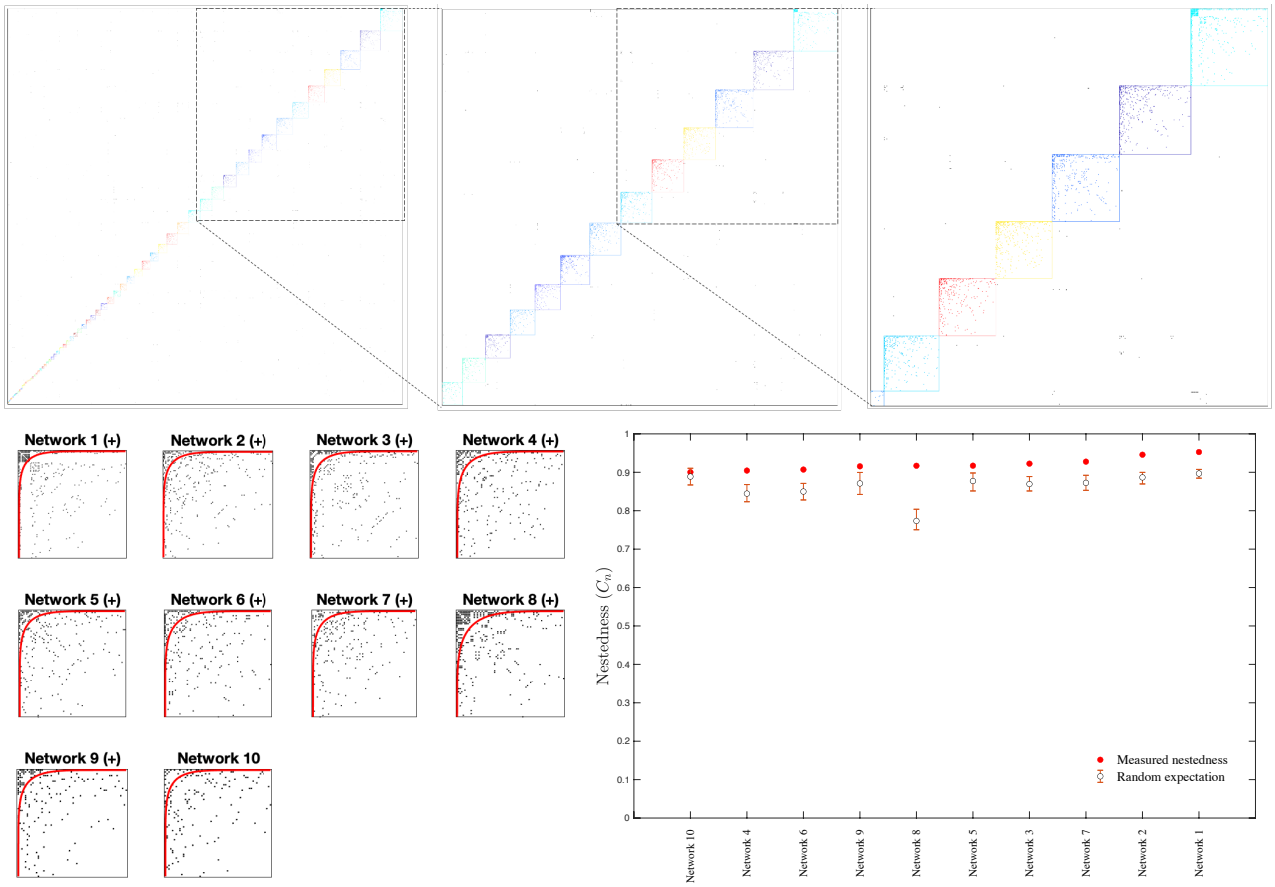


Figure B.16. The top left panel shows the adjacency matrix of the largest connected component of the R&D network in the year 2020 (see Fig. 2) where the modules identified are indicated with different colors. The top middle and top right panels zoom into the largest modules. The bottom left panel shows the adjacency matrices of each module (subnetwork) separately (showing only the first 10 modules with the highest modularity and size), where a (+) indicates that the matrix is statistically significantly nested at the 5% level (Almeida-Neto et al., 2008). The solid red line indicates the isocline (that is, a curve that divides the ones from the zeros of a perfectly nested matrix of the same size and connectivity). The bottom right panel shows the nestedness coefficients, C_n, of these 10 modules (sorted by C_n).

