
Oxford Martin Working Paper Series on
Economic and Technological Change



Demand-pull and technology-push: What drives the direction of technological change?

Kerstin Hötte

Working Paper No. 2021-4



Disclaimer: This is a working paper and represents research in progress. This paper represents the opinions of the authors and does not represent the position of the Oxford Martin School or other institutions or individuals. For more information on the Oxford Martin Programme on Technological and Economic Change, please visit: <https://www.oxfordmartin.ox.ac.uk/technological-economic-change/>
For more information on the Oxford Martin Programme on the Future of Work, please visit: <https://www.oxfordmartin.ox.ac.uk/future-of-work/>

Demand-pull and technology-push: What drives the direction of technological change?

An empirical network-based approach

Kerstin Hötte*

Oxford Martin School, University of Oxford, 34 Broad Street, Oxford OX1 3BD, UK
Faculty of Business Administration and Economics, University of Bielefeld, 33615, Germany

April 10, 2021

This paper was revised December 17, 2021.

Demand-pull and technology-push are linked to an empirical two-layer network-based on coupled cross-industrial input-output (IO) and patent citation links among 155 4-digit (NAICS) US-industries in 1976-2006 to study the evolution of industry hierarchies and link formation.

Both layers co-evolve, but differently: The patent network became denser and increasingly skewed, while market hierarchies are balanced and sluggish in change. Industries became more similar by patent citations, but less by IO linkages. Having similar R&D capabilities as other big industries is positively related to innovation and growth, but relying on the same market inputs is unfavorable but may incite industries to explore other technological pathways. A tentative interpretation is the non-rivalry of intangible knowledge. This may strengthen existing R&D trajectories. Growth in the market is constrained by competition and market pressure may trigger a re-direction in both layers. This work is limited by its reliance on endogenously evolving classifications.

*Corresponding author: kerstin.hotte@oxfordmartin.ox.ac.uk

I want to thank Angelo Secchi and Antoine Mandel who facilitated this work at an early stage. I also want to thank Herbert Dawid, the participants of the Annual GENED meeting in 2020, the INET Complexity group meeting and the OMPTEC-FoW seminar for helpful comments. Further Gratitude is owed to Cord Wiljes for research data management support. Moreover, I gratefully acknowledge the financial support by the German Academic Foundation and Deutsch-Französische Hochschule. The technical analysis of this paper was performed while I was PhD student at the University Paris 1 in 2018.

Contents

1 Introduction	4
2 Background and literature	7
3 A model of learning and technological change	9
3.1 The economy as a two-layer network	9
3.2 Technology-push and demand-pull as drivers of change	10
3.3 Technological similarity	11
3.4 Spillovers	11
3.5 Technological and economic importance	12
3.6 Technological change	12
4 Data	13
4.1 Input-output data	13
4.2 Patent data	14
4.3 Pre-processing	15
5 Results	17
5.1 Descriptive analysis	17
5.2 Demand-pull, technology-push and the evolution of industry hierarchies	21
5.3 What determines the formation of cross-industrial links?	23
5.3.1 Link formation at the extensive margin	24
5.3.2 Link formation at the intensive margin	27
6 Discussion	29
6.1 Demand-pull or technology-push: What is driving the direction of technological change?	32
6.2 Limitations	33
7 Conclusions	34
I APPENDIX	43
A Additional statistics and results	43
A.1 Additional descriptive information	43
A.2 Regressions to explain the evolution of sector hierarchies	50
A.3 Link formation at the extensive margin	51
A.4 Link formation at the intensive margin	54
II SUPPLEMENTARY MATERIAL	58
SI.1 Detailed information on data processing	58

SI.1.1 Patent data	58
SI.1.1.1 Overview	59
SI.1.1.2 Processing steps in detail	60
SI.1.1.3 Technical and conceptual issues of the data processing	63
SI.1.2 Input-output data	63
SI.1.2.1 Overview	64
SI.1.2.2 Processing steps in detail	64
SI.1.2.3 Technical and conceptual issues	65
SI.1.3 Combined datasets	67
SI.1.4 Regression analyses	69
SI.1.4.1 Data pre-processing for regression analyses	69
SI.1.4.2 Selection of explanatory variables	69
SI.2 Additional statistics and tables	70
SI.2.1 Descriptive statistics	70
SI.2.2 Network plots	73
SI.2.3 Additional regression results	78
SI.2.3.1 Industrial hierarchies	78
SI.2.3.2 Link formation at the extensive margin	81
SI.2.3.3 Link formation at the intensive margin	85

1. Introduction

Shaping the direction of technological change is essential to cope with major challenges of the 21st century such as climate change or digitization [Rogelj et al., 2016, Steffen et al., 2018, IPCC, 2018, Brynjolfsson and McAfee, 2012] and effective policy requires an understanding of the drivers of change such as technology-push and demand-pull. Technological change is path-dependent following trajectories that build on accumulated research and production experiences [Dosi, 1982], but trajectories may change in response to technological breakthroughs (technology-push) and changing market environments (demand-pull) [Geels and Schot, 2007, Di Stefano et al., 2012, Cohen, 2010, Nemet, 2009, Pavitt, 1984, Kline and Rosenberg, 1986].

Technology-push arises when technological and scientific advances allow the development and commercialization of novelties that diffuse if they outperform incumbent solutions. Demand-pull emerges from market needs if customers ask for improvement or face technical limitations while using existing technological solutions. When inventors recognize this, they may adapt R&D efforts in response to the perceived market potential [Di Stefano et al., 2012, Kline and Rosenberg, 1986]. Technology-push and demand-pull are interdependent because R&D objectives can be demand-selected and market-needs may arise in response to innovation [Cohen, 2010, Nemet, 2009].

In this paper, I associate technology-push and demand-pull with the dynamics of a co-evolving two-layered network of patent citations and flows of goods among industries to study their impact on technological and economic change.

At the microeconomic level demand-pull and technology-push are related to tacit and codified technological knowledge [Edmondson et al., 2003, Archibugi and Coco, 2005, Cowan et al., 2000]. Codified knowledge is embodied in technologies that are traded in the market. It is related to technology-push because (theoretically) it can be adopted immediately as soon as it becomes available. Tacit knowledge builds on everyday routines and is an enabling factor to exploit the benefits of adopted technology [Edmondson et al., 2003, Dosi and Nelson, 2010, Hötte, 2021b, Arndt and Pierce, 2017].

Knowledge is non-rival, i.e. it does not diminish when it is used [Dosi and Nelson, 2010]. Empirical studies relying on patent citation and input-output (IO) networks documented spillovers from one technology user to another if the adopter is technologically related [Jaffe and De Rassenfosse, 2017, Carvalho and Voigtländer, 2014, Acemoglu et al., 2016, Antony and Grebel, 2012, Boehm et al., 2019, Cohen and Levinthal, 2000, Holly and Petrella, 2012].

Here, I use a macro-level data-based approach linking demand-pull and technology-

push to the evolution of industries in an empirical two-layered (duplex) network. The layers are inferred from coupled cross-industrial patent citation and input-output (IO) networks covering a balanced panel of more than 155 US-industries classified by 4-digit North American Industrial Classification System (NAICS) codes over the period from 1976-2006.¹ Each node in the network is an industry characterized by its output of patents and goods, and by its position in the network. Cross-industrial links are directed, weighted flows of patent citations and intermediate goods normalized to shares. Interactions across layers capture the co-evolutionary nature of technology and markets.

Qualitatively, the technology of an industry is described by its links to other industries in each layer: The bundle of input links at the IO layer indicates that an industry has the technological know-how and equipment to make productive use of these inputs. Citation links at the patent layer indicate that R&D activities of the citing industry rely on patented inventions produced by other industries. Bundles of in- and outgoing links are used to quantify technological similarity of industry pairs at each layer. Demand-pull and technology-push effects arise from changes in an industry's output of goods and patents and may spill over to technologically similar industries. Spillovers flow upstream (downstream) to industries that are similar by in-going (out-going) links.

The coupled network is used to study the evolution of industry hierarchies and link formation processes at both layers. Hierarchies are described by the size ranking of industries measured by the level of output of goods and patents. The networks are used to compile industry-level panel data that are statistically analyzed.

It is found that both layers co-evolve but follow different dynamics: At the patent layer, industries became increasingly connected and the size distribution increasingly skewed. In the market, connectivity does not exhibit a clear trend and hierarchies are more sluggish in change. Industries became more similar by patent citations but less by customer links in the market indicating a process of market specialization while R&D activities converge. Patent similarity can be a driver of both, the technological con- and divergence of industries, in both layers. In contrast, market similarity is only predictive for the disconnection of industries.

The evolution of both layers is path-dependent: Big industries grow faster and new links are more likely formed if a link was formed in the preceding period. Well connected

¹The data is compiled from more disaggregate 6-digit level with 1179 different industries. Due to incomplete coverage, this number reduces to a balanced sample of 345 industries at the 6-digit level. To be rather conservative, the main analysis is based on the more aggregate 4-digit level, but robustness checks confirm that also the 6-digit level data and unbalanced samples yield consistent results.

and big industries connect to industries with a lower degree and smaller in size.

The impact of spillovers from similar industries differs across layers: In the market, upstream spillovers indicate that industries compete for the same type of inputs with other large industries. This is negatively associated with industrial growth. The opposite holds for downstream spillovers: Industries that have a large overlap in customer links with other big industries grow faster. Downstream similarity may indicate both, competition for similar customers or complementarity with positive mutual demand spillovers.

At the patent layer, the relations are opposite: Citing similar patents as other large industries is associated with higher growth. The effect of downstream spillovers is opposite: Having the same knowledge users as many other industries exhibits a negative relationship with size and growth. Downstream spillovers in the patent layer may indicate a lower degree of novelty of patents if the inventions of many other industries serve similar purposes.

These results indicate a conceptual difference between technology and markets: Goods traded in the market are scarce which imposes a barrier to growth and may explain the more balanced distribution. In contrast, inventions are not constrained by physical boundaries: Intangible knowledge required to invent patents is non-rival in use being associated with positive externalities from upstream links. This can be a source of increasing returns in the development of patented innovation.

One limitation must be emphasized: Classification systems are endogenous and evolve over time [cf. [Kay et al., 2014](#)]. Here, the evolution is studied using NAICS codes that were purposefully developed to describe industries in the market. This paper studies drivers of directed technological change manifested in the economy which justifies the use of NAICS codes rather than patent classes as means of description.

The results suggest that both demand-pull and technology-push influence the co-evolution of technology and markets. While market dynamics have a dampening effect on clustering and concentration, increasing returns from non-rival intangible knowledge give rise to the emergence of technology clusters. Increasing returns to innovation are not only engines of growth but can be also a reason for technological lock-in effects [[Arthur, 1989](#)]. The analysis shows that changing market conditions may lead to a redirection of R&D activities.

The remainder of the paper is structured as follows: In the next section, I give an overview of the related literature. The theoretical framework is explained in [Sec. 3](#). In [Sec. 4](#), I introduce the data. [Sec. 5](#) summarizes the results, [Sec. 6](#) offers a discussion, and [7](#) concludes.

2. Background and literature

Demand-pull versus technology-push as drivers of technological and economic change are subject of a long-lived debate. Demand-pull suggests that R&D activities follow the market, i.e. the perceived commercial potential of technological solutions offers an incentive for targeted R&D. Technology-push effects arise from technological opportunities that enable the development and commercialization of new products and processes. Both theories do not only focus on different allocation incentives but also different sources of ideas for technological improvements, emphasizing the role of users and customers (demand-pull) or external and internal research (technology-push) [Cohen and Levinthal, 1989, Kline and Rosenberg, 1986, Di Stefano et al., 2012].

A bibliometric study by Di Stefano et al. [2012] showed that more recent innovation studies abandon the traditional juxtaposition of demand-pull and technology-push, while firms' competences to absorb external and produce internal innovations became an increasingly relevant topic. External innovations originate from articulated user needs and internal build on firms' creativity to make use of capabilities.

In macroeconomic theory, technological knowledge plays a decisive role: Capabilities and absorptive capacities proxied by human capital are drivers of growth [Becker, 1994, Acemoglu, 2008, Nelson and Phelps, 1966, Romer, 1990, Aghion and Howitt, 1990]. Theories of non-neutral technological change argue that knowledge is technology-specific and accumulated by R&D. The direction of macroeconomic technological change is dependent on relative (factor) input costs that influence R&D allocation decisions and may differ across technology types [see Giovannoni, 2014, Giovannoni et al., 2014, Acemoglu, 2002, Popp et al., 2010].

Sector- and technology-specific progress can be empirically represented by learning curves describing the relationship between realized technological performance and cumulative production or research experiences [Arrow, 1962, Thompson, 2012]. Two-factor learning curve differentiate between *learning by using/doing* (LBD) arising from cumulative production and *learning by searching* (LBS) from cumulative R&D [Lundvall and Johnson, 1994, Di Comite et al., 2015, Wiesenthal et al., 2012]. LBS is a process of active search for (technological) knowledge, and LBU is a by-product of routine production processes [Lundvall and Johnson, 1994].

Firms invest in R&D to innovate themselves but also to build up the *absorptive capacity* to absorb external knowledge [Cohen and Levinthal, 1989]. Accumulated knowledge and technology-specific absorptive capacity are reasons for the path-dependence of technological change [Cohen and Levinthal, 2000]. Absorptive capacity can also explain

country-level patterns of technology adoption [cf. [Dosi, 1991](#), [Comin and Hobijn, 2010](#), [Lall, 1992](#), [Nelson and Phelps, 1966](#)].

This paper is based on a network approach to study directed technological change. In empirical production and patent citation networks, technology is qualitatively described by the position of an industry, firm or invention in the network: Knowledge (proxied by patents or physical output) is characterized by the bundle of adopted inputs, i.e. patent citations or physical production inputs. Two firms, industries or technologies are said to be similar if having a high share of overlapping in- and output links, i.e. if having the capability to make use of the same inputs and to serve the needs of similar customers. Similarity is a measure for the absorptive capacity of external knowledge [[Jaffe and De Rassenfosse, 2017](#), [Carvalho and Voigtländer, 2014](#), [Antony and Grebel, 2012](#), [Acemoglu et al., 2016](#), [Cai et al., 2017](#), [Huang, 2018](#), [Atalay et al., 2011](#)].

Using patent citation similarity as measure for absorptive capacity, [Antony and Grebel \[2012\]](#) showed that knowledge spillovers can explain firms' productivity improvements. [Kay et al. \[2014\]](#) used a patent-overlay mapping to illustrate evolving similarities as manifestation of technological change. Interpreting technological progress as expansion of technologies across fields, [Acemoglu et al. \[2016\]](#) used a patent citation network of US patents from 1975-2005 to show that upstream innovation creates positive spillovers on inventions that rely on citations to the upstream technology.² Using the same data, [Huang \[2018\]](#) models sectoral innovation and growth and finds that firms with many patents innovate more often and are more likely to innovate in technology fields that are similar to their pre-existing stock of patents.

Physical characteristics of an industry's production technology are reflected in the bundle of input used and outputs produced. The vicinity in IO networks may explain why firms and industries adopt specific inputs or acquire new customers. Using an empirical IO network relying quinquennial US data, [Carvalho and Voigtländer \[2014\]](#) observed that industries tend to adopt new inputs if they are similar to their pre-existing portfolio of inputs.³ [Boehm et al. \[2019\]](#) studied product line development of multi-product firms, i.e. the evolution of output links. Firms enter those product markets where they have core capabilities empirically proxied by firms' IO relationship. [Carvalho \[2014\]](#) have shown that relatedness through IO links can be a moderating factor of output fluctuations.

[Hausmann and Hidalgo \[2011\]](#) use a bipartite trade network of products and countries

²The same data is used in this paper, but the authors rely on IPC technology classes instead of NAICS codes as this paper does.

³They used the same data as used in this paper.

to describe technological capabilities revealed by a country’s comparative advantage in product exports. They find an inverse relationship between product ubiquity and countries’ export diversity and show how these measures may explain cross-country income differences.

In this paper, I use a coupled patent citation and IO network to describe technological knowledge empirically and to study the evolution of industries and cross-industrial links.

3. A model of learning and technological change

Technology is the capability to transform a bundle of inputs into outputs. Here, I consider two types of in- and outputs: Physical goods are in- and outputs of production processes and intangible knowledge encoded in patents is in- and output of R&D.

3.1. The economy as a two-layer network

Qualitative information about the technology used in an industry $i \in N$ is revealed by its IO connections in the market and patent citation patterns with N as set of industries. These relationships span a weighted, directed two-layered network. A node in the network represents an industry i which is connected with other industries $j \in N$ through patent citation links in the patent layer τ and IO links in the market layer μ . The layers $\alpha = \tau, \mu$ are coupled as a duplex network, i.e. each industry has a representation in each layer.

In-going (upstream) links indicate i ’s capability to transform a set of inputs supplied by industries j into an output (goods or patents). The output produced by i is supplied to other industries that are connected via out-going (downstream) links. Downstream links indicate that i has the capability to produce outputs that are useful to a specific set of other industries $j \in N$.

A network layer can be written as quadratic, asymmetric $|N| \times |N|$ matrix $W_t^{\alpha,d} = \{w_{ij,t}^{\alpha,d}\}_{i,j \in N}$ with positive non-zero entries if a link from i to j exists in time t . $d = in, out$ indicates the direction of a link, i.e. $w_{ij,t}^{\alpha,in}$ ($w_{ij,t}^{\alpha,out}$) indicates an input (output) link. The duplex network is given by the set of both matrices $W^d = \{W^{\tau,d}, W^{\mu,d}\}$. Links in the network are weighted $w_{ij,t}^{\alpha,d}$ and represent flows (goods and patent citations)

measured as shares

$$w_{ij,t}^{\alpha,d} = \frac{flow_{ij,t}^{\alpha,d}}{\sum_{k \in N} flow_{ik,t}^{\alpha,d}} \quad (1)$$

for $\alpha \in \{\mu, \tau\}$ and all $i, j \in N$. At the market layer, $flow_{ij,t}^{\mu,d}$ are physical goods (measured in monetary terms). At the patent layer, $flow_{ij,t}^{\tau,d}$ is the number of citations between patents held by i and j .

The entries of the matrices $W_t^{\alpha,d}$ representing a layer are column-wise normalized such that $\sum_j w_{ij,t}^{\alpha,d} = 1$.⁴ The weight $w_{ij,t}^{\alpha,in}$ reflects the relative importance of the output supplied by j in i 's input mix and $w_{ij,t}^{\alpha,out}$ measures the importance of j as customer or knowledge user to i . The normalization to shares makes different data types (patent counts and monetary flows) and industries that are very heterogeneous by size comparable.

The row-sum of weights $\sum_i w_{ij,t}^{\alpha,d}$ measures the *strength* of sector j . For $d = in$, this is a measure for the *input importance* of sector j describing the importance of j as input supplier to other industries, and for $d = out$, the strength measures the *output importance*, i.e. the importance of j as output user to other industries.

3.2. Technology-push and demand-pull as drivers of change

The industrial evolution manifests in the industry size $A_{i,t}^\alpha$ measured as i 's output of goods $A_{i,t}^\mu$ and patents $A_{i,t}^\tau$ in period t . Technology-push effects are positive shocks in the output of patents $A_{i,t}^\tau$ and may indicate a technological breakthrough that enables a surge of patented inventions. The effect may spill over to other R&D activities of other industries and to the market.⁵ Demand-pull effects are positive shocks in the output of goods $A_{i,t}^\mu$ indicating a rising demand for i 's output. This may spill over to the production activities of other industries and to the patent layer influencing the technological and economic evolution.

⁴Note that the *unweighted* output matrix equals the transposed unweighted input matrix but this does not hold for the *weighted* network, i.e. $flow_{ij,t}^{\alpha,in} = flow_{ji,t}^{\alpha,out}$ but $w_{ij,t}^{\alpha,in} \neq w_{ji,t}^{\alpha,out}$. The weighting factors differ for in- and outgoing links, i.e. $\sum_{k \in N} flow_{ik,t}^{\alpha,in} \neq \sum_{k \in N} flow_{jk,t}^{\alpha,out}$.

⁵Differently from other studies on patent citation networks [e.g. [Antony and Grebel, 2012](#), [Huang, 2018](#)], I do not use a perpetual inventory approach to compile patent stocks. Instead, the patents are aggregated into 5-year aggregates which is consistent with the quinquennial format of the IO data and has a similarly smoothing impact on industrial patent output (see Sec. 4).

3.3. Technological similarity

Technological similarity is measured in two ways: (1) Two industries i and j are similar if they have similar capabilities to make effective use of inputs, i.e. if i and j rely on similar physical production inputs and cite similar patents. (2) Industries can be also similar by outputs, i.e. if they are capable to serve the needs of similar customers.

The similarity is calculated separately for each layer and measured by the cosine similarity $\sigma_{ij,t}^{\alpha,d}$ of the vectors $\mathbf{w}_{k,t}^{\alpha,d} = (w_{k1,t}^{\alpha,d}, w_{k2,t}^{\alpha,d}, \dots, w_{kN,t}^{\alpha,d})$ of $k = i, j \in N$. The cosine similarity measures the angle between the two in- or output share vectors $\mathbf{w}_{i,t}^{\alpha,d}$ and $\mathbf{w}_{j,t}^{\alpha,d}$ on layer α in time t normalized to the length one. It is given by the normalized dot product

$$\sigma_{ij,t}^{\alpha,d} = \frac{\mathbf{w}_{i,t}^{\alpha,d} \mathbf{w}_{j,t}^{\alpha,d}}{\sqrt{(\mathbf{w}_{i,t}^{\alpha,d} \mathbf{w}_{i,t}^{\alpha,d})(\mathbf{w}_{j,t}^{\alpha,d} \mathbf{w}_{j,t}^{\alpha,d})}}. \quad (2)$$

The cosine similarity is a commonly used measure in network analysis and classification methods [Kay et al., 2014, McCune et al., 2002, Leydesdorff, 2005, Mikolov et al., 2013].

3.4. Spillovers

Technology-push and demand-pull may spill over to technologically related industries.

Knowledge that is created by R&D processes has public good characteristics and it is difficult to exclude others from using it once a new discovery is made. Hence, knowledge of j may spill over to i if i has the appropriate technology-specific *absorptive capacity* to make use of j 's knowledge [Cohen and Levinthal, 2000].

Analogously, positive market shocks in an industry j indicate the expansion of demand for j 's outputs but also an increasing demand for the production inputs used by j . Both effects may percolate through the market via price effects and spill over to industry i if it is technologically sufficiently related in terms of customer linkages and input use [Carvalho, 2014].

To which extent technology and market shocks spill over from j to i is moderated by the technological similarity $\sigma_{ij,t}^{\alpha,d}$.⁶ Similarity-weighted knowledge spillovers are

⁶Technological similarity as moderating factor for the adoption of new technologies and co-movement of business dynamics is well documented in the literature [e.g. Cohen and Levinthal, 2000, Carvalho and Voigtländer, 2014, Antony and Grebel, 2012, Carvalho, 2014].

calculated as

$$Spill(A)_{i,t}^{\alpha,d} = \sum_{j \neq i}^N \sigma_{ij,t}^{\alpha,d} \cdot A_{j,t}^{\alpha} \quad (3)$$

where $A_{j,t}^r$ is the stock of patents of industry j and $A_{j,t}^p$ is j 's output of physical goods.

3.5. Technological and economic importance

The technological and economic significance of an industry is reflected in its size measured by $A_{i,t}^{\alpha}$ and by its centrality in the network. An industry is central in the network if it is well connected to a great number of other (important) industries. Network centrality is an indicator for the economic relevance [cf. Jackson, 2008, Carvalho, 2014].

An industry can be important in two ways: (1) It can be the supplier of a good that is used by a great number of other industries and/ or that makes up a large share in the input bundle of other industries. (2) An industry can also be a critical customer of other industries. The former is associated with supply-side, the latter with demand-side market power.

Here, the centrality of an industry is measured by its in- and out-degree $D_{i,t}^{\alpha,d}$ given by the number of input (output) links of industry i for $d = in$ ($d = out$) which corresponds to the number of different inputs used (customers served) by i . I focus on the degree only because other centrality measures are highly correlated with the degree or industry size (see Fig. A.3).

3.6. Technological change

Technological change is a change in the production function that maps a set of inputs to outputs [Ruttan, 1959]. At the industry level, this is reflected in a changing composition of in- and outputs to physical production and R&D processes. Some industries grow, others shrink in relative and absolute terms. Within industries, patterns of input use and output supply change. This is observable in changing sector sizes, evolving IO and citation shares.

Here, technological change is analyzed in the following ways:

1. Relative aggregate stock variables $A_{i,t}^{\alpha}$ and their growth rates $gr(A_{i,t}^{\alpha})$ describe the rise and decline of industries, and

2. the formation of IO and citation links $w_{i,j,t}^{\alpha,d}$ that capture changes in the production technology.

These indicators are used in a series of regression analyses to identify drivers of change in the two network layers.

4. Data

The two-layer network is inferred from two different data sets on the US economy covering the 30-year period from 1976 to 2006. The market layer is compiled on national account data provided by the Bureau of Economic Analysis (BEA). The data are combined with the National Bureau of Economic Research (NBER) patent database covering all US patents and citations among US patents from 1976 to 2006 [Hall et al., 2001]. Patents are matched to firms that have an industry ID. This enables the compilation of a cross-industry patent citation network for different periods.

The compilation of the coupled network involved a series of steps. A very detailed description of the data processing is provided in SI.1 and the raw data and code use for the data compilation are available in the data publication Hötte [2021a]. Here, I explain only the most important steps.

4.1. Input-output data

BEA provides detailed current and historical benchmark IO tables in a quinquennial frequency dating back to 1947.⁷ During the data pre-processing, I used the most disaggregate data at the 6-digit level. The data are accounting data which are originally used to compile detailed national accounts (e.g. sectoral value-added, Social Accounting Matrices (SAM), etc.) [Horowitz and Planting, 2006]. The accounts show monetary flows between different industries including households' and government's final demand, and dummy positions that ensure the financial closure e.g. *Rest of the World* or *Scrap*. Accounting positions are largely but not perfectly compatible with NAICS or Standard Industrial Classification (SIC) codes. I converted the data step-wise into a time-consistent and convenient format. First, the data are transformed from accounting positions into industry codes, i.e. SIC codes for the 1977-1987 data and into NAICS

⁷The data were downloaded from <https://www.bea.gov/industry/benchmark-input-output-data> and <https://www.bea.gov/industry/historical-benchmark-input-output-tables> [Both accessed on Dec 21, 2020].

for later periods. The industry codes are harmonized to the NAICS 2002 version using concordance tables provided by BEA.⁸

After harmonizing the data, I obtain for each period a matrix of monetary flows between 1179 distinct 6-digit NAICS industries. The entries of the matrix are input flows $flow_{ij,t}^{\mu,in}$ indicating the monetary value of the inputs that i buys from j in time t . The transposed matrix represents output flows $flow_{ij,t}^{\mu,out}$. Division of the flows by the row sums $\sum_j flow_{ij,t}^{\mu,d}$ gives the input share matrix $W_t^{\mu,d}$. Note that some rows and columns are empty for some t . This is a result of the harmonization procedure to uniform NAICS codes and can happen when the classification changes. Industries can emerge or disappear over time. For example, industries associated with computer technologies were less granular in the 70s compared to the 90s. This is often associated with a split (merge) of pre-existing industries. For the main analysis, the data are aggregated to the 4-digit level which alleviates the problem of empty classes and concerns related to the precision and time consistency of the classification system.

4.2. Patent data

The patent citation layer is compiled from data provided and described by Hall et al. [2001].⁹ These data contain a mapping from patents to GVKEYs that serve as firm IDs in the Compustat database. Firms are assigned to industry codes (given by SIC and/or NAICS codes). Differently from other studies that use mappings from patent classes to industries [e.g. Dorner and Harhoff, 2018, Schmoch et al., 2003, Lybbert and Zolas, 2014, Kortum and Putnam, 1997], I map patents by ownership to publicly traded firms and obtain industry-specific patent and citation counts through aggregation of firms within an industry. This enables the compilation of a patent-industry mapping over a long time period at the disaggregate level.

The mapping of patents to industries is subject to crudity and exhibits some differences to classification of the IO data. In the IO data, the production activities are classified at the establishment level according to economic characteristics. If an establishment is active in multiple sectors, NAICS follows a majority principle and classifies the establishment by their “main business” [US Census Bureau, 2017]. Aggregation across establishments yields industry level data, but a single firm may run several establishments whose activities are classified differently. In contrast, one GVKEY in

⁸Detailed explanations of conceptual and technical issues (e.g. changing classification systems, ambiguous mappings) that arose during the compilation are available in SI.1.2.

⁹The data are downloaded from <https://sites.google.com/site/patentdataportproject/Home> [Accessed on Dec 21, 2020]. A mapping from firms to industries was obtained from Capital IQ.

the patent data is mapped to a unique NAICS code even though the firm may be active in multiple sectors. Moreover, in the available data, the mapping of firms to industries is fix over time while establishments are potentially re-classified if their major activity sufficiently changed. This inconsistency between the patent and IO data can not be overcome because establishment or firm level IO tables do not exist. The use of more aggregate 4-digit level data alleviates this problem.

To compile a network that has the same structure as the IO data, patent citations and counts are aggregated across 5 year windows from 1976-2006 at the industry level. For each period, the preceding 5 years were aggregated to compile the data of a given period. For example, for 1976, all patent citations between any two sectors i and j made in the years from 1972-1976 were summed up. Doing so, I obtained a symmetric patent citation flow matrix. An entry $flow_{ij,t}^{\tau,in}$ indicates the number of citations from i to j made in t . Again, the transposed matrix yields the out-going flows, i.e. $flow_{ij,t}^{\tau,out}$ indicates number of times that j cited i 's patents. As above, the entries of $flow_t^{\tau,d}$ are transformed to input shares $w_{ij,t}^{\tau,d}$ through division by the row sum $\sum_j flow_{ij,t}^{\tau,d}$. The final data builds on 5.79 M citation links between 1.15 unique patents.

4.3. Pre-processing

For the main analysis, I use 4-digit level data, but the data are also available at the more detailed 6-digit level and used for robustness checks. In contrast to many preceding studies [e.g. [Boehm et al., 2019](#), [Holly and Petrella, 2012](#), [Carvalho and Voigtländer, 2014](#), [Antony and Grebel, 2012](#)], this analysis covers all industries including services and primary sectors. The data are unbalanced panel data, i.e. some industries have no data entry for output flows or patent counts in some periods.¹⁰ For the main analysis, industries with incomplete coverage were removed, i.e. the final data is characterized by $A_{i,t}^\alpha > 0 \forall t, \alpha$. This reduces the sample size from 317 to 155 4-digit industries. The final data exhibits a bias to the manufacturing sector which accounts for 48% of the industries covered.¹¹

The 4-digit level is chosen to alleviate concerns related to the classification of industries arising from two conceptual difficulties: (1) Classification systems change over time. The earliest IO data and some of the firm IDs map to the Standard Industrial Classification (SIC) system. Throughout the compilation, I harmonized the data to

¹⁰The use of patents as means for Intellectual Property (IP) Right protection is heterogeneously spread across industries [[Jaffe and De Rassenfosse, 2019](#), [Blank and Kappos, 2012](#)]. Zero counts in IO tables may also occur as a result from reclassification (see above).

¹¹75 of 155 industries belong to the manufacturing sector.

2002 NAICS codes making sequentially use of concordance tables for reclassification. Whenever a sector maps to multiple other sectors, I distributed IO and patent citation flows equally. This hampers the precision at the detailed level. (2) Moreover, the mapping from patents to firm IDs and from IDs to industry codes is incomplete by coverage, i.e. some patents are not assigned to a firm, and in some cases firm IDs are mapped to coarse industry codes (e.g. 3- or 4-digit instead of 6-digit). In these cases, I distributed citation flows equally across sub-sectors.

The flow and weight matrices and the raw patent data are used to construct a panel of industry indicators. First, using the raw patent data, I computed aggregate patent stocks $A_{i,t}^{\tau}$.¹² Using the unweighted IO network, I extracted the sum of output given by the column sum $A_{i,t}^{\mu} = \sum_k flow_{ki,t}^{\mu,in}$ as measure for the market size. To cope with potential inconsistencies across time (e.g. potentially caused by classification system changes, price indices and general trends), I normalize the data by dividing these entries by the cross-industry average $\frac{1}{|N|} \sum_j A_{j,t}^{\alpha}$ for each t . The normalized size measures the size relative to other industries in t . Its cross-industry average equals one.

The network data are used to compile the degree $D_{i,t}^{\alpha,d}$, i.e. counting the number on in- and outgoing links.¹³

The weight matrices $W_t^{\alpha,d}$ are used to compute the cosine similarity matrices $\Sigma_t^{\alpha,d} = \{\sigma_{ij,t}^{\alpha,d}\}_{i,j \in N}$. The matrices $\Sigma_{i,t}^{\alpha,d}$ and industry sizes $A_{i,t}^{\alpha}$ are used to compute cross-industry spillovers $Spill(A)_{i,t}^{\alpha,d}$. Growth rates are calculated as $gr(A_{i,t}^{\alpha}) = \frac{A_{i,t}^{\alpha} - A_{i,t-1}^{\alpha}}{A_{i,t}^{\alpha} + A_{i,t-1}^{\alpha}}$ where the formula allows to deal with zero entries in the denominator.

For robustness checks, the unbalanced raw data is also aggregated to the more coarse 2- and 3-digit level. Note that all metrics that are derived from the network data are re-compiled for each aggregation level and data subset.¹⁴ To address concerns about the consistency and potential biases caused by the processing steps, all analyses presented in the paper were repeated at different aggregation levels, using the unbalanced data, and looking at manufacturing industries only (available in [A](#) and [SI.2](#)).

¹²To control for the heterogeneity of patents by value [[Hall et al., 2001](#)], I compiled citation weighted patent stocks for robustness checks.

¹³For data exploration purposes, also other metrics were computed, i.e. the strength $S_{i,t}^{\alpha,d}$, PageRank $PR_{i,t}^{\alpha,d}$ and betweenness centrality [see [Jackson, 2008](#)].

¹⁴For example, the pair-wise similarity $\sigma_{ij,t}^{\alpha,d}$ for the same digit level can differ across balanced versus unbalanced data.

5. Results

This section begins with a descriptive analysis of the two network layers and their overlap. Next, I show the results of two types of regression analyses: (1) It is analyzed how the evolution of sector hierarchies measured by relative size depends on pre-existing conditions and cross- and within-layer interactions. (2) I study the determinants of cross-industrial link formation processes.

5.1. Descriptive analysis

Fig. 1 shows a series of plots of the upstream network-layers and their overlap at different periods at the 4-digit level. A link between two nodes i and j is shown if j is a sufficiently important input supplier to i , i.e. if $w_{ij,t}^{\alpha,in}$ is higher than an average non-zero link of i in t . A link in the "overlap" network is shown if j is above average important to i as input provider in both layers. The distance of nodes is proportional to $w_{ij,t}^{\alpha,in}$, i.e. nodes are closer the stronger the connecting link. Industries that have strong mutual ties among each other tend to group together. The node colors indicate the broad industry class to which an industry belongs.¹⁵ In Fig. 1, it can be seen that industries belonging to the same broad sector group together. The node sizes in the figure are proportional to the industry size $A_{i,t}^{\alpha}$.¹⁶ Over time, agricultural industries and utilities (green and black color) decreased by size in the IO layer while information and service industries (blue color) grew. Industries became denser connected in the patent layer and their size distribution is more skewed.

Basic network statistics are summarized in Table 1. Network plots and statistics for other aggregation levels are available in A and SI.2.

Generally, Fig. 1 and Table 1 indicate a higher connectivity in the market compared to the patent layer.¹⁷ The IO network $W_t^{\mu,d}$ has a higher density, average degree and strength and a shorter diameter compared to $W_t^{\tau,d}$ but lower average weights in all periods. This indicates a higher level of in- and output diversification in μ relative to τ . The density and degree at layer τ are increasing over time which coincides with a decrease of average weights. The overlap network is least dense which must be true

¹⁵Green color is used for agriculture and mining, black for utilities, yellow for food processing, orange for non-metallic and red for metallic manufacturing, gray for retail, blue for information and management services and dark blue for other services.

¹⁶The figures show "average networks" given by the average $w_{ij,t}^{\alpha,d}$ and $A_{i,t}^{\alpha}$ over the periods 1976-1981, 1986-1991 and 1996-2006.

¹⁷Only significant connections are included where "significant" is defined by an above-average strength within the same layer but across time.