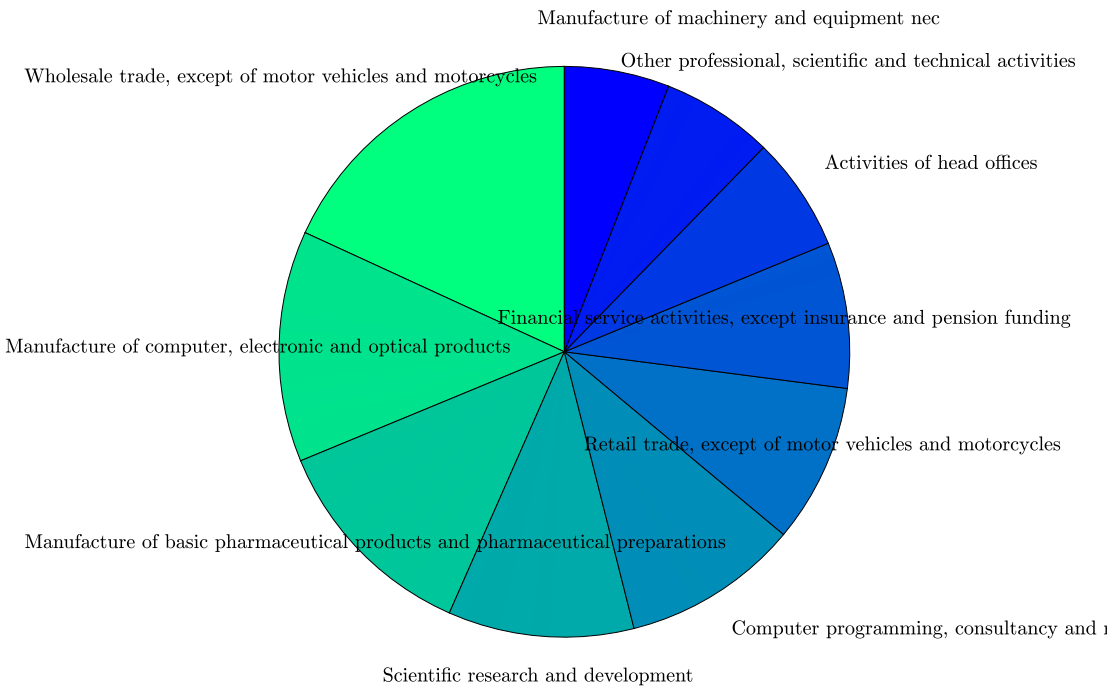


Figure B.17. The shares of the ten largest sectors at the 2-digit NACE Rev. 2 levels. See also Table 11, respectively. NACE refers to the current Statistical Classification of Economic Activities in the European Community (revision 2).

acquiror and target names, as well as the names of their ultimate parent (if applicable) to Orbis. See Figure B.18 for an illustration.

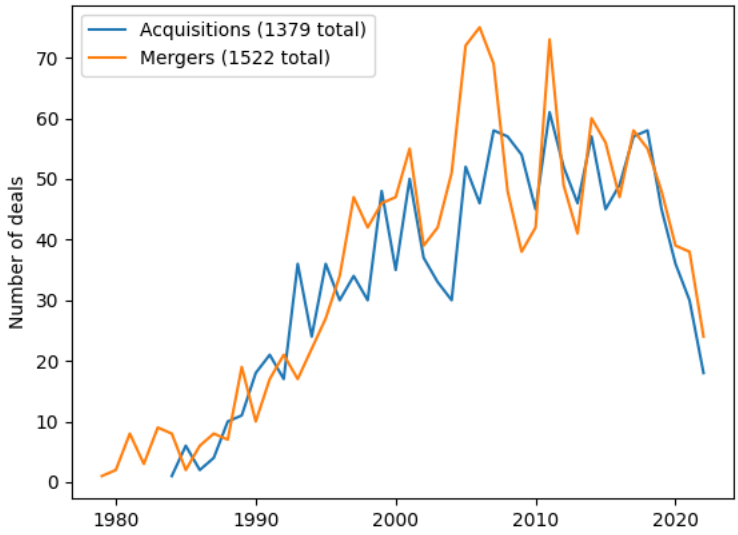


Figure B.18. Completed mergers, acquisitions, and acquisitions of majority interest, per year in the Thomson SDC Platinum M&A dataset of firms that participated also in R&D collaborations.

B.9 Compustat Data, Interpolation and Imputation

Figures B.19 and B.20 show various descriptive statistics for the balance sheet data from the Compustat North America Fundamentals Annual and Compustat Global databases merged with the firms participating in R&D collaborations. All variables reported in local currencies in the Compustat Global database have been translated into US dollars. Dollar values have then been deflated using the Consumer Price Index (CPI) from the US Bureau of Labor Statistics (`https://www.bls.gov/cpi/`) with 1983 as the base year with an index of 100. Output has been computed using the Producer Price Index (PPI) of the US Bureau of Labor Statistics (`https://www.bls.gov/ppi/`) matched to the year and NAICS code reported in the Compsutat sales item. Value added has been computed from Compustat data following the procedure in Liu (2018) and İmrohoroğlu and Tüzel (2014).[22] The linear interpolation uses multiple observations of the same firm over different years, when there are gaps between the observations across years for the same firm (using Stata's "ipolate" function). The imputation

[22]Value added can be computed from Compustat data using the following steps: $VALUE_ADDED = SALES - MATERIALS, MATERIALS = TOTAL_EXPENSES - LABOR_EXPENSES$, where $TOTAL_EXPENSES = SALES - OIBDP$, $LABOR_EXPENSES = EMP * WAGE$. $WAGE$ is the average wage from the US Social Security Administration, and $OIBDP$ is the Compustat item "Operating Income Before Depreciation and Amortization". Putting the above together yields $VALUE_ADDED = SALES - MATERIALS = SALES - (TOTAL_EXPENSES - LABOR_EXPENSES) = SALES - ((SALES - OIBDP) - LABOR_EXPENSES) = SALES - ((SALES - OIBDP) - EMP * WAGE) = OIBDP + EMP * WAGE$.

method is based on the predictive mean matching imputation method (using Stata's "mi impute pmm" function) using industry, location and covariates (sales, R&D, employment, output, etc.) as input (Rubin, 1986).