	Input-output			Patent			Overlap		
	1981	1991	2006	1981	1991	2006	1981	1991	2006
<i>Upstream network</i>									
Density	0.16	0.17	0.24	0.05	0.08	0.10	0.01	0.02	0.04
Avg. degree	25.09	26.18	37.25	8.08	12.01	15.09	2.38	3.41	5.40
Avg. strength	0.72	0.71	0.70	0.57	0.65	0.69	0.01	0.01	0.01
Avg. weight	0.79	0.76	0.74	2.43	1.76	1.38	8.43	5.45	4.41
Reciprocity	0.20	0.21	0.50	0.26	0.28	0.23	0.18	0.15	0.21
Diameter	2.00	2.00	2.00	4.00	3.00	3.00	5.00	4.00	3.00
Assort. by degree	-0.08	-0.10	-0.18	-0.25	-0.30	-0.31	-0.20	-0.25	-0.28
<i>Downstream network</i>									
Density	0.21	0.19	0.26	0.05	0.07	0.09	0.02	0.03	0.04
Avg. degree	32.39	29.53	40.93	7.48	11.54	13.55	2.83	4.02	5.39
Avg. strength	0.70	0.69	0.68	0.59	0.69	0.72	0.01	0.01	0.01
Avg. weight	0.79	0.76	0.74	2.57	1.82	1.41	9.51	6.69	4.78
Reciprocity	0.21	0.24	0.49	0.18	0.18	0.17	0.11	0.11	0.14
Diameter	2.00	2.00	2.00	4.00	3.00	3.00	5.00	4.00	3.00
Assort. by degree	-0.08	-0.10	-0.18	-0.25	-0.30	-0.31	-0.20	-0.25	-0.28

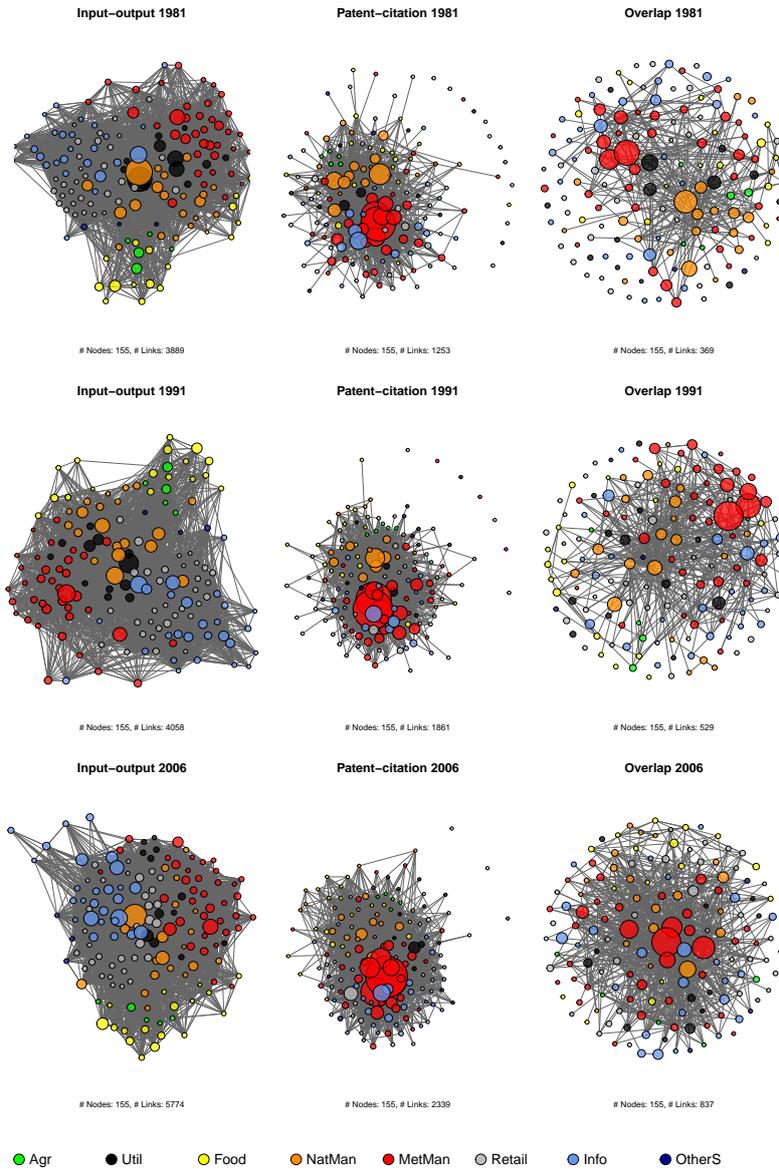
Notes: These statistics are computed on the weighted up- and downstream networks. For an introduction to network characteristics see [Jackson \[2008\]](#), [Csardi and Nepusz \[2006\]](#). "Assort." is the abbreviation for assortativity. Data: 4-digit balanced panel.

Table 1: Aggregate network statistics over time.

by definition: Industries that are connected at the patent layer are not necessarily connected at the market layer and vice versa.

The increasing connectivity at layer τ coincides with a growing similarity of in- and outgoing citation links which is reflected in an increasing density in the cosine similarity networks $\Sigma_t^{\tau,d}$ (see [Table A.1](#)).¹⁸ In contrast, in the market the connectivity does not exhibit a clear trend. Industries' input similarity is roughly constant but industries became more dissimilar by output links which indicates a process of specialization in product markets. These observations are robust across different levels of aggregation and data subsets (e.g. subset manufacturing sectors or unbalanced panel data, see [SI.2](#)).

¹⁸Similarity networks are compiled on the basis of symmetric matrices $\Sigma\alpha, d_t$ where the weight of link between i and j equals the cosine similarity $\sigma\alpha, d_{ij,t}$.



Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only in-going links are considered. Plots of the downstream network are available in [SI.2](#). Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Agr is Agriculture (1); Util are Mining and Utilities (2); Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33); Retail are Trade, Retail and Transportation sectors (4); Info are Information, Financial, Management and Administrative Services (5); OtherS are Other services and Public Sector (6-9). Data: Balanced panel of 4-digit industries.

Figure 1: Plots of the upstream networks at different periods for 4-digit level data.

<i>Top-10 industries by Aggregate output ($A_{i,t}^{\mu}$)</i>						
	1981		1991		2006	
1	Petroleum & Coal Prod.	3241 11.07	Elctr. Power Gen., Transm. & Distr.	2211 7.07	Mngm. Comp. & Enterpr.	5511 9.19
2	Elctr. Power Gen., Transm. & Distr.	2211 9.03	Motor Vhcl. Parts Mnft.	3363 6.10	Lessors of Real Estate	5311 5.89
3	Motor Vhcl. Parts Mnft.	3363 6.00	Petroleum & Coal Prod.	3241 5.87	Petroleum & Coal Prod.	3241 5.80
4	Architect., Engineer. & Srv.	5413 4.78	Architect., Engineer. & Srv.	5413 5.13	Motor Vhcl. Parts Mnft.	3363 5.20
5	Support Act. for Min.	2131 4.32	Basic Chem. Manufactur.	3251 4.36	Oil & Gas Extraction	2111 4.18
6	Oil & Gas Extraction	2111 3.90	Lessors of Real Estate	5311 3.73	Elctr. Power Gen., Transm. & Distr.	2211 3.70
7	Pulp, & Paper Mills	3221 3.09	Plastics Prod. Manufact.	3261 3.55	Basic Chem. Manufactur.	3251 3.70
8	Basic Chem. Manufactur.	3251 3.02	Semicond. & Oth. Elctr.	3344 3.32	Insurance Carriers	5241 3.60
9	Anim. Slaughter. & Prc.	3116 2.82	Pulp, & Paper Mills	3221 3.24	Plastics Prod. Manufact.	3261 3.45
10	Converted Paper Prod.	3222 2.74	Print. & Support Act.	3231 3.04	Semicond. & Oth. Elctr.	3344 3.33
Quartiles:						
	0.33, 0.6, 1.0625		0.36, 0.65, 1.14		0.26, 0.56, 1.16	

<i>Top 10 industries by Patent stock ($A_{i,t}^{\tau}$)</i>						
	1981		1991		2006	
1	Computer & Periph. Equ.	3341 12.81	Semicond. & Oth. Elctr.	3344 21.73	Semicond. & Oth. Elctr.	3344 25.52
2	Semicond. & Oth. Elctr.	3344 12.79	Computer & Periph. Equ.	3341 17.18	Computer & Periph. Equ.	3341 20.60
3	Pharm. & Medicine	3254 11.14	Communications Equ. Mnft.	3342 8.55	Communications Equ. Mnft.	3342 13.57
4	Petroleum & Coal Prod.	3241 8.70	Pharm. & Medicine	3254 8.55	Computer Syst. Design & Srv.	5415 8.55
5	Communications Equ. Mnft.	3342 5.99	Computer Syst. Design & Srv.	5415 6.29	Measur., Elctrmed. & Ctrl. Instr.	3345 7.28
6	Aerospace Prod. & Parts	3364 5.90	Measur., Elctrmed. & Ctrl. Instr.	3345 6.23	Pharm. & Medicine	3254 6.29
7	Motor Vehicle Manufactur.	3361 5.30	Petroleum & Coal Prod.	3241 6.15	Motor Vehicle Manufactur.	3361 5.47
8	Measur., Elctrmed. & Ctrl. Instr.	3345 5.27	Motor Vehicle Manufactur.	3361 5.14	Magnetic & Optical Media	3346 5.19
9	Computer Syst. Design & Srv.	5415 5.03	Chem. Prod. & Prepar.	3259 4.32	Motor Vhcl. Body & Trailer	3362 4.21
10	Chem. Prod. & Prepar.	3259 4.21	Resin, Synth. Rubber & Fibers	3252 4.00	Taxi & Limousine Service	4853 3.37
Quartiles:						
	0.03, 0.13, 1.055		0.02, 0.095, 0.89		0.02, 0.12, 0.74	

Notes: Each period shows a block of three columns indicating (1) the industry name, (2) the 4-digit NAICS code, (3) the normalized industry size $A_{i,t}^{\alpha}$. The average size normalized size equals one. Data: 4-digit level balanced panel.

Table 2: Ranking of the Top-10 industries by industry size ($A_{i,t}^{\alpha}$) over time.

Over time, the relative importance of industries has changed. Table 2 shows the Top-10 sectors ranked by $A_{i,t}^{\alpha}$ at different periods. $A_{i,t}^{\alpha}$ is normalized such that its average value equals one. The bottom line of the table shows the quartiles of $A_{i,t}^{\alpha}$. The median in both layers and for some years, also the 75%-quartile value are less than the average $\bar{A}_t^{\alpha} = 1$ which indicates a skewed size distribution with many small-sized industries and few large outliers. Comparing τ and μ , the size distribution in the market is less skewed compared to the patent layer which is also visible in Fig. 1.

The ranking by A_t^r documents the rise of information and communications technology (ICT) industries and related services.¹⁹ For example, in 2006, all Top-5 are occupied by industries related to ICT and similar services and equipment producers.

The rise of ICT sectors comes with the relative decline of others, especially of petroleum and coal product (3241), and aerospace manufacturing (3364) which were ranked as Top-4 and 6 in 1981, but disappeared from the Top-10 list by 2006. Pharmaceuticals and medicine (3254) gradually declined from rank 3 in 1981 to 6 in 2006. Other patent-intensive industries are related to chemical and synthetic product manufacturing and motor vehicles. Except from rising computer system services (5415), all patent-intensive industries belong to the manufacturing sector.²⁰

The rise of ICT related sectors is less pronounced in the market. In the Top-10 ranking, only the semiconductor industry appears in the Top-10 list in the years after 1991. The steady rise of ICT and supporting industries is better visible in the Top-50 ranking but the dominance of big petroleum industries, utilities and motor manufacturing is persistent (see A.2). These patterns of industrial change are robust in more aggregate data (cf. Tables SI.3-SI.4).

5.2. Demand-pull, technology-push and the evolution of industry hierarchies

Drivers of industrial change are analyzed in a series of regression analyses. Here, I study the evolution of industrial hierarchies. In the subsequent section 5.3, I analyze the determinants of cross-industrial link formation.

To understand the evolution of hierarchies, it is analyzed how lagged sector sizes $A_{i,t-1}^\alpha$, spillovers $Spill(A)_{i,t-1}^{\alpha,d}$, and network connectivity captured by the $D_{i,t-1}^{\alpha,d}$ are related to the industry size $A_{i,t}^\alpha$ and its growth rate $gr(A_{i,t}^\alpha)$ on both layers.²¹ Two separate sets of regressions are run distinguishing between network controls ($D_{i,t-1}^{\alpha,d}$,

¹⁹These are mainly the following industries: 3341 (Computer and Peripheral Equipment Manufacturing), 3333 (Commercial and Service Industry Machinery Manufacturing), 5415 (Computer Systems Design and Related Services), 3342 (Communications Equipment Manufacturing), 3344 (Semiconductor and Other Electronic Component Manufacturing).

²⁰The patent-intensity of manufacturing industries is in line with previous studies and points to an important problem when quantifying intellectual property (IP) of industries: Many commercial and service industry rely on other types of IP protection such as copyright and trademarks Blank and Kappos [2012].

²¹Tests with other network indicators were made but rejected to avoid multicollinearity. For example, the PageRank is strongly correlated with the industry size and measures computed on the basis of in- and outgoing links exhibit correlations of > 90%. The final model specification was selected in a step-wise model selection procedure using the Bayesian Information Criterion. Further information on the model selection procedure is available in the SI.1.4.2.

$Spill(A)_{i,t-1}^{\alpha,d}$) compiled on the basis of upstream $d = in$ or downstream $d = out$ links. Due to multicollinearity, I avoid to use both measures simultaneously. The regression equations are given by

$$Y_{i,t} = \sum_{\alpha=\mu,\tau} \left[\beta_A^\alpha A_{i,t-1}^\alpha + \beta_D^\alpha D_{i,t-1}^{\alpha,d} + \beta_S^\alpha Spill(A)_{i,t-1}^{\alpha,d} \right] \quad (4)$$

where $Y_{i,t} \in \{A_{i,t}^\alpha, gr(A_{i,t}^\alpha)\}_{\alpha=\mu,\tau}$ and $d = in, out$.

Industry and time fixed effects (FE) are included to account for unobserved heterogeneity and standard errors are two-ways clustered.²² Because some variables are highly skewed, they were log-linearized. To make the regression coefficients comparable, all variables were scaled and demeaned (see for additional detail [SI.1.4.1](#)).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
$A_{i,t-1}^\mu$	0.1688*** (0.0281)	0.5007*** (0.0484)	0.1613*** (0.0264)	0.5339*** (0.0521)	-0.012 (0.0111)	8e-04 (0.0063)	-0.0098 (0.0087)	0.0012 (0.0065)
$A_{i,t-1}^\tau$	0.0086 (0.0396)	-9e-04 (0.0482)	-0.0259 (0.0401)	-0.0305 (0.0435)	0.0955** (0.0295)	0.6727*** (0.0401)	0.0267 (0.0237)	0.6746*** (0.0385)
$D_{i,t-1}^{\mu,d}$	0.0128 (0.0173)	0.0793*** (0.0234)	-0.0524*** (0.0111)	-0.0352* (0.0144)	0.0144 (0.009)	0.005 (0.005)	0.0052 (0.0072)	-0.005* (0.0025)
$D_{i,t-1}^{\tau,d}$	0.0216 (0.0294)	0.0089 (0.0343)	-0.0022 (0.0272)	0.0152 (0.029)	-0.153*** (0.035)	0.0161* (0.0073)	0.1626*** (0.0276)	0.0353*** (0.0087)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0707*** (0.0145)	0.0071 (0.0133)	-0.0103 (0.0172)	0.0333 (0.0273)	-0.0099 (0.0104)	0.0039 (0.0038)	0.0149 (0.0139)	6e-04 (0.0069)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.0076 (0.0124)	0.0094 (0.0124)	0.0138 (0.016)	0.0059 (0.02)	0.037 (0.0191)	0.0037 (0.0036)	-0.0521** (0.018)	-0.003 (0.0036)
R^2	0.2316	0.2705	0.2362	0.2622	0.0615	0.6859	0.071	0.6912
N	930	930	930	930	930	930	930	930
Average	0.0066	0.5692	0.0066	0.5692	-0.0517	0.3878	-0.0517	0.3878

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 '.' 1

Notes: The model is estimated as linear OLS model including time and industry FE and using two-ways clustered SE. The explanatory variables have been selected in a step-wise selection procedure searching for the highest explanatory power measured by the BIC (see text for further detail). Regressions are run separately for spillovers and network-based variables computed on the basis of upstream ($d = in$) and downstream ($d = out$) links. Data: 4-digit, balanced panel.

Table 3: Regression results explaining the evolution of industrial hierarchies.

The results are shown in [Table 3](#). Additional results for more disaggregate 6-digit level data and the subset of manufacturing sectors are available in [A.2](#), and further

²²Additional analyses using time-only FE yield qualitatively consistent results.

robustness checks including trade data, using the unbalanced panel, and more aggregate 3-digit level data are provided in [SI.2.3.1](#). Columns (1)-(4) show the regression results at the market layer, and columns (5)-(8) at the patent layer. For each layer, the first two columns show the results using upstream, the latter two using downstream links for $D_{i,t-1}^{\alpha,d}$ and $Spill(A)_{i,t-1}^{\alpha,d}$.

At both layers, the results reveal path-dependency in the evolution of hierarchies: Industry size $A_{i,t}^\alpha$ and growth $gr(A_{i,t}^\alpha)$ are positively associated with lagged size $A_{i,t-1}^\alpha$. These relationships hold within both layers, but do not exhibit a significant effect across different layers. At the more disaggregate 6-digit level data (cf. [Table A.5](#)), a significant negative correlation between patent growth and market size is found. This is measured after controlling for industry FE, i.e. the correlations show the deviation from i 's average size and growth rate.

In the market, higher upstream connectivity (input diversification) measured by $D_{i,t-1}^{\mu,in}$ is positively associated with industry size and growth in both layers which is more significant if using 6-digit level data. The opposite holds for downstream connectivity $D_{i,t-1}^{\mu,d}$: It is negative related to industry size and growth which is also weakly significant at layer τ . In the patent layer, input diversification $D_{i,t-1}^{\tau,in}$ is negatively related to patent growth while output diversification (i.e. a high number of knowledge users) is positively associated with i 's size and growth in layer τ .

Upstream spillovers in the market $Spill(A)_{i,t-1}^{\mu,in}$ are negatively correlated with market growth. This effect is stronger in the subset of manufacturing sectors ([Table A.4](#)). Upstream (downstream) patent spillovers are positively (negatively) correlated with patent growth and size which is even more significant at the 6-digit level.

The predictive power of the model accounts for 24-28% in the market, and 3-7% for growth rates and 65% for the size in the patent layer. The regressions using downstream links perform slightly better than those using upstream links.

5.3. What determines the formation of cross-industrial links?

Changing patterns of cross-industrial IO flows and patent citations, reflected by $\Delta w_{ij,t}^{\alpha,d} = w_{ij,t}^{\alpha,d} - w_{ij,t-1}^{\alpha,d}$, indicate a changing direction of economic and technological change. The evolution of up- and downstream links is studied in two steps: First, I analyze link formation at the *extensive margin*, i.e. the probability that new links form and existing links break. Second, I analyze the *intensive margin* looking at the magnitude of $\Delta w_{ij,t}^{\alpha,d}$ conditional on $\Delta w_{ij,t}^{\alpha,d} \neq 0$.

5.3.1. Link formation at the extensive margin

The probability that a link from i to j on layer α changes is studied with a series of logistic regressions given by

$$Pr(\Delta w_{ij,t}^\alpha \gg 0) = \text{Logit} \left(\sum_{\alpha=\mu,\tau} \left(\beta_1^\gamma \mathbb{1}(w_{ij,t-1}^{\gamma,d}) + \beta_\sigma^\gamma \sigma_{ij,t-1}^{\gamma,d} + \beta_\Delta^\gamma \Delta \sigma_{ij,t-1}^{\gamma,d} \right) \right. \\ \left. + \sum_{k=i,j} \left(\beta_{A,k}^\gamma A_{k,t-1}^\gamma + \beta_{D,k}^\gamma D_{k,t-1}^{\gamma,d} + \beta_{S,k}^\gamma \text{Spill}(A)_{k,t-1}^{\gamma,d} \right) \right). \quad (5)$$

A mutual link from i to j is defined as being formed or broken if the magnitude of weight change exceeds a threshold level, i.e. if $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$. A positive (negative) sign of $\Delta w_{ij,t}^{\alpha,d}$ indicates that a link is formed (broken). The threshold level $c_t^{\alpha,d}$ is the average absolute change in t , i.e. $c_t^{\alpha,d} = \frac{1}{|N|^2} \sum_{i \in N} \sum_{j \in N} |\Delta w_{ij,t}^{\alpha,d}|$.