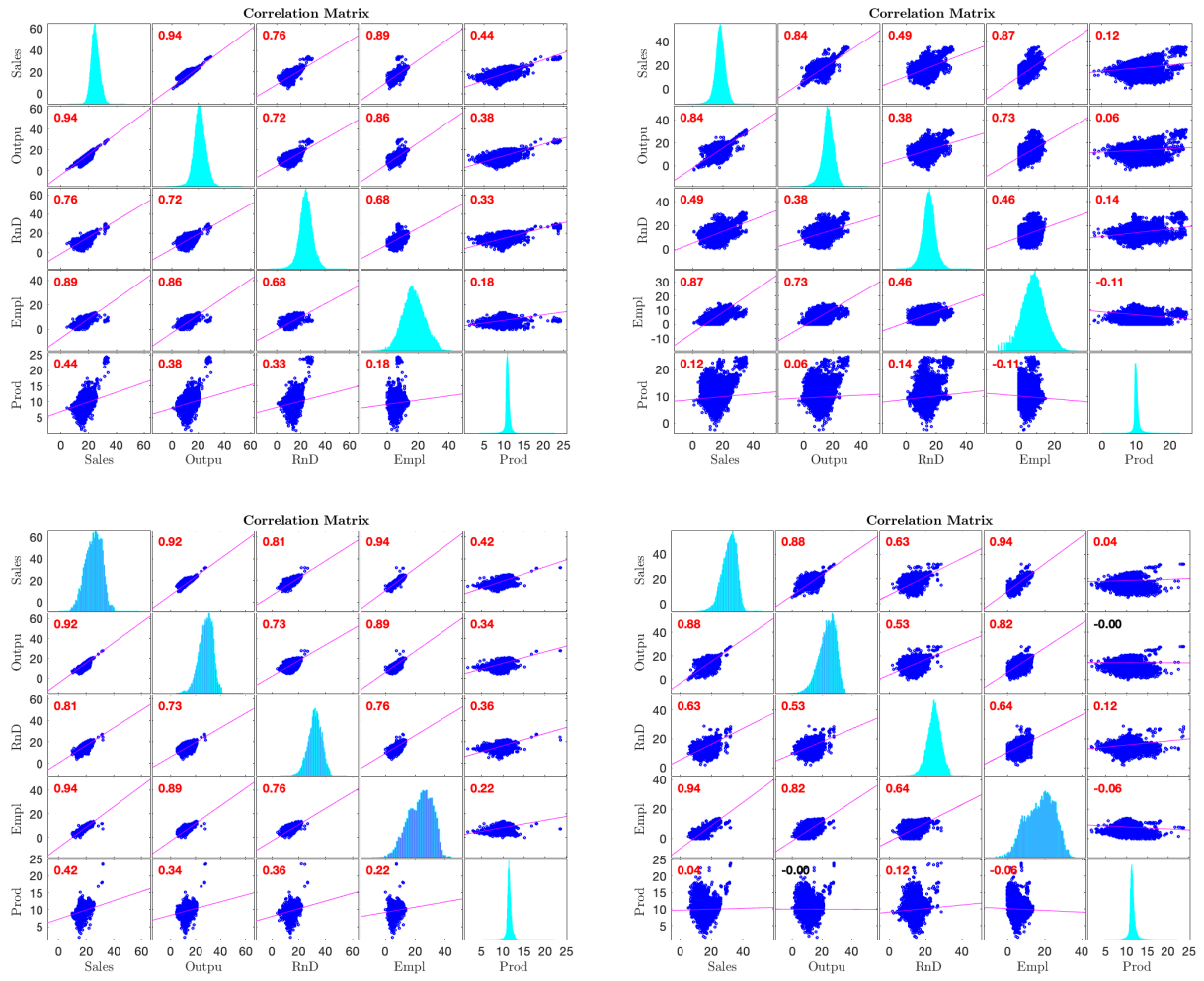


Figure B.19. The top left panel shows the correlation matrix of sales, output, R&D expenditures, employment and labor productivity (value added divided by the number of employees) for the merged Compustat Fundamentals Annual and Global data pooled across all years. The top right panel shows the corresponding correlation matrix for the linearly interpolated (over time) and imputed (for missing values) data of the merged Compustat North America Fundamentals Annual and Global databases. The bottom left panel shows the correlation matrix for the firms that participate in R&D collaborations in the merged Compustat North America Fundamentals Annual and Global data. The bottom right panel shows the correlation matrix for the linearly interpolated (over time) and imputed (for missing values) data for the R&D collaborating firms in the merged Compustat North America Fundamentals Annual and Global databases.

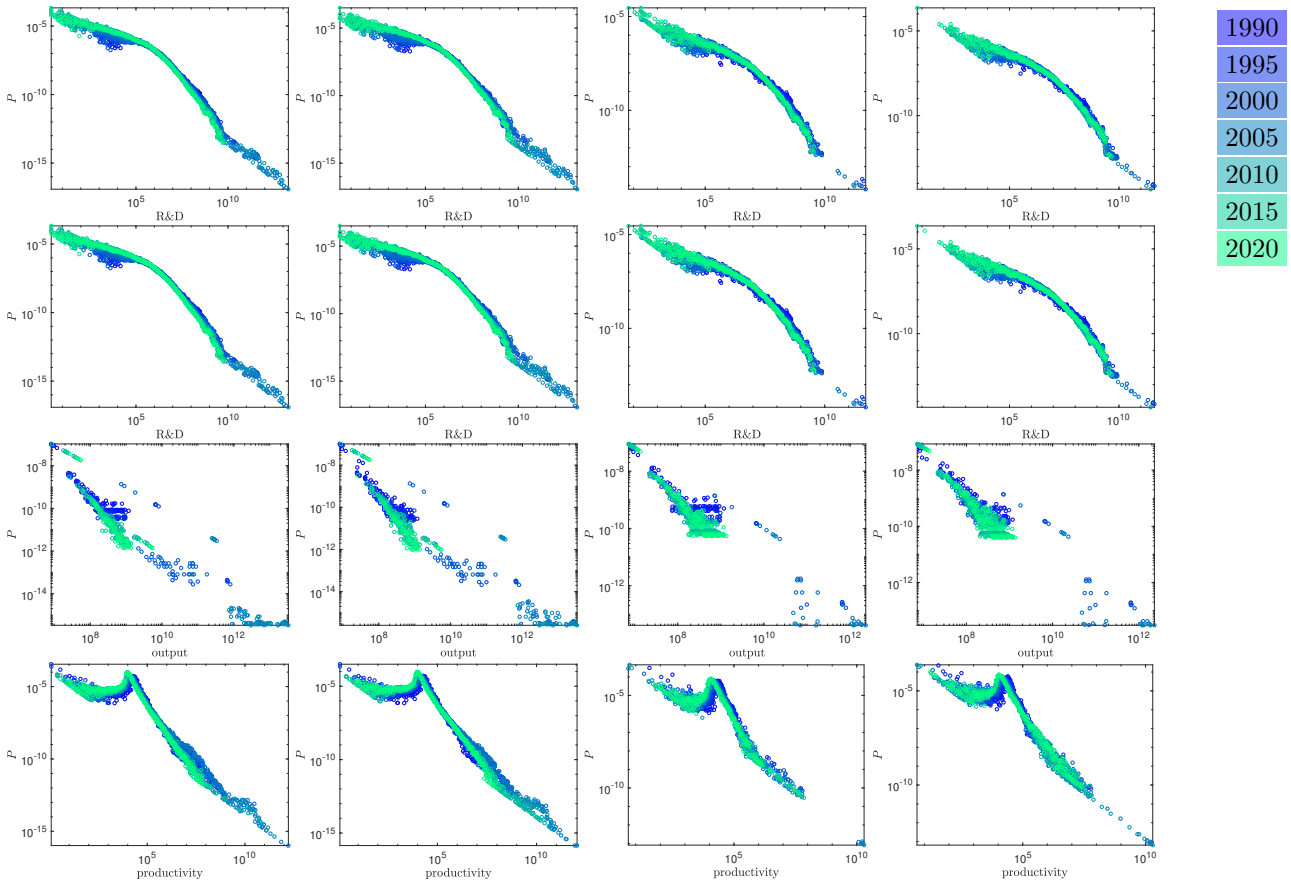


Figure B.20. The panels in the top row show the distribution of sales for different years. The panels in the second row the distribution os R&D expenditures, the third row the distribution of output and the bottom panels the distribution of labor productivity (value added divided by the number of employees). The first column shows the merged Compustat Fundamentals Annual and Global data. The second column shows the linearly interpolated (over time) and imputed (for missing values) data of the merged Compustat North America Fundamentals Annual and Global databases. The third column shows the firms that participate in R&D collaborations in the merged Compustat North America Fundamentals Annual and Global data. The last column shows the linearly interpolated (over time) and imputed (for missing values) data for the R&D collaborating firms in the merged Compustat North America Fundamentals Annual and Global databases.

C Cost Reducing R&D Collaborations

In the following we derive the firm's profit function in Equations (1) or (4), respectively, and the corresponding best response function in Equation (3) from a Cournot competition model (Appendix C.1) or a Bertrand competition model (Appendix C.3) with cost reducing R&D collaborations.[23]

C.1 Cournot Competition

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

$$c_i = \bar{c}_i - \alpha y_i - \beta \sum_{j=1}^{n} a_{ij} y_j. \tag{22}$$

The indicator variables $a_{ij} \in \{0, 1\}$ in Equation (22) indicate wether firms i and j are collaborating (or not), and can be represented by the symmetric adjacency matrix $\mathbf{A} = (a_{ij})_{1 \leq i,j \leq n}$ (with zero diagonal, $a_{ii} = 0$). The parameter $\alpha \geq 0$ measures the relative cost reduction due to a firms' 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. We further allow for ex ante heterogeneity (observed and unobserved) among firms in the cost parameter $\bar{c}_i \geq 0$ expressing their different technological and organizational capabilities related to heterogeneous firm productivities (Banerjee and Duflo, 2005; Blundell et al., 1995).

We further assume that firms incur a direct cost γy_i^2, with $\gamma \geq 0$, for their R&D effort and a fixed cost $\zeta_{ij} \geq 0$ for the R&D collaboration between firms i and j. The profit of firm i is then given by

$$\pi_i = (p_i - c_i) x_i - \gamma y_i^2 - \sum_{j=1}^{n} a_{ij} \zeta_{ij}, \tag{23}$$

where $x_i \geq 0$ is the output level and ζ_{ij} is the cost of collaboration between firms i and j. Inserting marginal costs from Equation (22) gives

$$\pi_i = p_i x_i - \bar{c}_i x_i + \alpha x_i y_i + \beta x_i \sum_{j=1}^{n} a_{ij} y_j - \gamma y_i^2 - \sum_{j=1}^{n} a_{ij} \zeta_{ij}.$$

[23] See also Goyal and Moraga-Gonzalez (2001), König et al. (2019) and Hsieh et al. (2024).

Next we consider the demand for goods produced by firm i. A representative consumer in market $\mathcal{M}_k \subseteq \mathcal{N}$, $k = 1, \ldots, K$ maximizes (Singh and Vives, 1984)

$$U_k = I_k + \upsilon \sum_{i \in \mathcal{M}_k} x_i - \frac{\psi}{2} \sum_{i \in \mathcal{M}_k} x_i^2 - \frac{\lambda}{2} \sum_{\substack{i,j \in \mathcal{M}_k, \\ j \neq i}} x_i x_j, \tag{24}$$

with the budget constraint $I_k + \sum_{i \in \mathcal{M}_k} p_i x_i \leq E_k$, endowment E_k and a numeraire good I_k. The parameter $\upsilon > 0$ captures the total size of the market, whereas $\lambda \in (0, 1]$, measures the degree of substitutability between products. In particular, $\lambda = 1$ depicts a market of perfect substitutable goods, while $\lambda \to 0$ represents the case of almost independent markets. The parameter λ is therefore a measure of the degree of competition between firms. We assume that $\psi > \lambda > 0$ to ensure that the utility function is concave. The budget constraint is binding and the utility maximization of the representative consumer gives the inverse demand function for firm $i \in \mathcal{M}_k$:[24]

$$p_i = \upsilon - \psi x_i - \lambda \sum_{j \in \mathcal{M}_k, j \neq i} x_j. \tag{25}$$