Lagged pair-wise and individual characteristics of both i and j are used as explanatory variables. Pair-wise variables in the regression are $\sigma_{ij,t-1}^{\alpha,d}$, its change in the preceding period $\Delta \sigma_{ij,t-1}^{\alpha,d} = \sigma_{ij,t-1}^{\alpha,d} - \sigma_{ij,t-2}^{\alpha,d}$ and lagged binary indicators $\mathbb{1}(\Delta w_{ij,t-1}^{\alpha,d} \gg 0)$ ($\mathbb{1}(w_{ij,t-1}^{\alpha,d} = 0)$) that equal 1 if a link was formed (no significant change occurred) in $t-1$ and zero otherwise. The regression coefficients β_1^α indicate the difference in $Pr(\Delta w_{ij,t}^\alpha \gg 0)$ compared to an observation with a link broken in $t-1$.²³ Analogous regressions are run to investigate the determinants of link breaking probability $Pr(\Delta w_{ij,t}^\alpha \ll 0)$ using the same model specification.

²³As above (Sec. 5.2), the final model equation is selected in a step-wise procedure using the BIC as selection criterion. Additional detail about the model selection procedure is available in SI.1.4.

Other authors who studied link formation in isolated patent citation and IO networks used Probit, linear probability and hazard duration models [Carvalho and Voigtländer, 2014, Boehm et al., 2019, Huang, 2018, Acemoglu et al., 2016].

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Pr(\Delta w_{ij,t}^{\mu} \geq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \leq 0)$	$Pr(\Delta w_{ij,t}^{\pi} \geq 0)$	$Pr(\Delta w_{ij,t}^{\pi} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$
	upstream		downstream		upstream		downstream	
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	-0.0258*** (0.0058)	-0.0383*** (0.0046)	-0.0095. (0.0055)	-0.0111* (0.0044)	-4e-04 (0.0034)	-0.0057. (0.0034)	-0.0036 (0.0036)	-0.0016 (0.0033)
$1(\Delta w_{ij,t-1}^{\mu,d} \geq 0)$	0.0698*** (0.0044)	0.0911*** (0.0036)	0.0637*** (0.0043)	0.0533*** (0.0034)	-0.0016 (0.0015)	-0.0021 (0.0013)	-0.0029. (0.0015)	-0.0027* (0.0014)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	0.0095 (0.0074)	-0.0042 (0.0064)	0.0038 (0.0068)	-0.0056 (0.006)	0.0658*** (0.0048)	-0.1318*** (0.0051)	0.0685*** (0.0047)	-0.161*** (0.0053)
$1(\Delta w_{ij,t-1}^{\tau,d} \geq 0)$	0.0075 (0.0053)	0.0018 (0.0046)	0.0038 (0.0049)	0.002 (0.0043)	0.1215*** (0.0063)	0.1473*** (0.0048)	0.1287*** (0.0063)	0.1419*** (0.0047)
$\sigma_{ij,t-1}^{\mu,d}$	0.0143* (0.0064)	0.0564*** (0.0056)	-0.0076 (0.0066)	0.0617*** (0.0056)	0.0018 (0.0024)	-0.0011 (0.0021)	-0.0043* (0.0022)	-0.0043* (0.0021)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	0.0194*** (0.0034)	0.0356*** (0.0029)	0.0024 (0.003)	0.0313*** (0.0026)	0 (0.0012)	4e-04 (0.0011)	-0.001 (0.0011)	-0.0013 (0.001)
$\sigma_{ij,t-1}^{\tau,d}$	-0.0044 (0.0033)	-0.0036 (0.0029)	0.0036 (0.0034)	5e-04 (0.0031)	0.0061*** (0.0015)	0.001 (0.0013)	0.0054*** (0.0015)	0.0045*** (0.0013)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	-9e-04 (0.002)	-0.0041* (0.0018)	-0.002 (0.0021)	-0.0015 (0.0019)	0.0026** (9e-04)	-4e-04 (7e-04)	0.0025** (8e-04)	0.0033*** (7e-04)
$A_{i,t-1}^{\mu}$	0.0786*** (0.0035)	0.0735*** (0.0037)	0.0385*** (0.0043)	0.1379*** (0.0034)	-0.0045** (0.0017)	-0.0052*** (0.0014)	-0.0058*** (0.0016)	-0.003* (0.0014)
$A_{j,t-1}^{\mu}$	-0.0062 (0.0043)	0.0203*** (0.0036)	0.0054. (0.0032)	-0.0021 (0.0031)	-2e-04 (0.0017)	0.0018 (0.0014)	0 (0.0017)	0.0038** (0.0015)
$A_{i,t-1}^{\tau}$	-0.0513*** (0.0104)	-0.0592*** (0.0089)	-0.0286** (0.0098)	0.0091 (0.0082)	0.0272*** (0.0081)	0.0613*** (0.0074)	0.0665*** (0.0078)	0.0505*** (0.0066)
$A_{j,t-1}^{\tau}$	-0.0175. (0.01)	-0.0132 (0.0083)	-0.0166* (0.0083)	0.0058 (0.0075)	-0.0117** (0.0042)	-0.0199*** (0.0044)	-0.0104* (0.0043)	-0.0043 (0.0046)
$D_{i,t-1}^{\mu,d}$	0.0194*** (0.0028)	-0.0037 (0.0023)	0.1092*** (0.0092)	0.0669*** (0.0083)	-0.0037* (0.0018)	0.0033. (0.0017)	-0.0155*** (0.0034)	0.0013 (0.0033)
$D_{j,t-1}^{\mu,d}$	0.014*** (0.0026)	0.0141*** (0.0023)	0.1225*** (0.0087)	0.073*** (0.0079)	0.0021 (0.0016)	0.003. (0.0016)	0.0018 (0.004)	0.0175*** (0.0038)
$D_{i,t-1}^{\tau,d}$	-0.0032 (0.0049)	0.0147*** (0.0044)	-0.0088 (0.0054)	-0.0125** (0.0048)	-0.0042* (0.0017)	-0.0085*** (0.0014)	-0.02*** (0.0016)	-0.0077*** (0.0016)
$D_{j,t-1}^{\tau,d}$	0.0104. (0.0055)	-0.0191*** (0.0047)	0.0123* (0.0052)	0.0086. (0.0049)	0.0198*** (0.0021)	0.0299*** (0.0017)	0.0249*** (0.0024)	0.0342*** (0.0018)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0214*** (0.0046)	-0.0049 (0.0039)	0.0207*** (0.0038)	0.0173*** (0.0034)	0.0067** (0.0021)	-0.0016 (0.0019)	0.0022 (0.0016)	0.0064*** (0.0015)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.0117** (0.0045)	-0.0137*** (0.0039)	0.0174*** (0.0037)	-0.0108*** (0.0032)	-0.0026 (0.0022)	-0.0071*** (0.002)	0.0021 (0.0019)	-0.0025 (0.0017)
$Spill(A)_{i,t-1}^{\tau,d}$	0.0079** (0.0028)	0.0091*** (0.0026)	0.0086*** (0.0025)	-6e-04 (0.0022)	-0.0059*** (6e-04)	-0.0015** (6e-04)	-0.0044*** (5e-04)	-0.0023*** (5e-04)
$Spill(A)_{j,t-1}^{\tau,d}$	0.0055. (0.003)	0.0101*** (0.0027)	0.0027 (0.0021)	3e-04 (0.0019)	-9e-04 (0.0012)	0.0051*** (9e-04)	0.0025** (8e-04)	0.0026*** (7e-04)
R^2	0.0326	0.0453	0.0323	0.0848	0.0218	0.1334	0.028	0.1472
N	66831	73141	67669	72302	96546	96723	95942	97332
Sample avg.	0.3895	0.4367	0.3989	0.4273	0.1857	0.168	0.1816	0.1721
Subset avg.	0.087	0.0966	0.0839	0.0887	0.0638	0.0575	0.0587	0.0542

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). The sample average is the probability of link formation or breaking in the full sample, while the subset average is the frequency of significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$. Data: 4-digit balanced panel.

Table 4: Regression of link formation & breaking.

The results are shown in Table 4. All regressions control for industry-pair and time FEs and use clustered standard errors (SE). As before, the regressions are run on different data subsets and aggregation levels (see A.3 and SI.2.3.2). Column (1)-(4) ((5)-(8)) show the results for the market (patent) layer where the first (last) two columns

show the results for upstream (downstream) links. Odd (even) numbered columns show the drivers of link formation (breaking). The probability to observe a link being formed or broken as defined above accounts for roughly 9% in the market and 6% in the patent layer. The major findings are row-wise summarized highlighting cross-layer interactions.

In both layers, it is more likely that a new link is formed or broken if a link was formed in $t - 1$ at the same layer, i.e. $\mathbb{1}(\Delta w_{ij,t-1}^{\alpha,d} \gg 0)$. In the market, stability is auto-correlated, i.e. it is less likely that a link forms or breaks if it was stable before, i.e. $\Delta w_{ij,t-1}^{\alpha,d} = 0$. Differently, in the patent layer, link breaking is auto-correlated: If a link was weakened in $t - 1$, it is more likely that the link continues to weaken (indicated by the negative coefficient of $\mathbb{1}(\Delta w_{ij,t-1}^{\alpha,d}) = 0$). Stability in mutual links, i.e. $\Delta w_{ij,t-1}^{\tau,d} = 0$, is positively associated with the formation probability compared to a link being broken before.

Looking at more disaggregate 6-digit data (Table A.8), patent linking is positively (negatively) correlated with input link formation (breaking) in the market: It is less (more) likely that market links form (break) if industries connected in the patent layer in $t - 1$. Lagged market link formation is negatively correlated with the probability patent input links break.

Market input similarity $\sigma_{ij,t}^{\mu,in}$ has an inconsistent impact on the link formation probability: Using aggregate data (2-4 digit), it is significantly positively associated with $Pr(\Delta w_{ij,t}^{\mu} \gg 0)$, but at the 6-digit level it has the opposite effect and also has a negative effect on the probability that patent links form. Output similarity $\sigma_{ij,t}^{\mu,out}$ is consistently negatively associated with the link formation probability at all aggregation levels and both in- and output similarity make it more likely that in- and output links break in both layers.

Patent similarity (up- and downstream) shows opposite associations: It is positively correlated with patent link formation and breaking dynamics. Moreover, it appears to have a stabilizing effect on input link formation processes in the market but a destabilizing effect on market output links making it less likely that market links change (see also Table A.8).

Big industries are more actively rewiring: $A_{i,t-1}^{\alpha}$ is positively correlated with the probability to loose or create in- and output links within the same layer. Across layers, industry size has the opposite effect: Industries that are big in the market less likely rewire in τ and industries with many patents less likely rewire in μ . In the market, input links are more likely formed and broken to industries j that are big in μ , while there is weak evidence that output links with big market customers are less likely to

be broken. The relationship between $A_{j,t-1}^\mu$ is inconsistent across layers: At the most disaggregate level, links are less likely formed to big industries while the opposite effect is observed at the 2-4 digit level. In the patent layer, the j 's size positively associated with the stability of links: Patent citation links are less likely formed or broken if $A_{j,t-1}^\alpha$, $\alpha = \mu, \tau$ is high.

An industry with a high degree in the market indicated by $D_{i,t-1}^{\mu,d}$ is more likely to form new market but less likely to form new patent links. The opposite holds for industries with a high patent indegree: They are less likely to form and break links, but patent links are more likely formed to industries j characterized by a high degree. Within both layers, up- and downstream links are more likely formed and broken to well connected industries indicated $D_{j,t-1}^{\alpha,d}$.

Market input links are less likely formed between industries that are subject to high market spillovers, i.e. if $Spill(A)_{k,t-1}^{\mu,d}$ is high $\forall k = j, i$. The opposite holds for market output links. Market spillovers have a positive correlation with the probability that patent links form and break. In contrast, patent spillovers $Spill(A)_{i,t-1}^{\tau,d}$ to the benefit of i have a stabilizing effect, i.e. $Spill(A)_{i,t-1}^{\tau,d}$ is negatively correlated with the probability that patent links change. Patent spillovers to the benefit of the target industry j are positively correlated with the probability that market and patent links change.

Generally, the R^2 accounts for only 3-8% at the market layer. At the patent layer, it is low for the formation of links (2-3%) and slightly higher for the probability of link breaking (13-15%).

5.3.2. Link formation at the intensive margin

Link formation at the intensive margin $\Delta w_{ij,t}^\alpha$ describes the magnitude of change if a link changes, i.e. given $\Delta w_{ij,t}^\alpha \neq 0$. The intensive margin is analyzed in an OLS regressions including time and industry-pair FE, clustered SE and using similar explanatory variables as above (Sec. 5.3.1), i.e.

$$\Delta w_{ij,t}^{\alpha,d} = \sum_{\alpha=\mu,\tau} \left[\beta_{\Delta w,\alpha}^\alpha \Delta w_{ij,t-1}^{\alpha,d} + \beta_\sigma^\alpha \sigma_{ij,t-1}^{\alpha,d} + \beta_\Delta^\alpha \Delta \sigma_{ij,t-1}^{\alpha,d} \right. \\ \left. + \sum_{k=i,j} \left[\beta_{A,k}^\alpha A_{k,t-1}^\alpha + \beta_{D,k}^\alpha D_{k,t-1}^{\alpha,d} + \beta_{S,k}^\alpha Spill(A)_{k,t-1}^{\alpha,d} \right] \right]. \quad (6)$$

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	0.0751 (0.0436)	-0.1504** (0.0456)	-0.1196* (0.0508)	-0.2098*** (0.0284)	-0.0088 (0.0121)	-0.0018 (0.0197)	0.0089 (0.0232)	-0.0043 (0.0366)
$\Delta w_{ij,t-1}^{\tau,d}$	0.0053 (0.004)	-0.0024 (0.003)	-9e-04 (0.003)	-0.0046 (0.0028)	0.0977*** (0.0267)	-0.1718*** (0.0187)	0.1361*** (0.0365)	-0.2143*** (0.0198)
$\sigma_{ij,t-1}^{\mu,d}$	9e-04 (6e-04)	-0.0011 (6e-04)	-6e-04 (6e-04)	-0.0035*** (5e-04)	0 (9e-04)	-2e-04 (6e-04)	9e-04 (0.001)	-2e-04 (6e-04)
$\Delta \sigma_{ij,t-1}^{\mu,d}$	6e-04 (3e-04)	-9e-04*** (2e-04)	-5e-04 (3e-04)	-0.0012*** (2e-04)	0 (4e-04)	-2e-04 (3e-04)	0 (4e-04)	3e-04 (3e-04)
$\sigma_{ij,t-1}^{\tau,d}$	3e-04 (2e-04)	-4e-04 (2e-04)	5e-04** (2e-04)	-6e-04*** (2e-04)	0 (6e-04)	-0.0011* (5e-04)	-4e-04 (9e-04)	-0.0017* (8e-04)
$\Delta \sigma_{ij,t-1}^{\tau,d}$	-1e-04 (1e-04)	0 (1e-04)	1e-04 (1e-04)	-3e-04** (1e-04)	2e-04 (4e-04)	-4e-04 (3e-04)	-1e-04 (6e-04)	-0.0012* (5e-04)
$A_{i,t-1}^{\mu}$	0.0039*** (5e-04)	-0.0028*** (4e-04)	0.0014*** (3e-04)	-0.0062*** (3e-04)	-6e-04 (5e-04)	8e-04 (4e-04)	7e-04 (6e-04)	-1e-04 (5e-04)
$A_{j,t-1}^{\mu}$	-4e-04 (2e-04)	7e-04*** (2e-04)	-4e-04 (2e-04)	5e-04* (2e-04)	8e-04 (6e-04)	-1e-04 (3e-04)	1e-04 (5e-04)	6e-04 (4e-04)
$A_{i,t-1}^{\tau}$	-0.0049*** (0.0011)	0.0017** (5e-04)	-0.0015 (8e-04)	5e-04 (7e-04)	0.0014 (0.0011)	8e-04 (0.001)	0.0104*** (0.0017)	-0.0052*** (0.0012)
$A_{j,t-1}^{\tau}$	1e-04 (0.001)	-7e-04 (4e-04)	6e-04 (5e-04)	6e-04 (7e-04)	0.0015* (6e-04)	-0.0021*** (6e-04)	-2e-04 (6e-04)	-3e-04 (4e-04)
$D_{i,t-1}^{\mu,d}$	-2e-04 (3e-04)	1e-04 (2e-04)	-0.0035*** (6e-04)	0.0013** (5e-04)	-4e-04 (4e-04)	-4e-04 (3e-04)	-0.001 (9e-04)	-8e-04 (9e-04)
$D_{j,t-1}^{\mu,d}$	0.001** (4e-04)	9e-04 (8e-04)	-0.0014* (7e-04)	0.0015** (5e-04)	-0.001* (4e-04)	0 (6e-04)	-0.0012 (0.0015)	0.0045** (0.0014)
$D_{i,t-1}^{\tau,d}$	0.001** (3e-04)	-3e-04 (2e-04)	2e-04 (3e-04)	-1e-04 (2e-04)	-3e-04 (6e-04)	-0.0011 (7e-04)	-0.0019** (6e-04)	1e-04 (5e-04)
$D_{j,t-1}^{\tau,d}$	1e-04 (4e-04)	2e-04 (2e-04)	3e-04 (3e-04)	-4e-04 (3e-04)	-0.0205*** (0.0026)	0.0297*** (0.0034)	-0.0205*** (0.0032)	0.0385*** (0.0043)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0011** (4e-04)	5e-04 (3e-04)	0.0011*** (3e-04)	-1e-04 (3e-04)	0 (4e-04)	8e-04* (4e-04)	0 (5e-04)	-9e-04 (5e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.0013** (4e-04)	8e-04** (3e-04)	2e-04 (3e-04)	6e-04** (2e-04)	2e-04 (6e-04)	1e-04 (5e-04)	4e-04 (7e-04)	-0.0012* (5e-04)
$Spill(A)_{i,t-1}^{\tau,d}$	-5e-04** (2e-04)	2e-04* (1e-04)	2e-04 (1e-04)	-3e-04* (1e-04)	2e-04 (5e-04)	0.002** (7e-04)	2e-04 (7e-04)	0.0013* (7e-04)
$Spill(A)_{j,t-1}^{\tau,d}$	3e-04 (2e-04)	1e-04 (1e-04)	-3e-04** (1e-04)	2e-04 (1e-04)	-6e-04 (0.0011)	0.0011 (8e-04)	-2e-04 (0.001)	8e-04 (0.001)
R^2	0.0221	0.0507	0.0102	0.1505	0.0583	0.2481	0.0747	0.3064
N	46209	52519	47048	51681	22627	22804	22018	23408
Subset avg.*	0.005	-0.0044	0.0048	-0.0046	0.01	-0.0097	0.0121	-0.011
Subset avg.**	0.0183	-0.0158	0.0183	-0.0182	0.0336	-0.0336	0.0418	-0.0413

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are OLS regressions including industry-pair and time FE and using two-ways clustered SE estimating the magnitude of change conditional on $\Delta w_{ij,t}^{\alpha,d} \neq 0$. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta w_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.

Data: 4-digit balanced panel.

Table 5: Regression of link formation & breaking at the intensive margin.

The results are shown in Table 5. Additional results are available in A.4 and SI.2.3.3. The average $\Delta w_{ij,t}^{\alpha,d}$ is small and accounts for roughly 0.4-0.5 (1.2-1.4) percentage points for $\alpha = \mu$ ($\alpha = \tau$).²⁴ The predictive power of the regressions is low in the market with an adjusted R^2 of 1 – 15%. It is higher at the patent layer ranging between 6 – 31%. Again, the results are summarized row-wise.

In the patent layer, $\Delta w_{ij,t}^{\tau,d}$ is auto-correlated for both weakening and strengthening links. In the market, the magnitude of change is autocorrelated if links are weakened.

Pairwise similarity has only weak explanatory power, but similar industries tend to adjust links more heavily if $\Delta w_{ij,t}^{\alpha,in} < 0$.

Within the same layer, the size of i is positively correlated with the magnitude of change, i.e. strengthening (weakening) links become stronger (weaker) if i is large. Across layers, industries with a large patent stock $A_{i,t-1}^{\tau}$ adjust market links more slowly and industries that are big in the market adjust patent links more slowly.

New market output links are adjusted by a smaller magnitude if both i and j are already well connected. Citation link adjustments in both directions are smaller if j has many in- and outgoing links.

Spillovers exhibit only weak interactions with the size of link adjustments: Both market and patent link adjustments tend to be more moderate if i and j are exposed to high spillovers.

6. Discussion

The main insights of this analysis can be summarized as follows:

1. During the period of study, both network layers exhibit different dynamics: The patent layer becomes increasingly connected and the size distribution of industries increasingly skewed. In contrast, connectivity is higher in the market but stable and hierarchies change more sluggishly. Both layers co-evolve (for example reflected by the rise of ICTs), but changes are more pronounced at the patent layer. Industries became more similar by patent citation patterns, but less by customer links in the market. This indicates a diversification process in the market while industrial R&D activities converge.
2. Both layers are subject to path-dependence and auto-correlation: Big industries grow faster and new links are more likely formed if a link was formed before. The

²⁴It is significantly higher in subset of links with significant changes as studied in Tab. SI.10, i.e. accounting for roughly 2 – 4 percentage points.

magnitude of link adjustments is larger if it was large before, for both links being formed and broken.

Links are more likely formed and broken to industries that are well connected, but the magnitude of change is smaller. Big industries are more likely to form new in- and output links, but links are less likely formed if the target industry j is big. This matches with the within-layer negative assortativity (Table 1): Well connected and big industries connect with industries with a lower degree and smaller in size.

Across layers, the industry size is negatively correlated with rewiring dynamics: Industries that are big in the market are less likely to change patent citation patterns and industries with many patents are less likely to revise their IO linkages.

3. Patent in- and output similarity is positively related to the probability that industries form mutual links. This matches with the observation that industrial R&D activities converged over the period of study (cf. Table A.1).

In contrast, similarity in the market appears to be a driver of industrial divergence: Similar industries are more likely to disconnect. At the disaggregate 6-digit level, it is also found that industries are less likely to form new links if they are similar.

Across layers, similarity seems to be a stabilizing factor: Industries are less likely to change market links if they are similar by patent citations and vice versa.

4. Differently from other studies [Acemoglu et al., 2016, Antony and Grebel, 2012], the analysis does not support an unambiguously positive effect of (knowledge) spillovers from technologically similar industries.

In the market, upstream spillovers exhibit a negative association with industry growth and with the probability that new input links are formed. One potential explanation for the negative impact of upstream spillovers is scarcity in markets: Input similarity in the market indicates that industries compete for the same type of physical inputs. High levels of upstream spillovers in the market indicates that many other (large) industries that rely on the same inputs. Competition may undermine the potential to grow.

The opposite effect is observed for downstream spillovers, i.e. industries that have a large overlap in customer links with other big industries grow and grow faster and are more likely to form and attract new customer links. Downstream

similarity may indicate both, substitutability associated with competition for customers or complementarity with positive mutual demand spillovers. Here, I find that complementarity dominates.

At the patent layer, the effects of up- and downstream spillovers are reverse: Upstream spills have a positive and downstream a negative association with industrial growth (and both negatively with link formation dynamics).

Patent upstream spillovers indicate that an industry relies on similar knowledge inputs as many other industries. This may create a positive externality. Differently from input competition in the market, intangible knowledge encoded in patents is non-rival.

Downstream spillovers imply that the inventions produced by an industry are similar to those by many other industries. This may indicate that the patents of an industry are not very radical since there are many other industries whose inventions serve similar purposes. Other research has shown that more radical innovations bear a higher potential for growth [Jaffe and De Rassenfosse, 2019].

These observations are more significant for the disaggregate 6-digit level data and for the subset of manufacturing industries.

5. Input diversification in the market, reflected in the in-degree, is positively associated with market size while the opposite is observed for output diversification. This pattern is reverse at the patent layer: Many diverse output links are positively related to patent growth but the citing patents from many diverse industries exhibits a negative association with industry growth.

A high level of output diversification in the patent layer may indicate a high value of patented knowledge: Citations from diverse users is positively correlated with the commercial value of a patent [Jaffe and De Rassenfosse, 2019], and inventions that are used by many different industries can be interpreted as general purpose technologies. This is one explanation for the positive relationship between the out-degree and industrial patent growth.

In contrast, in the market, industries specialized to serve the needs of few customers exhibit a higher potential to grow.

The differences across both network layers may be explained by a conceptual difference: Goods traded in the market are exhaustible but knowledge is not. Competition and scarcity are balancing mechanisms imposing a barrier to the cross-industrial

divergence by market size and industries that become too similar diverge in their technological trajectories.

In contrast, the use of intangible knowledge is non-rival and the formation of links to well connected industries is associated with access to diverse sources of technological knowledge. This may strengthen existing technological trajectories [cf. [Dosi, 1982](#)] and the cross-industrial distribution of innovative activities becomes increasingly skewed.

6.1. Demand-pull or technology-push: What is driving the direction of technological change?

Descriptively, both network layers co-evolve. But change in industrial hierarchies at the patent layer is more pronounced which could support the hypothesis of technology-push.

Direct cross-layer interactions are only weakly significant for the evolution of hierarchies and patterns of link formation when using aggregate data (see [Sec. 5.2](#)). In tendency, layers seem to be inversely related which is more significant at the 6-digit level. For example size in the market is negatively associated with size and link formation processes in the patent layer. Pairwise similarity seems to have a stabilizing effect across layers: Industries similar in the market are less likely to change mutual patent citation patterns and industries similar in the patent layer are less likely to revise mutual market links. Industries that are big by patents are less likely to acquire new market input connections in- and output links.

Within both layers, I find evidence for path-dependence: If a pair of industries connected before, it is more likely to further strengthen its link. Using more disaggregate data, I also observe a spillover effect from patent link formation to IO links: Industry pairs that formed citation links in the past, are more likely to form and less likely to break input links in the market. (The opposite holds for output links.) This offers support for the hypothesis that technology-push influences industrial capabilities to make productive use of inputs.

Regarding the connectivity, demand-pull exhibits a weak stabilizing effect on the technological evolution while technology-push may enable new upstream connections. However, the explanatory power of the regression analyses is rather low which is potentially due to dynamic industrial heterogeneity [[Pavitt, 1984](#)]. Some industries rely more on patents than others and likely, the relative importance demand-pull compared to technology-push is industry-specific.

Here, I associate demand-pull and technology-push with different layers of a network.

An alternative approach is the association of demand-pull with downstream and technology-push with upstream links within a network layer. It was observed that downstream links are more influential in the market and upstream links in the patent layer. Following this view, it could be said that demand matters more for the evolution of the market while supply-side push effects drive the evolution of patented innovation.

6.2. Limitations

Studying innovation and industrial evolution over time is challenging because of non-static classification systems. This analysis relies on industry codes that are purposely designed to describe industries by their production processes. I harmonized NAICS and SIC codes from different years using BEA concordance tables (see 4 and SI.1). NAICS is designed as a means for the description of industry and regularly (quasi-endogenously) updated to meet this purpose. This can be one explanation for the less skewed sector-size distribution, and possibly also for the relatively higher stability of the IO network. In the regression analysis, I control for industry and time FE hoping to capture potential distortions.

Aim of this analysis is the study of economic change where the use of intangible, patented knowledge is a means to enable economic production. This justifies the choice of NAICS codes instead of an approach based on patent classes [cf. Lybbert and Zolas, 2014, Van Looy et al., 2014]. A systematic, dynamic comparison between concordances based on patent classes and those based firm IDs (as used in this study or by Dorner and Harhoff [2018]) is an interesting avenue for future research. It would be also interesting to compare the results of this study with an approach using patent-classes as means of description: Classifying IO flows by their correspondence in patent-classes can be insightful to the impact of demand-pull on the dynamics of patented innovations. But this is beyond the scope of this paper.

Moreover, the two layers differ by the "level of classification": The IO data represents aggregate trade among all establishments whose primary activity is best described by the corresponding NAICS class description. Establishments are re-classified over time which changes the establishment composition of a NAICS industry. patent citation flows are derived at the firm-level while firms may have > 1 establishment whose production activities are assigned to different NAICS codes. Firms' NAICS class as patent-owner is assigned to their primary sector of activity and constant over time.²⁵ The impact of this difference across layers is difficult, if not impossible, to depict

²⁵An exception is made for patents that were acquired through M&A (see SI.1.1).

from the available data, but I expect this to be of minor importance, especially when using more aggregate digit levels. Of course, detailed firm-level time series of IO flows covering the whole economy are desirable, but these data do not exist.

The analysis in this paper studies demand-pull and technology-push at the aggregate level. But patterns of innovation, sources of knowledge and the relevance of patents as means of appropriation differ across single firms, industries and technology fields [Pavitt, 1984, Carlsson and Stankiewicz, 1991, Blank and Kappos, 2012]. The static dimension of sector heterogeneity is captured by the FE approach in the regressions. Moreover, the sample of industries is limited to those that have non-zero patent counts in all periods leading to an over-representation of manufacturing sectors.²⁶ Nevertheless, it should be kept in mind that the drivers of change in specific industries might be different from those analyzed in this study.

7. Conclusions

In this paper, demand-pull and technology-push effects arise from market dynamics and successful innovation in a networked economy. An empirical two-layer patent-IO network is compiled on US data covering more than 155 different industries classified by 4-digit NAICS codes in the period from 1976-2006. The patent network describes the evolution of technology and the IO network captures dynamics in the market. Interactions across layers capture the co-evolutionary nature of technological development and market needs. The coupled network is used to study the evolution of industry hierarchies and patterns of link formation as a manifestation of directed technological change.

To the best of my knowledge, this is the first paper that analyzes the co-evolution of technology and market networks at the detailed industry level covering a time period of three decades and all sectors of the US economy. Methodologically, this is a novel approach to map technology-push and demand-pull effects to data.

What is learned from this analysis? First, the analysis highlights a conceptual difference between the evolution of markets and innovation: While growth in the market is constrained by scarcity and competition, knowledge needed as input to innovation is non-rival. This can be a source of increasing returns leading to a skewed distribution. Second, and related to this: Relying on the same knowledge inputs as other big industries can be beneficial in the patent layer because this provides access to resources of knowledge. In contrast, being reliant on the same types of market

²⁶Additional robustness checks using the subset of manufacturing industries were made.

inputs as competitors may be unfavorable because it intensifies competition for scarce inputs. This can incite industries to explore other directions of research. Third, having many diverse users of knowledge is an indicator for high quality innovation with the potential for future growth. In contrast, having many diverse users of market inputs is associated with lower market growth in the same industry but with diversification into other sectors. This indicates different patterns of specialization: In the market, industries grow through output specialization and input diversification, while this is reversed in the patent layer.

Both demand-pull and technology-push have an impact on the direction of technological change. Increasing returns to innovation may strengthen existing trajectories, but in both layers industries tend to diverge if becoming too similar. Change in the market manifests more sluggishly than in patented innovation.

One important limitation needs to be mentioned that points to a relevant avenue for future research: Classification systems are endogenous and evolve over time [Kay et al., 2014]. Societal economic challenges manifest in the market layer and the language of industrial classifications seems appropriate to describe the economic evolution. The analysis shows that both layers co-evolve: Hence, policies that alter the conditions of competition in the market may spill over to the patent layer and vice versa. Increasing returns in patented innovation could be a source of technological lock-in effects. The analysis shows that market pressure may induce a redirection of R&D activities.

References

- Daron Acemoglu. Directed technical change. *The Review of Economic Studies*, 69(4): 781–809, 2002.
- Daron Acemoglu. *Introduction to modern economic growth*. Princeton University Press, 2008.
- Daron Acemoglu, Ufuk Akcigit, and William R Kerr. Innovation network. *Proceedings of the National Academy of Sciences*, page 201613559, 2016.
- Philippe Aghion and Peter Howitt. A model of growth through creative destruction. Technical report, National Bureau of Economic Research, 1990.
- Jürgen Antony and Thomas Grebel. Technology flows between sectors and their impact on large-scale firms. *Applied Economics*, 44(20):2637–2651, 2012.
- Daniele Archibugi and Alberto Coco. Measuring technological capabilities at the country level: A survey and a menu for choice. *Research policy*, 34(2):175–194, 2005. doi: 10.1016/j.respol.2004.12.002.
- Felix Arndt and Lamar Pierce. The behavioral and evolutionary roots of dynamic capabilities. *Industrial and Corporate Change*, 27(2):413–424, 2017. doi: 10.1093/icc/dtx042.
- Kenneth J Arrow. The economic implications of learning by doing. *The review of economic studies*, 29(3):155–173, 1962.
- W Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394):116–131, 1989. doi: 10.2307/2234208.
- Enghin Atalay, Ali Hortacsu, James Roberts, and Chad Syverson. Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202, 2011. doi: 10.1073/pnas.1015564108.
- Gary S Becker. Human capital revisited. In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*, pages 15–28. The university of Chicago press, 1994.
- Andrew B Bernard, J Bradford Jensen, and Peter K Schott. Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of us manufacturing plants.

- Journal of international Economics*, 68(1):219–237, 2006. doi: 10.1016/j.jinteco.2005.06.002.
- C Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st edn. 2006. corr. 2nd printing edn. *Springer, New York*, 2007.
- Rebecca M Blank and David J Kappos. Intellectual property and the US economy: Industries in focus. Economics and Statistics Administration & US Patent and Trademark Office, 2012.
- Johannes Boehm, Swati Dhingra, and John Morrow. The comparative advantage of firms. CEPR Discussion Paper No. DP13699, Centre for Economic Policy Research, 2019. URL <https://ssrn.com/abstract=3379615>. Accessed on 17 Dec. 2020.
- Erik Brynjolfsson and Andrew McAfee. *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee, 2012.
- Jie Cai, Nan Li, and Ana Maria Santacreu. Knowledge Diffusion, Trade and Innovation across Countries and Sectors. 2017.
- Benny Carlsson and Rikard Stankiewicz. On the nature, function and composition of technological systems. *Journal of Evolutionary Economics*, 1(2):93–118, 1991. doi: 10.1007/BF01224915.
- Vasco M Carvalho. From micro to macro via production networks. *Journal of Economic Perspectives*, 28(4):23–48, 2014. doi: 10.1257/jep.28.4.23.
- Vasco M Carvalho and Nico Voigtländer. Input diffusion and the evolution of production networks. Technical report, National Bureau of Economic Research, 2014.
- Wesley M Cohen. Fifty years of empirical studies of innovative activity and performance. In *Handbook of the Economics of Innovation*, volume 1, pages 129–213. Elsevier, 2010.
- Wesley M Cohen and Daniel A Levinthal. Innovation and learning: the two faces of R & D. *The economic journal*, 99(397):569–596, 1989. doi: 10.2307/2233763.
- Wesley M Cohen and Daniel A Levinthal. Absorptive capacity: A new perspective on learning and innovation. In *Strategic Learning in a Knowledge economy*, pages 39–67. Elsevier, 2000.

- Diego Comin and Bart Hobijn. An exploration of technology diffusion. *American Economic Review*, 100(5):2031–59, 2010. doi: 10.1257/aer.100.5.2031.
- Robin Cowan, Paul A David, and Dominique Foray. The explicit economics of knowledge codification and tacitness. *Industrial and corporate change*, 9(2):211–253, 2000. doi: 10.1093/icc/9.2.211.
- Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL <http://igraph.org>.
- Francesco Di Comite, d’Artis Kancs, et al. Macro-Economic Models for R&D and Innovation Policies-A Comparison of QUEST, RHOMOLO, GEM-E3 and NEMESIS. Technical report, Joint Research Centre (Seville site), 2015.
- Giada Di Stefano, Alfonso Gambardella, and Gianmario Verona. Technology push and demand pull perspectives in innovation studies: Current findings and future research directions. *Research Policy*, 41(8):1283–1295, 2012. doi: 10.1016/j.respol.2012.03.021.
- Matthias Dorner and Dietmar Harhoff. A novel technology-industry concordance table based on linked inventor-establishment data. *Research Policy*, 47(4):768–781, 2018. doi: 10.1016/j.respol.2018.02.005.
- Giovanni Dosi. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3):147–162, 1982.
- Giovanni Dosi. The research on innovation diffusion: An assessment. In *Diffusion of technologies and social behavior*, pages 179–208. Springer, 1991. doi: 10.1007/978-3-662-02700-4_7.
- Giovanni Dosi and Richard R Nelson. Technical change and industrial dynamics as evolutionary processes. In *Handbook of the Economics of Innovation*, volume 1, pages 51–127. Elsevier, 2010. doi: 10.1016/S0169-7218(10)01003-8.
- Amy C Edmondson, Ann B Winslow, Richard MJ Bohmer, and Gary P Pisano. Learning how and learning what: Effects of tacit and codified knowledge on performance improvement following technology adoption. *Decision Sciences*, 34(2):197–224, 2003. doi: 10.1111/1540-5915.02316.
- Nancy T Gallini. The economics of patents: Lessons from recent US patent reform. *Journal of Economic Perspectives*, 16(2):131–154, 2002. doi: 10.1257/0895330027292.