Inserting the marginal cost from Equation (22) and inverse demand from Equation (25) into profits yields

$$\pi_i = \upsilon x_i - \psi x_i^2 - \lambda \sum_{j \neq i}^{n} b_{ij} x_j x_i - \bar{c}_i x_i + \alpha x_i y_i + \beta x_i \sum_{j=1}^{n} a_{ij} y_j - \gamma y_i^2 - \sum_{j=1}^{n} a_{ij} \zeta_{ij}, \tag{26}$$

where $b_{ij} \in \{0, 1\}$ is an indicator variable for whether firms i and j are competitors (or not), i.e. $i, j \in \mathcal{M}_k$ for some k. This can be represented by the symmetric block diagonal competition matrix $\mathbf{B} = (b_{ij})_{1 \leq i,j \leq n}$. The optimal R&D effort level follows from the first-order condition

$$\frac{\partial \pi_i}{\partial y_i} = \alpha x_i - 2\gamma y_i = 0.$$

Solving for y_i delivers[25]

$$y_i = \frac{\alpha}{2\gamma} x_i. \tag{27}$$

This equation can be viewed as reflecting learning-by-doing effects of production on R&D. It is consistent with various empirical studies which have found that the R&D effort is correlated with firm size (Cohen and Klepper, 1996a,b).

[24] With the budget constraint $I_k = E_k - \sum_{i \in \mathcal{M}_k} p_i x_i$ in the consumer's utility in Equation (24), the FOC for $i \in \mathcal{M}_k$ is given by $\frac{\partial U_k}{\partial x_i} = -p_i + \upsilon - \psi x_i - \lambda \sum_{j \in \mathcal{M}_k, j \neq i}^{n} x_j = 0$, from which we obtain Equation (25).

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

Similarly, the first-order condition w.r.t. to output, x_i, is given by

$$\frac{\partial \pi_i}{\partial x_i} = \upsilon - 2\psi x_i - \lambda \sum_{j \neq i}^{n} b_{ij} x_j - \bar{c}_i + \alpha y_i + \beta \sum_{j=1}^{n} a_{ij} y_j = 0.$$

This gives

$$x_i = \frac{\upsilon - \bar{c}_i}{2\psi} - \frac{\lambda}{2\psi} \sum_{j \neq i}^{n} b_{ij} x_j + \frac{\alpha}{2\psi} y_i + \frac{\beta}{2\psi} \sum_{j=1}^{n} a_{ij} y_j.$$

Further, inserting Equation (27) yields

$$x_i = \frac{\upsilon - \bar{c}_i}{2\psi} - \frac{\lambda}{2\psi} \sum_{j \neq i}^{n} b_{ij} x_j + \frac{\alpha^2}{4\psi\gamma} x_i + \frac{\alpha\beta}{4\psi\gamma} \sum_{j=1}^{n} a_{ij} x_j.$$

Solving for x_i then gives

$$x_i = \frac{2\gamma(\upsilon - \bar{c}_i)}{4\psi\gamma - \alpha^2} - \frac{2\lambda\gamma}{4\psi\gamma - \alpha^2} \sum_{j \neq i}^{n} b_{ij} x_j + \frac{\alpha\beta}{4\psi\gamma - \alpha^2} \sum_{j=1}^{n} a_{ij} x_j. \tag{28}$$

Further, with Equation (27) the profit function in Equation (34) of firm i can be written as

$$\pi_i = (\upsilon - \bar{c}_i) x_i - \psi x_i^2 - \lambda \sum_{j \neq i}^{n} b_{ij} x_i x_j + \frac{\alpha^2}{2\gamma} x_i^2 + \frac{\alpha\beta}{2\gamma} \sum_{j=1}^{n} a_{ij} x_i x_j - \frac{\alpha^2}{4\gamma} x_i^2 - \sum_{j=1}^{n} a_{ij} \zeta_{ij}.$$

This is

$$\pi_i = (\upsilon - \bar{c}_i) x_i - (\psi - \frac{\alpha^2}{4\gamma}) x_i^2 - \lambda \sum_{j \neq i}^{n} b_{ij} x_i x_j + \frac{\alpha\beta}{2\gamma} \sum_{j=1}^{n} a_{ij} x_i x_j - \sum_{j=1}^{n} a_{ij} \zeta_{ij}. \tag{29}$$

If we denote by $\eta_i \equiv \upsilon - \bar{c}_i$, $\nu \equiv \psi - \frac{\alpha^2}{4\gamma}$ and $\rho \equiv \frac{\alpha\beta}{2\gamma}$ then we can write Equation (29) more compactly as follows

$$\pi_i = \eta_i x_i - \nu x_i^2 - \lambda x_i \sum_{j \neq i}^{n} x_j + \rho \sum_{j=1}^{n} a_{ij} x_i x_j - \sum_{j=1}^{n} a_{ij} \zeta_{ij}. \tag{30}$$

This gives Equation (1). Similarly, Equation (28) can then be written as

$$x_i = \frac{\eta_i}{2\nu} - \frac{\lambda}{2\nu} \sum_{j \neq i}^{n} b_{ij} x_j + \frac{\rho}{2\nu} \sum_{j=1}^{n} a_{ij} x_j, \tag{31}$$

which is Equation (3) in Section 3 if we normalize $\nu = 1/2$. From Equations (31) and (30) it

then immediately follows that

$$\pi_i = \nu x_i^2 - \sum_{j=1}^{n} a_{ij}\zeta_{ij},$$

which gives Equation (4) in Section 3 if we normalize $\nu = 1/2$. Producer surplus is given by firms' aggregate profits $\Pi = \sum_{i=1}^{n} \pi_i$.