- Frank W Geels and Johan Schot. Typology of sociotechnical transition pathways. *Research policy*, 36(3):399–417, 2007. doi: 10.1016/j.respol.2007.01.003.
- Olivier Giovannoni. What do we know about the labor share and the profit share? Part I: Theories. 2014.
- Olivier Giovannoni, Lei Lu, Dam Nguyen, and Alex Xu. What do we know about the labor share and the profit share? Part II: Empirical studies. 2014.
- Bronwyn H Hall, Adam B Jaffe, and Manuel Trajtenberg. The NBER patent citation data file: Lessons, insights and methodological tools. National Bureau of Economic Research Working Paper w8498, 2001.
- Ricardo Hausmann and César A Hidalgo. The network structure of economic output. *Journal of Economic Growth*, 16(4):309–342, 2011.
- Sean Holly and Ivan Petrella. Factor demand linkages, technology shocks, and the business cycle. *Review of Economics and Statistics*, 94(4):948–963, 2012. doi: 10.1162/REST_a.00253.
- Karen Horowitz and Mark Planting. Concepts and Methods of the U.S. Input-Output Accounts. Measuring the Nation’s Economy. Bureau of Economic Analysis (BEA), U.S. Department of Commerce: Input-Output Manual, 2006. URL https://apps.bea.gov/papers/pdf/IOmanual_092906.pdf.
- Kerstin Hötte. Data publication: Demand-pull and technology-push: What drives the direction of technological change? Empirical data on a coupled two-layer input-output and patent-citation network. Data Publication, Bielefeld University, 2021a.
- Kerstin Hötte. Skill transferability and the stability of transition pathways- a learning-based explanation for patterns of diffusion. *Journal of Evolutionary Economics*, forthcoming, 2021b. doi: 10.1007/s00191-020-00710-7.
- Jingong Jim Huang. Technology Network, Innovation and Growth. In *2018 Meeting Papers*, number 178. Society for Economic Dynamics, 2018.
- IPCC. Summary for Policymakers. In V Masson-Delmotte, P Zhai, HO Pörtner, D Roberts, J Skea, P R Shukla, A Pirani, W Moufouma-Okia, C Péan, R Pidcock, S Connors, J B R. Matthews, Y Chen, X Zhou, M I Gomis, E Lonnoy, T Maycock, M Tignor, and T Waterfield, editors, *An IPCC Special Report on the impacts of global warming of 1.5 degree C above pre-industrial levels and related global greenhouse gas*

- emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, Summary for Policymakers.* World Meteorological Organization, Geneva, Switzerland, 2018.
- Matthew O Jackson. *Social and economic networks.* Princeton university press, 2008.
- Adam B Jaffe and Gaétan De Rassenfosse. Patent citation data in social science research: Overview and best practices. *Journal of the Association for Information Science and Technology*, 68(6):1360–1374, 2017. doi: 10.1002/asi.23731.
- Adam B Jaffe and Gaétan De Rassenfosse. Patent citation data in social science research: Overview and best practices. In *Research handbook on the economics of intellectual property law.* Edward Elgar Publishing, 2019. doi: 10.4337/9781789903997.00043.
- Luciano Kay, Nils Newman, Jan Youtie, Alan L Porter, and Ismael Rafols. Patent overlay mapping: Visualizing technological distance. *Journal of the Association for Information Science and Technology*, 65(12):2432–2443, 2014.
- Stephen J Kline and Nathan Rosenberg. An overview of innovation. In *The positive sum strategy: Harnessing technology for economic growth.* The National Academy of Science, USA, 1986.
- Samuel Kortum and Jonathan Putnam. Assigning patents to industries: tests of the Yale technology concordance. *Economic Systems Research*, 9(2):161–176, 1997.
- Sanjaya Lall. Technological capabilities and industrialization. *World development*, 20(2):165–186, 1992. doi: 10.1016/0305-750X(92)90097-F.
- Loet Leydesdorff. Similarity measures, author cocitation analysis, and information theory. *Journal of the American society for Information Science and Technology*, 56(7):769–772, 2005.
- Bengt-åke Lundvall and Björn Johnson. The learning economy. *Journal of industry studies*, 1(2):23–42, 1994. doi: 10.1080/13662719400000002.
- Travis J Lybbert and Nikolas J Zolas. Getting patents and economic data to speak to each other: An ‘algorithmic links with probabilities’ approach for joint analyses of patenting and economic activity. *Research Policy*, 43(3):530–542, 2014. doi: 10.1016/j.respol.2013.09.001.
- Bruce McCune, James B Grace, and Dean L Urban. *Analysis of ecological communities*, volume 28. MjM software design Gleneden Beach, OR, 2002.

- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- Richard R Nelson and Edmund S Phelps. Investment in humans, technological diffusion, and economic growth. *American Economic Review*, 56(1/2):69–75, 1966. URL www.jstor.org/stable/1821269.
- Gregory F Nemet. Demand-pull, technology-push, and government-led incentives for non-incremental technical change. *Research policy*, 38(5):700–709, 2009. doi: 10.1016/j.respol.2009.01.004.
- Keith Pavitt. Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13:343–373, 1984. doi: 10.1016/0048-7333(84)90018-0.
- David Popp, Richard G Newell, and Adam B Jaffe. Energy, the environment, and technological change. In *Handbook of the Economics of Innovation*, volume 2, pages 873–937. Elsevier, 2010.
- Robert A Rigby and D Mikis Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554, 2005. doi: 10.1111/j.1467-9876.2005.00510.x.
- Joeri Rogelj, Michel Den Elzen, Niklas Höhne, Taryn Fransen, Hanna Fekete, Harald Winkler, Roberto Schaeffer, Fu Sha, Keywan Riahi, and Malte Meinshausen. Paris Agreement climate proposals need a boost to keep warming well below 2 C. *Nature*, 534(7609):631, 2016. doi: 10.1038/nature18307.
- Paul M Romer. Endogenous technological change. *Journal of political Economy*, 98(5, Part 2):71–102, 1990.
- Vernon W Ruttan. Usher and Schumpeter on invention, innovation, and technological change. *The Quarterly Journal of Economics*, pages 596–606, 1959. doi: 10.2307/1884305.
- Ulrich Schmoch, Françoise Laville, Pari Patel, and Rainer Frietsch. Linking technology areas to industrial sectors. Final Report to the European Commission, DG Research, European Commission, 2003.
- Will Steffen, Johan Rockström, Katherine Richardson, Timothy M Lenton, Carl Folke, Diana Liverman, Colin P Summerhayes, Anthony D Barnosky, Sarah E Cornell, Michel Crucifix, et al. Trajectories of the Earth System in the Anthropocene.

- Proceedings of the National Academy of Sciences*, 115(33):8252–8259, 2018. doi: 10.1073/pnas.1810141115.
- Peter Thompson. The relationship between unit cost and cumulative quantity and the evidence for organizational learning-by-doing. *Journal of Economic perspectives*, 26(3):203–24, 2012.
- US Census Bureau. The North American Industrial Classification System, Manual. Technical report, United States Census Bureau, Executive office of the president office of management and budget, url=census.gov/naics, 2017.
- Bart Van Looy, Caro Vereyen, and Ulrich Schmoch. Patent Statistics: Concordance IPC V8–NACE Rev. 2. Report, Eurostat, European Commission, 2014.
- T Wiesenthal, P Dowling, J Morbee, C Thiel, B Schade, P Russ, S Simoes, S Petevs, K Schoots, M Londo, et al. Technology learning curves for energy policy support. *JRC scientific and policy reports*, 332, 2012.

Part I.

APPENDIX

A. Additional statistics and results

A.1. Additional descriptive information

	Input-output			Patent			Combined		
	1981	1991	2006	1981	1991	2006	1981	1991	2006
<i>Upstream network</i>									
Density	0.39	0.39	0.39	0.34	0.42	0.44	0.36	0.41	0.42
Avg. degree	60.21	60.41	59.96	52.36	64.77	67.32	55.34	62.53	64.53
Avg. strength	13.25	9.70	4.16	2.62	10.40	15.92	5.10	9.13	12.14
Avg. weight	27.32	23.35	21.79	15.60	24.49	28.34	1460.32	1970.94	2282.58
Diameter	0.14	0.16	0.08	0.03	0.05	0.10	0.07	0.08	0.13
Avg. dist.	1.00	1.00	1.00	1.15	1.04	1.01	1.00	1.00	1.00
<i>Downstream network</i>									
Density	0.40	0.38	0.36	0.39	0.46	0.48	0.37	0.45	0.44
Avg. degree	60.93	58.39	56.13	59.41	70.54	74.03	56.44	68.93	67.08
Avg. strength	11.50	10.52	4.14	4.54	13.51	15.34	4.67	13.95	12.59
Avg. weight	24.27	22.40	18.73	23.10	41.06	42.67	1976.58	3347.28	3257.67
Diameter	0.05	0.03	0.04	0.07	0.13	0.23	0.07	0.08	0.24
Avg. dist.	1.01	1.01	1.00	1.04	1.00	1.00	1.00	1.00	1.00

Notes: These statistics are computed on the weighted upstream networks. For an introduction to network characteristics see Jackson [2008], Csardi and Nepusz [2006]. "Assort." is the abbreviation for assortativity. Data: 4-digit balanced panel, cosine similarity network.

Table A.1: Aggregate cosine similarity network statistics over time (upstream).

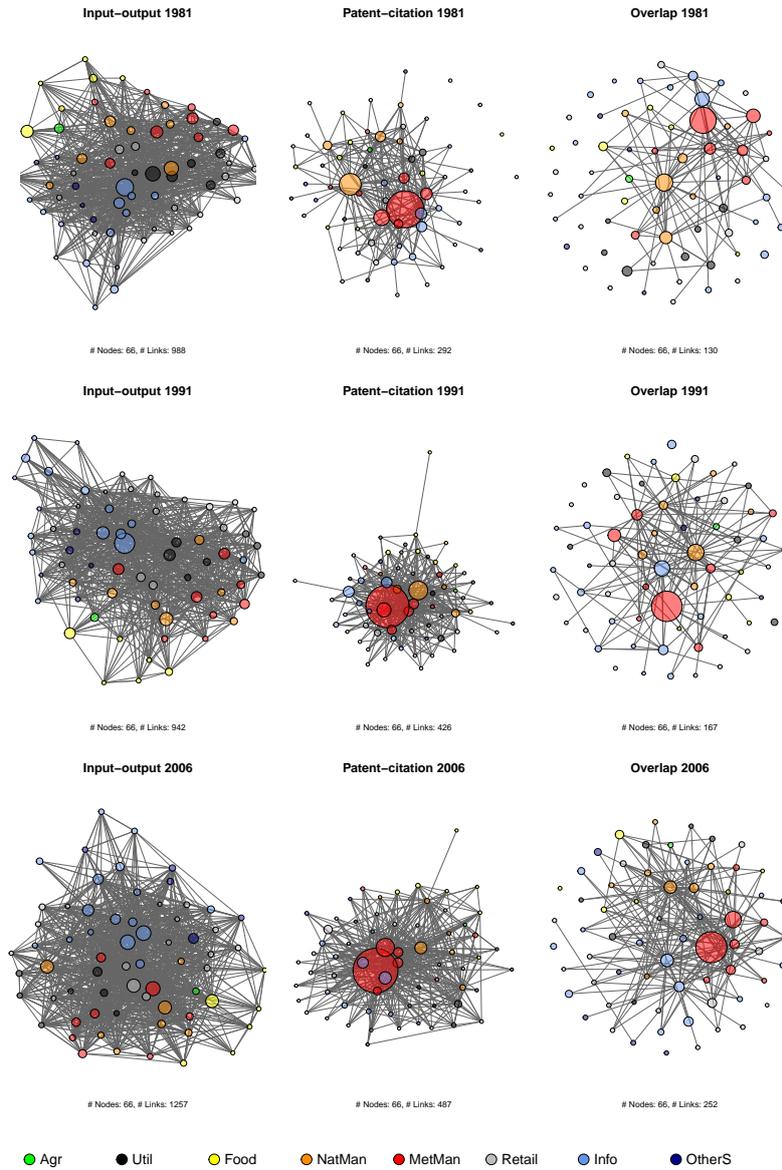
Table A.1 summarizes the network characteristics of the cosine (upstream) similarity network. The network is given by the symmetric $N \times N$ cosine similarity matrix $\Sigma_t^{\alpha, in}$ where the pairwise input similarities $\sigma_{ij,t}^{\alpha, in}$ are the weights of a link connecting i and j .

For the purpose of calculating aggregate network statistics and plotting the network, two nodes i, j are shown as being connected in the similarity network in period t if their pairwise similarity is higher than the average of pairwise similarity of all sectors and all periods $\sigma_{ij,t}^{\alpha, out} > \frac{1}{|N|(|N|-1) \cdot |T|} \sum_{i \in N} \sum_{j \in N, j \neq i} \sum_{t \in T} \sigma_{ij,t}^{\alpha, out}$ with $T = 1981, 1991, 2006$. The density in the patent citation similarity network is increasing which suggests that industries became more similar in their up- and downstream patent citation behavior. Over time, the similarity of industries by their input-linkages has decreased by output similarity which may suggest specialization in the market.

In Figure [A.1](#) and [A.2](#), additional network illustrations are provided showing the input network for the balanced panel of 3-digit data and for the unbalanced panel of 4-digit data. The isolated nodes in Fig. [A.2](#) are industries only rely on inputs produced by themselves, i.e. physical inputs produced within the same 4-digit industry and self-citations.

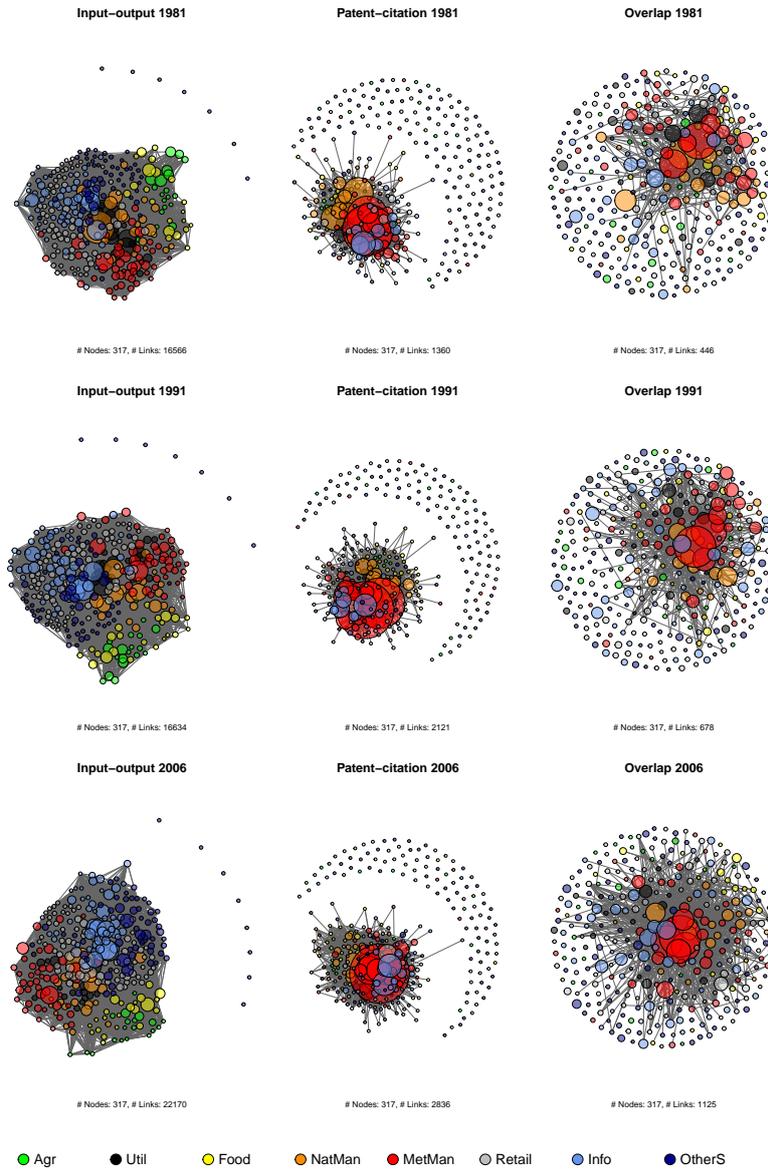
Figure [A.3](#) shows the pairwise correlation of different indicators used to describe the network across and within layers. The figure illustrates two observations: (1) the degree is least correlated with other variables, and (2) variable computed on the basis of in- and out-going links are highly correlated.

In Table [A.2](#) and [A.3](#), the sector rankings by sum of output and by patent stock are shown for the Top-50 largest industries.



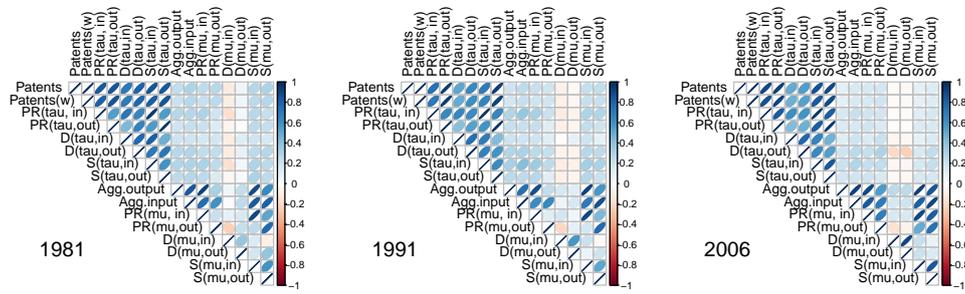
Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only in-going links are considered. Plots of the downstream network are available in [SI.2](#). Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Agr is Agriculture (1); Util are Mining and Utilities (2); Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33); Retail are Trade, Retail and Transportation sectors (4); Info are Information, Financial, Management and Administrative Services (5); OtherS are Other services and Public Sector (6-9). Data: Balanced panel of 3-digit industries.

Figure A.1: Network plots at different periods for 3-digit level data.



Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only in-going links are considered. Plots of the downstream network are available in [SI.2](#). Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Agr is Agriculture (1); Util are Mining and Utilities (2); Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33); Retail are Trade, Retail and Transportation sectors (4); Info are Information, Financial, Management and Administrative Services (5); OtherS are Other services and Public Sector (6-9). Data: Unbalanced panel of 4-digit industries.

Figure A.2: Network plots at different periods for 4-digit level data (unbalanced).



Notes: This figure shows a correlation plot between different pairs of indicators. PR is short for PageRank, D for degree, S for strength. $Patents(w)$ is the weighted patent stock. The correlation at the diagonal is by definition equal one. The colors and the shape of the ellipses indicate the strength of correlation. Data: 4-digit, balanced panel.

Figure A.3: Pairwise correlations of network indicators

Top 50 industries by Aggr. output ($A_{i,t}^{\mu}$)						
1981			1991		2006	
1	Petroleum & Coal Prod.	3241 11.07	Elctr. Power Gen., Transm. & Distr.	2211 7.07	Mngm. Comp. & Enterpr.	5511 9.19
2	Elctr. Power Gen., Transm. & Distr.	2211 9.03	Motor Vhcl. Parts Mnft.	3363 6.10	Lessors of Real Estate	5311 5.89
3	Motor Vhcl. Parts Mnft.	3363 6.00	Petroleum & Coal Prod.	3241 5.87	Petroleum & Coal Prod.	3241 5.80
4	Architect., Engineer. & Srv.	5413 4.78	Architect., Engineer. & Srv.	5413 5.13	Motor Vhcl. Parts Mnft.	3363 5.20
5	Support Act. for Min.	2131 4.32	Basic Chem. Manufactur.	3251 4.36	Oil & Gas Extraction	2111 4.18
6	Oil & Gas Extraction	2111 3.90	Lessors of Real Estate	5311 3.73	Elctr. Power Gen., Transm. & Distr.	2211 3.70
7	Pulp, & Paper Mills	3221 3.09	Plastics Prod. Manufact.	3261 3.55	Basic Chem. Manufactur.	3251 3.70
8	Basic Chem. Manufactur.	3251 3.02	Semicond. & Oth. Elctr.	3344 3.32	Insurance Carriers	5241 3.60
9	Anim. Slaughter. & Prc.	3116 2.82	Pulp, & Paper Mills	3221 3.24	Plastics Prod. Manufact.	3261 3.45
10	Converted Paper Prod.	3222 2.74	Print. & Support Act.	3231 3.04	Semicond. & Oth. Elctr.	3344 3.33
11	Plastics Prod. Manufact.	3261 2.65	Converted Paper Prod.	3222 2.92	Oth. Fin. Investm. Act.	5239 3.32
12	Resin, Synth. Rubber & Fibers	3252 2.63	Resin, Synth. Rubber & Fibers	3252 2.63	Architect., Engineer. & Srv.	5413 3.13
13	Utility System Constr.	2371 2.41	Support Act. for Min.	2131 2.41	Mngm., Scient. & Tech. Consult. Srv.	5416 2.64
14	Other Crop Farming	1119 2.40	Anim. Slaughter. & Prc.	3116 2.31	Securities & Commodity Contracts Intermed. & Brokerage	5231 2.51
15	Lessors of Real Estate	5311 2.36	Mngm., Scient. & Tech. Consult. Srv.	5416 2.20	Print. & Support Act.	3231 2.20
16	Print. & Support Act.	3231 2.30	Oil & Gas Extraction	2111 2.15	Iron & Steel Mills & Ferroalloy	3311 2.19
17	Oilseed & Grain Farming	1111 2.21	Oth. Wood Prod. Manufact.	3219 2.05	Lessors of Nonfin. Intangible Assets (exc. Copyrighted Works)	5331 2.17
18	Semicond. & Oth. Elctr.	3344 2.21	Oth. Fin. Investm. Act.	5239 2.00	Anim. Slaughter. & Prc.	3116 2.00
19	Iron & Steel Mills & Ferroalloy	3311 2.08	Fabricated Metal Prod.	3329 2.00	Resin, Synth. Rubber & Fibers	3252 1.95
20	Oth. Wood Prod. Manufact.	3219 2.02	Utility System Constr.	2371 1.92	Converted Paper Prod.	3222 1.88
21	Nonferrous Metal (exc. Aluminum) Prod. & Prc.	3314 2.02	Other Crop Farming	1119 1.89	Computer Syst. Design & Srv.	5415 1.87
22	Fabricated Metal Prod.	3329 1.98	Srv. to Build.s & Dwell.s	5617 1.83	General Freight Trucking	4841 1.83
23	Steel Prod. from Purch. Steel	3312 1.97	Architect. & Struct. Metals	3323 1.70	Pulp, & Paper Mills	3221 1.81
24	Architect. & Struct. Metals	3323 1.91	Oilseed & Grain Farming	1111 1.69	Nondep. Cred. Intermed.	5222 1.78
25	Grain & Oilseed Milling	3112 1.72	Nonferrous Metal (exc. Aluminum) Prod. & Prc.	3314 1.57	Wireless Telecomm. Carriers (exc. Satellite)	5172 1.72
26	Boiler, Tank & Shipp. Container	3324 1.70	Chem. Prod. & Prepar.	3259 1.56	Pharm. & Medicine	3254 1.69
27	Cem. & Concrete Prod.	3273 1.52	Nondep. Cred. Intermed.	5222 1.53	Motor Vehicle Manufactur.	3361 1.59
28	Natural Gas Distribution	2212 1.51	Computer Syst. Design & Srv.	5415 1.51	Srv. to Build.s & Dwell.s	5617 1.57
29	Aluminum Prod. & Process.	3313 1.43	Securities & Commodity Contracts Intermed. & Brokerage	5231 1.49	Scientific R & D Serv.	5417 1.54
30	Residential Build. Constr.	2361 1.29	Cem. & Concrete Prod.	3273 1.41	Architect. & Struct. Metals	3323 1.53
31	Fabric Mills	3132 1.29	Natural Gas Distribution	2212 1.39	Aerospace Prod. & Parts	3364 1.45
32	Chem. Prod. & Prepar.	3259 1.28	Boiler, Tank & Shipp. Container	3324 1.33	Comm. Equ. & Suppl. Wholes.	4234 1.42
33	Electrical Equ. Manufact.	3353 1.27	Iron & Steel Mills & Ferroalloy	3311 1.30	Natural Gas Distribution	2212 1.40
34	Dairy Product Manufactur.	3115 1.26	Grain & Oilseed Milling	3112 1.29	Grocery & Prod. Wholes.	4244 1.31
35	Coal Mining	2121 1.19	Steel Prod. from Purch. Steel	3312 1.24	Oilseed & Grain Farming	1111 1.29
36	Gen. Purpose Machinery	3339 1.19	Aluminum Prod. & Process.	3313 1.20	Machin., Equ. & Suppl. Wholes.	4238 1.23
37	Nonresidential Build. Constr.	2362 1.07	Electrical Equ. Manufact.	3353 1.19	Fabricated Metal Prod.	3329 1.23
38	Pestic., Fertl. & Agr. Chem.	3253 1.07	Wireless Telecomm. Carriers (exc. Satellite)	5172 1.17	Motion Pict. & Video Industries	5121 1.18
39	Srv. to Build.s & Dwell.s	5617 1.06	Grocery & Prod. Wholes.	4244 1.14	Machine Shops, Turned Prod. & Screw, Nut & Bolt	3327 1.17
40	Paint, Coat. & Adhesive	3255 1.06	Residential Build. Constr.	2361 1.14	Internet Srv. Providers & Web Search Portals	5181 1.15
41	Mngm., Scient. & Tech. Consult. Srv.	5416 1.04	Dairy Product Manufactur.	3115 1.14	Business Support Services	5614 1.15
42	Glass & Glass Prod. Mnft.	3272 1.03	Other Food Manufacturing	3119 1.13	Dairy Product Manufactur.	3115 1.12
43	Other Food Manufacturing	3119 1.02	Paint, Coat. & Adhesive	3255 1.13	Cem. & Concrete Prod.	3273 1.10
44	Grocery & Prod. Wholes.	4244 1.00	Oth. Misc. Manufactur.	3399 1.12	Data Prc., Host. & Serv.	5182 1.08
45	Rubber Prod. Manufactur.	3262 0.97	General Freight Trucking	4841 1.09	Grain & Oilseed Milling	3112 1.07
46	Heat., AC & Comm. Refrig.	3334 0.96	Measur., Elctrmed. & Ctrl. Instr.	3345 1.07	Nonferrous Metal (exc. Aluminum) Prod. & Prc.	3314 1.03
47	Nondep. Cred. Intermed.	5222 0.92	Fabric Mills	3132 1.04	Misc. Dur. Goods Wholes.	4239 1.01
48	Oth. Fin. Investm. Act.	5239 0.91	Aerospace Prod. & Parts	3364 1.03	Aluminum Prod. & Process.	3313 0.98
49	Cutlery & Handtool Mnft.	3322 0.90	Act. to Cred. Intermed.	5223 1.01	Computer & Periph. Equ.	3341 0.94
50	Oth. Misc. Manufactur.	3399 0.88	Glass & Glass Prod. Mnft.	3272 0.99	Measur., Elctrmed. & Ctrl. Instr.	3345 0.94
Quartiles:		0.33, 0.6, 1.0625	0.36, 0.65, 1.14	0.26, 0.56, 1.16		

Notes: Each period shows a block of three columns indicating (1) the industry name, (2) the 4-digit NAICS code, (3) the normalized industry market size $A_{i,t}^{\mu}$. The average size normalized size equals one. Data: 4-digit level balanced panel. The data are averages across the sub-periods 1976-1981, 1986-1991, 1996-2006.

Table A.2: Ranking of the Top-50 industries by aggregate output $A_{i,t}^{\mu}$

Top 50 industries by Patent stock ($A_{i,t}^+$)						
	1981		1991		2006	
1	Computer & Periph. Equ.	3341 12.81	Semicond. & Oth. Electr.	3344 21.73	Semicond. & Oth. Electr.	3344 25.52
2	Semicond. & Oth. Electr.	3344 12.79	Computer & Periph. Equ.	3341 17.18	Computer & Periph. Equ.	3341 20.60
3	Pharm. & Medicine	3254 11.14	Communications Equ. Mnft.	3342 8.55	Communications Equ. Mnft.	3342 13.57
4	Petroleum & Coal Prod.	3241 8.70	Pharm. & Medicine	3254 8.55	Computer Syst. Design & Srv.	5415 8.55
5	Communications Equ. Mnft.	3342 5.99	Computer Syst. Design & Srv.	5415 6.29	Measur., Elctrmed. & Ctrl. Instr.	3345 7.28
6	Aerospace Prod. & Parts	3364 5.90	Measur., Elctrmed. & Ctrl. Instr.	3345 6.23	Pharm. & Medicine	3254 6.29
7	Motor Vehicle Manufactur.	3361 5.30	Petroleum & Coal Prod.	3241 6.15	Motor Vehicle Manufactur.	3361 5.47
8	Measur., Elctrmed. & Ctrl. Instr.	3345 5.27	Motor Vehicle Manufactur.	3361 5.14	Magnetic & Optical Media	3346 5.19
9	Computer Syst. Design & Srv.	5415 5.03	Chem. Prod. & Prepar.	3259 4.32	Motor Vhcl. Body & Trailer	3362 4.21
10	Chem. Prod. & Prepar.	3259 4.21	Resin, Synth. Rubber & Fibers	3252 4.00	Taxi & Limousine Service	4853 3.37
11	Basic Chem. Manufactur.	3251 4.03	Aerospace Prod. & Parts	3364 3.99	Chem. Prod. & Prepar.	3259 3.33
12	Heat., AC & Comm. Refrig.	3334 3.87	Magnetic & Optical Media	3346 3.34	Boiler, Tank & Shipp. Container	3324 3.19
13	Resin, Synth. Rubber & Fibers	3252 3.77	Motor Vhcl. Body & Trailer	3362 2.95	Wireless Telecomm. Carriers (exc. Satellite)	5172 2.68
14	Wireless Telecomm. Carriers (exc. Satellite)	5172 3.38	Support Act. for Min.	2131 2.71	Resin, Synth. Rubber & Fibers	3252 2.37
15	Oil & Gas Extraction	2111 3.07	Wireless Telecomm. Carriers (exc. Satellite)	5172 2.68	Aerospace Prod. & Parts	3364 2.31
16	Motor Vhcl. Body & Trailer	3362 3.00	Fabricated Metal Prod.	3329 2.63	Audio & Video Equ. Mnft.	3343 2.10
17	Fabricated Metal Prod.	3329 2.80	Electrical Equ. Manufact.	3353 2.46	Support Act. for Min.	2131 2.06
18	Electrical Equ. Manufact.	3353 2.79	Heat., AC & Comm. Refrig.	3334 2.20	Petroleum & Coal Prod.	3241 2.04
19	Agr., Constr. & Min. Machin.	3331 2.78	Boiler, Tank & Shipp. Container	3324 2.14	Comm. & Srv. Ind. Machin.	3333 2.03
20	Soap, Clean. & Toilet Prepar.	3256 2.73	Audio & Video Equ. Mnft.	3343 2.14	Fabricated Metal Prod.	3329 1.89
21	Gen. Purpose Machinery	3339 2.58	Pulp, & Paper Mills	3221 2.11	Electrical Equ. Manufact.	3353 1.69
22	Support Act. for Min.	2131 2.57	Medical Equ. & Supplies	3391 1.97	Pulp, & Paper Mills	3221 1.66
23	Paint, Coat. & Adhesive	3255 1.98	Taxi & Limousine Service	4853 1.87	Soap, Clean. & Toilet Prepar.	3256 1.52
24	Pulp, & Paper Mills	3221 1.86	Soap, Clean. & Toilet Prepar.	3256 1.79	Agr., Constr. & Min. Machin.	3331 1.47
25	Magnetic & Optical Media	3346 1.66	Comm. & Srv. Ind. Machin.	3333 1.76	Oth. Electrical Equ.	3359 1.41
26	Converted Paper Prod.	3222 1.62	Basic Chem. Manufactur.	3251 1.72	Investm. Pools & Funds	5259 1.40
27	Internet Srv. Providers & Web Search Portals	5181 1.56	Paint, Coat. & Adhesive	3255 1.70	Motion Pict. & Video Industries	5121 1.32
28	Nondep. Cred. Intermed.	5222 1.51	Oth. Electrical Equ.	3359 1.48	Heat., AC & Comm. Refrig.	3334 1.26
29	Boiler, Tank & Shipp. Container	3324 1.41	Motion Pict. & Video Industries	5121 1.47	Internet Srv. Providers & Web Search Portals	5181 1.25
30	Medical Equ. & Supplies	3391 1.36	Converted Paper Prod.	3222 1.40	Medical Equ. & Supplies	3391 1.06
31	Audio & Video Equ. Mnft.	3343 1.33	Agr., Constr. & Min. Machin.	3331 1.37	Household Appl. Manufact.	3352 1.02
32	Cutlery & Handtool Mnft.	3322 1.27	Oil & Gas Extraction	2111 1.22	Paint, Coat. & Adhesive	3255 0.91
33	Securities & Commodity Contracts Intermed. & Brokerage	5231 1.24	Household Appl. Manufact.	3352 1.16	Gen. Purpose Machinery	3339 0.76
34	Comm. & Srv. Ind. Machin.	3333 1.22	Gen. Purpose Machinery	3339 1.06	Converted Paper Prod.	3222 0.74
35	Oth. Electrical Equ.	3359 1.17	Nondep. Cred. Intermed.	5222 0.99	Cutlery & Handtool Mnft.	3322 0.74
36	Motion Pict. & Video Industries	5121 1.16	Internet Srv. Providers & Web Search Portals	5181 0.95	Electric Light. Equ.	3351 0.69
37	Rubber Prod. Manufactur.	3262 1.15	Securities & Commodity Contracts Intermed. & Brokerage	5231 0.90	Lessors of Nonfin. Intangible Assets (exc. Copyrighted Works)	5331 0.57
38	Gasoline Stations	4471 1.06	Investm. Pools & Funds	5259 0.86	Oil & Gas Extraction	2111 0.56
39	Household Appl. Manufact.	3352 1.05	Gasoline Stations	4471 0.86	Oth. Transportation Equ.	3369 0.54
40	Glass & Glass Prod. Mnft.	3272 0.83	Cutlery & Handtool Mnft.	3322 0.84	Plastics Prod. Manufact.	3261 0.52
41	Oth. Textile Prod. Mills	3149 0.80	Electr. & Electr. Goods Wholes.	4236 0.83	Electr. & Electr. Goods Wholes.	4236 0.51
42	Plastics Prod. Manufact.	3261 0.80	Oth. Textile Prod. Mills	3149 0.75	Basic Chem. Manufactur.	3251 0.45
43	Taxi & Limousine Service	4853 0.78	Electric Light. Equ.	3351 0.70	Oth. Textile Prod. Mills	3149 0.43
44	Motor Vhcl. & Suppl. Wholes.	4231 0.73	Plastics Prod. Manufact.	3261 0.63	Industrial Machinery	3332 0.40
45	Drugs & Druggists' Sundries Wholes.	4242 0.60	Rubber Prod. Manufactur.	3262 0.54	Scientific R & D Serv.	5417 0.39
46	Electr. & Electr. Goods Wholes.	4236 0.58	Tobacco Manufacturing	3122 0.42	Nondep. Cred. Intermed.	5222 0.38
47	Electric Light. Equ.	3351 0.56	Glass & Glass Prod. Mnft.	3272 0.40	Securities & Commodity Contracts Intermed. & Brokerage	5231 0.38
48	Pestic., Fertl. & Agr. Chem.	3253 0.50	Lessors of Nonfin. Intangible Assets (exc. Copyrighted Works)	5331 0.40	Beverage Manufacturing	3121 0.35
49	Architect., Engineer. & Srv.	5413 0.48	Aluminum Prod. & Process.	3313 0.36	Medical & Diagnostic Lab.	6215 0.35
50	Investm. Pools & Funds	5259 0.47	Scientific R & D Serv.	5417 0.36	Motor Vhcl. Parts Mnft.	3363 0.31
Quartiles:						
	0.03, 0.13, 1.055		0.02, 0.095, 0.89		0.02, 0.12, 0.74	

Notes: Each period shows a block of three columns indicating (1) the industry name, (2) the 4-digit NAICS code, (3) the normalized patent stock size $A_{i,t}^+$. The average size normalized size equals one. Data: 4-digit level balanced panel. The data are averages across the sub-periods 1976-1981, 1986-1991, 1996-2006.

Table A.3: Ranking of the Top-50 industries by patent stock $A_{i,t}^+$

A.2. Regressions to explain the evolution of sector hierarchies

Here, I provide additional results of the hierarchy regression using other data compared to those in the text. The results for a data subset that covers manufacturing industries only are summarized in Table A.4. Table A.5 shows the results for a balanced panel at the 6-digit level of aggregation.

	market layer				Patent-layer			
	(1) $gr(A_{i,t}^\mu)$	(2) $A_{i,t}^\mu$	(3) $gr(A_{i,t}^\mu)$	(4) $A_{i,t}^\mu$	(5) $gr(A_{i,t}^\tau)$	(6) $A_{i,t}^\tau$	(7) $gr(A_{i,t}^\tau)$	(8) $A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
$A_{i,t-1}^\mu$	0.2421*** (0.0222)	0.2603* (0.1102)	0.2261*** (0.0227)	0.3157** (0.1131)	-0.0067 (0.014)	0.0108 (0.0135)	-6e-04 (0.0131)	-0.0012 (0.0137)
$A_{i,t-1}^\tau$	0.0251 (0.0691)	-0.0103 (0.0725)	-0.0151 (0.0725)	-0.0971 (0.0657)	0.0581 (0.0346)	0.6924*** (0.0609)	0.0325 (0.0293)	0.6871*** (0.0591)
$D_{i,t-1}^{\mu,d}$	0.0166 (0.0156)	0.082** (0.0251)	-0.0442*** (0.0126)	-0.0871*** (0.0209)	2e-04 (0.0129)	-0.0025 (0.0068)	9e-04 (0.0093)	-0.0113* (0.005)
$D_{i,t-1}^{\tau,d}$	0.0461 (0.0305)	0.0013 (0.0289)	0.0637** (0.0222)	0.008 (0.029)	-0.0826 (0.0447)	0.0062 (0.0096)	0.1039** (0.034)	0.0251* (0.0103)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0508** (0.0189)	-0.0576** (0.0198)	0.0169 (0.0194)	-0.0047 (0.0149)	0.0118 (0.0147)	0.0203* (0.0086)	-0.0053 (0.0146)	-0.0038 (0.0096)
$Spill(A)_{i,t-1}^{\tau,d}$	0.0064 (0.0181)	-0.0218 (0.0194)	0.0083 (0.0108)	0.0355* (0.0161)	0.0024 (0.0234)	0.0089 (0.0064)	-0.0502** (0.0177)	-0.0043 (0.0041)
R^2	0.386	0.1677	0.405	0.2099	0.0286	0.7274	0.074	0.7285
N	450	450	450	450	450	450	450	450
Average	-0.0043	0.5871	-0.0043	0.5871	-0.0446	0.4423	-0.0446	0.4423

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Notes: The model is estimated as linear OLS model including time and industry FE and using two-ways clustered SE. The explanatory variables have been selected in a step-wise selection procedure searching for the highest explanatory power measured by the BIC (see text for further detail). Regressions are run separately for spillovers and network-based variables computed on the basis of upstream ($d = in$) and downstream ($d = out$) links. Data: 4-digit, balanced panel of the subset of manufacturing industries.