Using the budget constraint $I_k + \sum_{i\in\mathcal{M}_k} p_i x_i = E_k$ and inserting Equation (25) into the utility (24) of the consumer gives:

$$\begin{aligned} U_k &= E_k - \sum_{i\in\mathcal{M}_k} p_i x_i + \upsilon \sum_{i\in\mathcal{M}_k} x_i - \frac{\psi}{2}\sum_{i\in\mathcal{M}_k} x_i^2 - \frac{\lambda}{2}\sum_{i,j\in\mathcal{M}_k, j\neq i} x_i x_j \\ &= E_k - \sum_{i\in\mathcal{M}_k}\left(\upsilon - \psi x_i - \lambda\sum_{j\neq i}^{n} b_{ij}x_j\right)x_i + \upsilon\sum_{i\in\mathcal{M}_k} x_i - \frac{\psi}{2}\sum_{i\in\mathcal{M}_k} x_i^2 - \frac{\lambda}{2}\sum_{i,j\in\mathcal{M}_k, j\neq i} x_i x_j \\ &= E_k + \frac{\psi}{2}\sum_{i\in\mathcal{M}_k} x_i^2 + \frac{\lambda}{2}\sum_{i\in\mathcal{M}_k}\sum_{j\neq i}^{n} b_{ij}x_i x_j. \end{aligned}$$

Summation over all markets gives

$$U = \sum_{k=1}^{K} U_k = \sum_{k=1}^{K} E_k + \frac{\psi}{2}\sum_{k=1}^{K}\sum_{i\in\mathcal{M}_k} x_i^2 + \frac{\lambda}{2}\sum_{k=1}^{K}\sum_{i\in\mathcal{M}_k}\sum_{j\neq i}^{n} b_{ij}x_i x_j = E + \frac{\psi}{2}\sum_{i=1}^{n} x_i^2 + \frac{\lambda}{2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} b_{ij}x_i x_j, \tag{32}$$

where we have denoted by $E = \sum_{k=1}^{K} E_k$, and U in Eq. (32) defines consumer surplus. Note that in the special case of non-substitutable goods, when $\lambda \to 0$, we obtain $U = E + \frac{\psi}{2}\sum_{i=1}^{n} x_i^2$, while in the case of perfectly substitutable goods, when $\lambda \to 1$, we get $U = E + \frac{1+\psi}{2}\sum_{i,j=1}^{n} b_{ij}x_i x_j$.

C.2 Herfindahl Index

The Herfindahl-Hirschman industry concentration index is defined as $H_k = \sum_{i\in\mathcal{M}_k} s_i^2$, where the market share of firm i in market $\mathcal{M}_k$ is given by $s_i = \frac{x_i}{\sum_{j\in\mathcal{M}_k}^{n} x_j}$ (cf. e.g. Tirole, 1988; Hirschman, 1964). When there is only a single market, i.e. $\mathcal{M}_k = \mathcal{N}$, we can write

$$H = \sum_{i=1}^{n}\left(\frac{x_i}{\sum_{j=1}^{n} x_j}\right)^2 = \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x}\|_1^2}, \tag{33}$$

With $\mathbf{x} = (\mathbf{I}_n + \lambda\mathbf{B} - \rho\mathbf{A})^{-1}\boldsymbol{\eta} = \mathbf{b}_{\boldsymbol{\eta}}(G, \lambda, \rho) = \mathbf{M}(G, \lambda, \rho)\boldsymbol{\eta}$ in the Nash equilibrium,[26] we can write the Herfindahl index of Equation (33) as follows

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

which is the inverse of the *participation ratio* $\gamma(\cdot)$. The participation ratio $\gamma(\mathbf{x})$ measures the number of elements of $\mathbf{x}$ which are dominant. We have that $1 \leq \gamma(\mathbf{x}) \leq n$, where a value of $\gamma(\mathbf{x}) = n$ corresponds to a fully homogeneous case, while $\gamma(\mathbf{x}) = 1$ corresponds to a fully concentrated case (note that, if all x_i are identical then $\gamma(\mathbf{x}) = n$, while if one x_i is much larger than all others we have $\gamma(\mathbf{x}) = 1$). Moreover, $\gamma(\mathbf{x})$ is scale invariant, that is, $\gamma(\alpha\mathbf{x}) = \gamma(\mathbf{x})$ for any $\alpha \in \mathbb{R}_+$.[27] The participation ratio $\gamma(\mathbf{x})$ is further related to the *coefficient of variation* $\mathrm{CV}(\mathbf{x}) = \frac{\sigma(\mathbf{x})}{\mu(\mathbf{x})}$, where $\sigma(\mathbf{x})$ is the standard deviation and $\mu(\mathbf{x})$ the mean of the components of $\mathbf{x}$, via the relationship $\mathrm{CV}(\mathbf{x})^2 = \frac{n}{\gamma(\mathbf{x})} - 1$. This implies that

$$H(G) = \frac{\boldsymbol{\eta}^\top \mathbf{M}(G, \lambda, \rho)^2 \boldsymbol{\eta}}{(\boldsymbol{\eta}^\top \mathbf{M}(G, \lambda, \rho)\boldsymbol{\eta})^2} = \frac{\mathrm{CV}(\mathbf{b})^2 + 1}{n} \sim \frac{\mathrm{CV}(\mathbf{b})^2}{n}.$$

Hence, the Herfindhal index is maximized for the graph G with the highest coefficient of variation in the components of the Bonacich centrality $\mathbf{b}_{\boldsymbol{\eta}}(G, \lambda, \rho)$.

C.3 Bertrand Competition

In the case of price setting firms we obtain from the profit function

$$\pi_i = (p_i - c_i)x_i - \gamma y_i^2 - \sum_{j=1}^n a_{ij}\zeta_{ij}, \tag{34}$$

the following first order condition for firm i:

$$\frac{\partial \pi_i}{\partial p_i} = (p_i - c_i)\frac{\partial x_i}{\partial p_i} + x_i = 0. \tag{35}$$

[26] In the case of a single market the matrix $\mathbf{B}$ is an $n \times n$ matrix of ones.

[27] The inverse participation ratio (IPR) is also used in the network science literature (cf. e.g. Martin et al., 2014; Alves et al., 2022). An IPR near zero indicates a minimal localization effect, meaning there is little dominance by individual nodes and centrality is evenly spread across all nodes. In contrast, an IPR close to one suggests that centrality is highly concentrated in a small number of nodes, with the system being dominated by a few hub nodes.

The inverse demand function is given by

$$p_i = \upsilon - \psi x_i - \lambda \sum_{j \neq i}^{n} b_{ij} x_j,$$

from which it can be shown that

$$x_i = \frac{\upsilon}{\lambda(m_i - 1) + \psi} - \frac{\lambda(m_i - 2) + \psi}{(\lambda(m_i - 1) + \psi)(\psi - \lambda)} p_i - \frac{\lambda}{(\lambda(m_i - 1) + \psi)(\psi - \lambda)} \sum_{j=1, j \neq i}^{n} b_{ij} p_j,$$

where $m_i = 1 + \sum_{j=1, j \neq i}^{n} b_{ij}$ counts one plus the number of firms competing with firm i (i.e. the total number of firms in the market in which firm i is operating). It then follows that

$$\frac{\partial x_i}{\partial p_i} = -\frac{\lambda(m_i - 2) + \psi}{(\lambda(m_i - 1) + \psi)(\psi - \lambda)}. \tag{36}$$

Inserting Equation (36) into Equation (35) gives

$$x_i = -(p_i - c_i)\frac{\partial x_i}{\partial p_i} = \frac{\lambda(m_i - 2) + \psi}{(\lambda(m_i - 1) + \psi)(\psi - \lambda)}(p_i - c_i). \tag{37}$$

Further, using the fact that

$$c_i = \bar{c}_i - \alpha y_i - \beta \sum_{j=1}^{n} a_{ij} y_j,$$

and

$$p_i = \upsilon - \psi x_i - \lambda \sum_{j=1}^{n} b_{ij} x_j,$$

we get that

$$p_i - c_i = \upsilon - \bar{c}_i - \psi x_i - \lambda \sum_{j=1}^{n} b_{ij} x_j + \alpha y_i + \beta \sum_{j=1}^{n} a_{ij} y_j. \tag{38}$$

Inserting Equation (38) into (37) gives

$$x_i = \frac{\lambda(m_i - 2) + \psi}{(\lambda(m_i - 1) + \psi)(\psi - \lambda)} \left(\upsilon - \bar{c}_i - \psi x_i - \lambda \sum_{j=1}^{n} b_{ij} x_j + \alpha y_i + \beta \sum_{j=1}^{n} a_{ij} y_j \right).$$

Solving for x_i yields

$$x_i = \frac{(\lambda(m_i-2)+\psi)(\upsilon-\bar{c}_i)}{\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i} - \frac{(\lambda(m_i-2)+\psi)\lambda}{\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i}\sum_{j=1}^{n} b_{ij}x_j$$
$$+ \frac{(\lambda(m_i-2)+\psi)\alpha}{\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i}y_i + \frac{(\lambda(m_i-2)+\psi)\beta}{\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i}\sum_{j=1}^{n} a_{ij}y_j. \quad (39)$$

Further, inserting Equation (38) into the profit function of Equation (34) gives

$$\pi_i = \upsilon x_i - \psi x_i^2 - \lambda\sum_{j=1}^{n} b_{ij}x_i x_j - \bar{c}_i x_i + \alpha x_i y_i + \beta\sum_{j=1}^{n} a_{ij}x_i y_j - \gamma y_i^2 - \sum_{j=1}^{n}\zeta_{ij}.$$

The first order condition with respect to R&D effort is given by

$$\frac{\partial \pi_i}{\partial y_i} = \alpha x_i - 2\gamma y_i = 0,$$

from which we get that

$$y_i = \frac{\alpha}{2\gamma}x_i. \quad (40)$$

Inserting Equation (40) into Equation (39) gives

$$x_i = \frac{(\lambda(m_i-2)+\psi)(\upsilon-\bar{c}_i)}{\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i} - \frac{(\lambda(m_i-2)+\psi)\lambda}{\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i}\sum_{j=1}^{n} b_{ij}x_j$$
$$+ \frac{(\lambda(m_i-2)+\psi)\alpha^2}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i)}x_i + \frac{(\lambda(m_i-2)+\psi)\alpha\beta}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i)}\sum_{j=1}^{n} a_{ij}x_j. \quad (41)$$

This can be written as

$$x_i = \frac{2\gamma(\lambda(m_i-2)+\psi)(\upsilon-\bar{c}_i)}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i)-\alpha^2(\psi+\lambda(m_i-2))}$$
$$- \frac{2\gamma(\lambda(m_i-2)+\psi)\lambda}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i)-\alpha^2(\psi+\lambda(m_i-2))}\sum_{j=1}^{n} b_{ij}x_j$$
$$+ \frac{(\lambda(m_i-2)+\psi)\alpha\beta}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)m_i)-\alpha^2(\psi+\lambda(m_i-2))}\sum_{j=1}^{n} a_{ij}x_j. \quad (42)$$

In the case of a single market, $m_i = n$, Equation (42) can be written as

$$\begin{aligned} x_i = & \frac{2\gamma(\lambda(n-2)+\psi)(\upsilon-\bar{c}_i)}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)n)-\alpha^2(\psi+\lambda(n-2))} \\ & - \frac{2\gamma(\lambda(n-2)+\psi)\lambda}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)n)-\alpha^2(\psi+\lambda(n-2))}\sum_{j=1}^{n} b_{ij}x_j \\ & + \frac{(\lambda(n-2)+\psi)\alpha\beta}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)n)-\alpha^2(\psi+\lambda(n-2))}\sum_{j=1}^{n} a_{ij}x_j. \end{aligned} \tag{43}$$

If we further denote by $\tilde{\eta}_i \equiv (\upsilon - \bar{c}_i)C$, $\tilde{\lambda} \equiv \lambda C$ and $\tilde{\rho} \equiv \frac{\alpha\beta}{2\gamma}C$, where

$$C = \frac{2\gamma(\lambda(n-2)+\psi)}{2\gamma(\lambda^2-4\lambda\psi+2\psi^2-\lambda(\lambda-2\psi)n)-\alpha^2(\psi+\lambda(n-2))}, \tag{44}$$

we can write Equation (43) compactly as follows

$$x_i = \tilde{\eta}_i - \tilde{\lambda}\sum_{j=1}^{n} b_{ij}x_j + \tilde{\rho}\sum_{j=1}^{n} a_{ij}x_j,$$

which has the same functional form as Equation (3) in Section 3. Next, note that after inserting the optimal R&D effort from Equation (40), the profit function of the firm becomes

$$\pi_i = (p_i - \bar{c}_i)x_i - \frac{\alpha^2}{4\gamma}x_i^2 - \sum_{j=1}^{n}\zeta_{ij}. \tag{45}$$

Similarly, inserting Equation (40) into Equation (38), we get that

$$p_i - c_i = \upsilon - \psi x_i - \lambda\sum_{j=1}^{n} b_{ij}x_j - \bar{c}_i + \frac{\alpha^2}{2\gamma}x_i + \frac{\alpha\beta}{2\gamma}\sum_{j=1}^{n} a_{ij}x_j. \tag{46}$$

Inserting Equation (46) into Equation (45) gives

$$\pi_i = \left(\upsilon - \bar{c}_i - \lambda\sum_{j=1}^{n} b_{ij}x_j + \frac{\alpha\beta}{2\gamma}\sum_{j=1}^{n} a_{ij}x_j\right)x_i - \left(\psi - \frac{\alpha^2}{2\gamma}\right)x_i^2 - \sum_{j=1}^{n}\zeta_{ij}.$$

Using the definition in Equation (44) this can be written as

$$\pi_i = \left(\frac{1}{C} - \left(\psi - \frac{\alpha^2}{2\gamma}\right)\right)x_i^2 - \sum_{j=1}^{n}\zeta_{ij}.$$

which gives the same functional form as in Equation (4) in Section 3.

D Matrix Perturbation

The following lemma is useful to compute $\mathbf{y}(G) = (\mathbf{I}_n + \lambda\mathbf{B} - \rho\mathbf{A})^{-1}\boldsymbol{\eta}$ resulting from the addition or removal of a link ij in the network G in a computationally efficient way.

Lemma 1. *Let $\mathbf{e}_i$ and $\mathbf{e}_j$ be the i-th and j-th unit basis vectors. Then the matrix $\mathbf{e}_i\mathbf{e}_j^\top$ has a one in the (i,j)-th position and zeros everywhere else. Adding an entry ij and ji with weight α to the matrix $\mathbf{A}$ can be written as $\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top$. Then*

$$(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top)^{-1} = \mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i} - \alpha\frac{\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)}{1+\alpha\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j}, \tag{47}$$

assuming that the corresponding matrix inverses exist.

Proof of Lemma 1. The Sherman–Morrison formula states that (cf. 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 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

$$\mathbf{B}^{-1} = (\mathbf{C} + \alpha\mathbf{e}_j\mathbf{e}_i^\top)^{-1} = \mathbf{C}^{-1} - \alpha\frac{\mathbf{C}^{-1}\mathbf{e}_j\mathbf{e}_i^\top\mathbf{C}^{-1}}{1+\alpha\mathbf{e}_i^\top\mathbf{C}^{-1}\mathbf{e}_j}, \tag{48}$$

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 (48) yields

$$\mathbf{B}^{-1} = \mathbf{A}^{-1} - \alpha \frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i} - \alpha \frac{\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)}{1+\alpha\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j}. \quad (49)$$

□

Lemma 2. *Let $\mathbf{e}_i$ and $\mathbf{e}_j$ be the i-th and j-th unit basis vectors. Then the matrix $\mathbf{e}_i\mathbf{e}_j^\top$ has a one in the (i,j)-th position and zeros everywhere else. Adding an entry ij and ji with weight α to the matrix $\mathbf{A}$ can be written as $\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top$. Then*

$$\det\left(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top\right) = \left(1 + \alpha\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j\right)\left(1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i\right)\det\left(\mathbf{A}\right),$$

assuming that the inverses $\mathbf{A}^{-1}$ exist.

Proof of Lemma 2. We have that $\det\left(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top\right) = \det\left(\mathbf{C} + \alpha\mathbf{e}_j\mathbf{e}_i^\top\right)$, 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)

$$\det\left(\mathbf{A} + \mathbf{u}\mathbf{v}^\top\right) = (1 + \mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u})\det(\mathbf{A}).$$

It then follows that

$$\det\left(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top\right) = \det\left(\mathbf{C} + \alpha\mathbf{e}_j\mathbf{e}_i^\top\right) = (1 + \alpha\mathbf{e}_i^\top\mathbf{C}^{-1}\mathbf{e}_j)\det\left(\mathbf{C}\right),$$

Similarly, from the matrix determinant lemma, we have that

$$\det\left(\mathbf{C}\right) = \det\left(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top\right) = (1 + \alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i)\det\left(\mathbf{A}\right).$$

Further, the Sherman–Morrison formula states that (cf. 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

$$\begin{aligned}\det\left(\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^\top + \alpha\mathbf{e}_j\mathbf{e}_i^\top\right) &= (1 + \alpha\mathbf{e}_i^\top\mathbf{C}^{-1}\mathbf{e}_j)\det\left(\mathbf{C}\right) \\ &= \left(1 + \alpha\mathbf{e}_i^\top\left(\mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^\top\mathbf{A}^{-1}}{1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i}\right)\mathbf{e}_j\right)\left(1+\alpha\mathbf{e}_j^\top\mathbf{A}^{-1}\mathbf{e}_i\right)\det\left(\mathbf{A}\right).\end{aligned}$$

□