Table A.4: Regression results explaining the evolution of industrial hierarchies (manufacturing).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
$A_{i,t-1}^\mu$	0.1699*** (0.0167)	0.548*** (0.0299)	0.1679*** (0.0168)	0.5618*** (0.031)	-0.0241** (0.0076)	-7e-04 (0.0038)	-0.017* (0.0075)	8e-04 (0.0039)
$A_{i,t-1}^\tau$	-0.0144 (0.0271)	0.0167 (0.039)	-0.0374 (0.0285)	0.0022 (0.0382)	0.0592*** (0.0171)	0.6327*** (0.0294)	0.0353* (0.0162)	0.6321*** (0.0292)
$D_{i,t-1}^{\mu,d}$	0.0569*** (0.0104)	0.0809*** (0.0111)	-0.0505*** (0.0069)	-0.0355*** (0.0104)	0.0286*** (0.0086)	0.0073** (0.0028)	-0.0091 (0.0055)	-0.0026 (0.0018)
$D_{i,t-1}^{\tau,d}$	0.0089 (0.0184)	0.0083 (0.0172)	0.012 (0.0118)	0.0123 (0.016)	-0.0659*** (0.0169)	0.0053 (0.0042)	0.1434*** (0.0163)	0.0167*** (0.0033)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0481*** (0.0096)	-0.0102 (0.0091)	0.0482*** (0.0131)	0.0563*** (0.0147)	0.0024 (0.008)	0.0028 (0.0028)	0.0106 (0.0095)	3e-04 (0.0038)
$Spill(A)_{i,t-1}^{\tau,d}$	0.0039 (0.0105)	0.026* (0.0121)	0.0067 (0.0112)	0.0301** (0.0094)	0.0842*** (0.0153)	0.009** (0.0029)	-0.0474** (0.0172)	-0.0035 (0.0026)
R^2	0.2437	0.2766	0.2619	0.2753	0.0431	0.6472	0.0715	0.6477
N	2340	2340	2340	2340	2340	2340	2340	2340
Average	0.0105	0.5456	0.0105	0.5456	-0.0873	0.3165	-0.0873	0.3165

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Notes: The model is estimated as linear OLS model including time and industry FE and using two-ways clustered SE. The explanatory variables have been selected in a step-wise selection procedure searching for the highest explanatory power measured by the BIC (see text for further detail). Regressions are run separately for spillovers and network-based variables computed on the basis of upstream ($d = in$) and downstream ($d = out$) links. Data: 6-digit, balanced panel.

Table A.5: Regression results explaining the evolution of industrial hierarchies (6-digit, balanced panel).

A.3. Link formation at the extensive margin

Here, I show the results on the drivers of link formation at the extensive margin for more aggregate and disaggregate data. Table A.6 and A.7 show the results for the 2- and 3-digit level and Table A.8 for the 6-digit level.

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Pr(\Delta w_{ij,t}^{\mu} \gg 0)$	$Pr(\Delta w_{ij,t}^{\mu} \ll 0)$	$Pr(\Delta w_{ij,t}^{\mu} \gg 0)$	$Pr(\Delta w_{ij,t}^{\mu} \ll 0)$	$Pr(\Delta w_{ij,t}^{\tau} \gg 0)$	$Pr(\Delta w_{ij,t}^{\tau} \ll 0)$	$Pr(\Delta w_{ij,t}^{\tau} \gg 0)$	$Pr(\Delta w_{ij,t}^{\tau} \ll 0)$
	upstream		downstream		upstream		downstream	
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	0.2435 (0.1362)	0.57* (0.2231)	0.0401 (0.1502)	0.1295 (0.1065)	-0.0211 (0.035)	0.0398 (0.0401)	0.0053 (0.0397)	0.2294* (0.1036)
$1(\Delta w_{ij,t-1}^{\mu,d} \gg 0)$	0.0222 (0.0322)	0.0597 (0.0306)	-0.0108 (0.0367)	0.0557* (0.0283)	-0.0168 (0.0147)	0.0343* (0.0133)	0.0175 (0.0146)	-6e-04 (0.0127)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	0.0179 (0.0719)	-0.011 (0.0746)	-0.0335 (0.0688)	0.0127 (0.0626)	0.1298*** (0.0335)	-0.0502 (0.0333)	0.0444 (0.029)	-0.133*** (0.0392)
$1(\Delta w_{ij,t-1}^{\tau,d} \gg 0)$	0.0028 (0.0444)	-0.0237 (0.0376)	-0.0218 (0.044)	0.005 (0.038)	0.0895** (0.0324)	0.1067*** (0.0273)	0.0501 (0.0306)	-0.0176 (0.0254)
$\sigma_{ij,t-1}^{\mu,d}$	0.1073 (0.0617)	0.0601 (0.0675)	0.0649 (0.0511)	0.0757 (0.0428)	0.0292 (0.0264)	4e-04 (0.0311)	0.0368* (0.0165)	-0.0248 (0.0157)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	0.0052 (0.0395)	0.0882* (0.0383)	0.0017 (0.0312)	0.0379 (0.0272)	0.0132 (0.0136)	0.0271 (0.0183)	0.021* (0.01)	0.0025 (0.0096)
$\sigma_{ij,t-1}^{\tau,d}$	0.0268 (0.0294)	-0.0287 (0.0341)	-0.0425 (0.0218)	0.0277 (0.0193)	-0.0069 (0.011)	-0.0196 (0.0118)	0.0071 (0.0092)	-0.0023 (0.0055)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	0.013 (0.0185)	-0.0069 (0.0216)	-0.0188 (0.0197)	0.0312* (0.0155)	-0.0021 (0.0063)	-0.0078 (0.0067)	0.0076 (0.0077)	0.004 (0.0057)
$A_{i,t-1}^{\mu}$	0.075 (0.0516)	0.2534*** (0.0534)	0.0892 (0.0575)	0.1589*** (0.0459)	-0.0195 (0.0174)	-0.0179 (0.0215)	0.0154 (0.0153)	-0.0146 (0.0161)
$A_{j,t-1}^{\mu}$	-0.0121 (0.0617)	-0.0941* (0.0473)	0.1205** (0.04)	-0.0829 (0.0508)	-0.0287 (0.0216)	0.0354 (0.0226)	0.0057 (0.0182)	0.0392 (0.0205)
$A_{i,t-1}^{\tau}$	-0.0582 (0.1688)	0.0559 (0.1564)	0.1393 (0.1812)	0.125 (0.1353)	-0.0072 (0.1338)	0.1258 (0.1291)	0.1188 (0.1407)	0.2162 (0.1408)
$A_{j,t-1}^{\tau}$	0.0915 (0.1724)	0.1091 (0.1252)	0.0972 (0.1466)	0.1297 (0.1524)	0.0883 (0.0707)	-0.0749 (0.059)	-0.0443 (0.0987)	-0.0595 (0.0659)
$D_{i,t-1}^{\mu,d}$	-0.028 (0.0244)	0.0149 (0.0267)	0.0823 (0.0446)	-0.0073 (0.0358)	0.0097 (0.0089)	-0.0146 (0.0089)	0.0144 (0.0131)	-0.0144 (0.0153)
$D_{j,t-1}^{\mu,d}$	0.0222 (0.0261)	-0.0549* (0.0229)	0.0135 (0.0557)	0.0048 (0.0379)	-0.004 (0.009)	-0.0021 (0.0126)	-0.0225 (0.0172)	-0.0259 (0.0166)
$D_{i,t-1}^{\tau,d}$	-0.0338 (0.0534)	0.0284 (0.0568)	-0.1044* (0.0529)	-0.0897 (0.0507)	-0.0156 (0.0126)	-0.0097 (0.0127)	-0.0333 (0.0202)	-0.0202 (0.0211)
$D_{j,t-1}^{\tau,d}$	0.0085 (0.0615)	0.0977 (0.0588)	0.0635 (0.0524)	0.05 (0.0505)	-0.0178 (0.0212)	0.0236 (0.0261)	-0.0013 (0.0235)	0.0496* (0.0222)
$Spill(A)_{i,t-1}^{\mu,d}$	0.0477 (0.0775)	-0.1248 (0.0819)	-0.0372 (0.0507)	0.0106 (0.0394)	0.0501 (0.0308)	-0.0258 (0.0393)	-0.0322** (0.0119)	-7e-04 (0.0121)
$Spill(A)_{j,t-1}^{\mu,d}$	0.0316 (0.0854)	-0.0833 (0.0748)	-0.0755 (0.0412)	0.0035 (0.0376)	0.0081 (0.027)	-0.0329 (0.0337)	-0.0586*** (0.0158)	-0.0231 (0.0146)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.0454 (0.0277)	0.0395 (0.0281)	0.1623* (0.0699)	-0.0216 (0.0595)	-0.0024 (0.0078)	0.0033 (0.0094)	-0.0259 (0.0345)	0.0117 (0.033)
$Spill(A)_{j,t-1}^{\tau,d}$	0.0325 (0.0271)	0.0349 (0.0298)	0.0593 (0.073)	-0.093 (0.0707)	-0.0239* (0.0117)	-0.0085 (0.0112)	0.061 (0.0347)	-0.0487 (0.0337)
R^2	0.0351	0.0767	0.0553	0.0695	0.0468	0.0698	0.0391	0.0477
N	1020	1084	1027	1077	1340	1392	1323	1413
Sample avg.	0.4917	0.506	0.4885	0.5091	0.3452	0.3266	0.3333	0.3369
Subset avg.	0.1369	0.1504	0.1381	0.1456	0.0972	0.0845	0.0929	0.0845

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). Data: 2-digit balanced panel.

Table A.6: Regression of link formation & breaking (2-digit).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Pr(\Delta w_{ij,t}^{\mu} \geq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \leq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \geq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$
	upstream		downstream		upstream		downstream	
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	-0.0104 (0.0208)	-0.0525** (0.0188)	0.017 (0.0204)	-6e-04 (0.0158)	-0.0113 (0.0115)	-0.0031 (0.013)	0.0233 (0.0165)	-0.0155 (0.0127)
$1(\Delta w_{ij,t-1}^{\mu,d} \geq 0)$	0.0587*** (0.0101)	0.0831*** (0.0085)	0.0423*** (0.0099)	0.049*** (0.0078)	0.0038 (0.0038)	-0.0013 (0.0032)	0.0021 (0.0036)	-0.0026 (0.0034)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	0.0279 (0.0199)	-0.0028 (0.0183)	0.0573** (0.0187)	0.0099 (0.0165)	0.087*** (0.0113)	-0.1512*** (0.0117)	0.039*** (0.0107)	-0.1585*** (0.0124)
$1(\Delta w_{ij,t-1}^{\tau,d} \geq 0)$	-0.0019 (0.0134)	-5e-04 (0.0126)	0.0132 (0.0139)	0.0154 (0.0119)	0.1171*** (0.0124)	0.1416*** (0.0103)	0.0876*** (0.0127)	0.1085*** (0.0097)
$\sigma_{ij,t-1}^{\mu,d}$	0.0217 (0.0195)	0.1047*** (0.0167)	-0.0034 (0.0148)	0.0809*** (0.0123)	0.0025 (0.0066)	0.0018 (0.0064)	-0.0089 (0.0056)	-0.0077 (0.0047)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	0.0143 (0.011)	0.0527*** (0.0094)	-0.0119 (0.0081)	0.039*** (0.0065)	-5e-04 (0.0039)	-0.0082* (0.0037)	-1e-04 (0.0029)	-5e-04 (0.0026)
$\sigma_{ij,t-1}^{\tau,d}$	0 (0.0097)	-0.0169 (0.0087)	-0.0052 (0.0082)	0.0173* (0.0071)	0.0081* (0.0038)	0.0092** (0.0033)	0.007* (0.0033)	0.0012 (0.0028)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	0.0094 (0.0061)	-0.0099 (0.0053)	0.002 (0.0057)	0.0066 (0.0051)	0.0052* (0.0021)	0.0036* (0.0017)	0.0034 (0.0019)	0.0011 (0.0016)
$A_{i,t-1}^{\mu}$	0.1202*** (0.0116)	0.1221*** (0.0115)	0.0609*** (0.0129)	0.1969*** (0.0096)	-0.0063 (0.0037)	-0.0021 (0.0033)	-0.0055 (0.0041)	-0.0181*** (0.004)
$A_{j,t-1}^{\mu}$	-0.0082 (0.0136)	-0.0255* (0.011)	0.0194 (0.011)	-0.0041 (0.0104)	0.0013 (0.0047)	-0.0031 (0.0046)	-0.0061 (0.0046)	-0.0027 (0.0042)
$A_{i,t-1}^{\tau}$	-0.1171** (0.0361)	-0.0658 (0.0345)	-0.0277 (0.0374)	0.015 (0.0274)	0.0557 (0.029)	0.124*** (0.0255)	0.1216*** (0.027)	0.1122*** (0.0248)
$A_{j,t-1}^{\tau}$	-0.0284 (0.039)	-0.0299 (0.0285)	-0.0307 (0.0294)	-0.0146 (0.0266)	-0.0302 (0.0155)	-0.0258 (0.0143)	0.0032 (0.0141)	-0.0133 (0.0128)
$D_{i,t-1}^{\mu,d}$	0.0187*** (0.0053)	-0.0045 (0.0043)	0.1226*** (0.0221)	0.0039 (0.0221)	-0.0047 (0.0032)	0.0018 (0.0026)	0.0013 (0.0077)	0.0275*** (0.0075)
$D_{j,t-1}^{\mu,d}$	0.0196*** (0.0047)	0.0246*** (0.0041)	0.0568*** (0.0171)	0.0057 (0.0147)	0.005* (0.0023)	0.0102*** (0.0026)	0.0166 (0.0115)	-0.0037 (0.0099)
$D_{i,t-1}^{\tau,d}$	0.0113 (0.0137)	-0.02 (0.0117)	-0.0163 (0.0186)	-0.0328* (0.0136)	-0.0083* (0.0041)	-0.0141*** (0.0041)	-0.0394*** (0.0058)	-0.0153** (0.0052)
$D_{j,t-1}^{\tau,d}$	-0.0264 (0.017)	-0.0415** (0.0129)	-0.008 (0.016)	0.0144 (0.0153)	0.0138** (0.0054)	0.0207*** (0.0041)	0.0284*** (0.0071)	0.0511*** (0.0056)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0175 (0.0163)	-0.0067 (0.0142)	0.0119 (0.009)	0.0087 (0.0065)	0.0089 (0.0073)	0.0075 (0.007)	-0.0083* (0.0042)	0.0146*** (0.0036)
$Spill(A)_{j,t-1}^{\mu,d}$	0.0021 (0.0177)	-8e-04 (0.0139)	4e-04 (0.0094)	-0.0145* (0.0068)	-0.0108 (0.0074)	0.01 (0.0067)	0.0034 (0.0038)	-1e-04 (0.0037)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.019** (0.0069)	0.0126* (0.0058)	0.0125* (0.0064)	0.0073 (0.006)	-0.0044** (0.0014)	-0.0022 (0.0013)	-0.0029* (0.0012)	-7e-04 (0.0011)
$Spill(A)_{j,t-1}^{\tau,d}$	0.0044 (0.0071)	0.0107 (0.0064)	0.005 (0.0053)	-0.0034 (0.0056)	0.0034 (0.0027)	0.0076*** (0.0022)	0.0036 (0.002)	6e-04 (0.0016)
R^2	0.0403	0.0514	0.0199	0.0917	0.0271	0.1422	0.0252	0.1199
N	10249	12142	10462	11929	16483	16495	16225	16752
Sample avg.	0.441	0.5184	0.4477	0.5117	0.2281	0.2053	0.2176	0.2158
Subset avg.	0.1017	0.1135	0.1024	0.1092	0.0743	0.0673	0.0699	0.0642

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). The sample average is the probability of link formation or breaking in the full sample, while the subset average is the frequency of significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$.
Data: 3-digit balanced panel.

Table A.7: Regression of link formation & breaking (3-digit).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Pr(\Delta w_{ij,t}^{\mu} \geq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$
	upstream		downstream		upstream		downstream	
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	-0.0409*** (0.0021)	-0.0723*** (0.0018)	-0.0418*** (0.002)	-0.0563*** (0.0018)	0.0033*** (9e-04)	-0.0034*** (9e-04)	0.0014 (0.001)	0.0013 (9e-04)
$1(\Delta w_{ij,t-1}^{\mu,d} \geq 0)$	0.0956*** (0.0023)	0.1217*** (0.0017)	0.1156*** (0.002)	0.0666*** (0.0017)	-4e-04 (5e-04)	3e-04 (5e-04)	-8e-04 (5e-04)	0.0018*** (5e-04)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	0.0127*** (0.0024)	-0.0089*** (0.0022)	0.0102*** (0.0022)	0.0148*** (0.0021)	0.0555*** (0.002)	-0.1503*** (0.0022)	0.0656*** (0.0021)	-0.182*** (0.0025)
$1(\Delta w_{ij,t-1}^{\tau,d} \geq 0)$	0.01*** (0.0018)	-0.0018 (0.0015)	-0.0042** (0.0016)	0.0035* (0.0016)	0.1279*** (0.0028)	0.1064*** (0.0021)	0.1251*** (0.0027)	0.1319*** (0.0022)
$\sigma_{ij,t-1}^{\mu,d}$	-0.0158*** (0.0021)	0.0492*** (0.0018)	-0.0193*** (0.0025)	0.0554*** (0.0023)	-0.0012* (6e-04)	-5e-04 (7e-04)	-9e-04 (7e-04)	-0.0022** (8e-04)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	0.0024* (0.001)	0.0251*** (9e-04)	0.001 (0.001)	0.0338*** (9e-04)	-0.0014*** (3e-04)	-0.0024*** (3e-04)	-8e-04* (3e-04)	-3e-04 (3e-04)
$\sigma_{ij,t-1}^{\tau,d}$	-0.0134*** (0.001)	-0.0078*** (8e-04)	0.0188*** (0.0012)	0.0122*** (0.0011)	0.006*** (5e-04)	0.0077*** (6e-04)	0.0049*** (5e-04)	0.0083*** (5e-04)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	-0.0024*** (5e-04)	-0.0019*** (5e-04)	0.0054*** (6e-04)	0.0069*** (6e-04)	0.0014*** (3e-04)	0.0019*** (3e-04)	-3e-04 (3e-04)	0.0028*** (3e-04)
$A_{i,t-1}^{\mu}$	0.0541*** (0.001)	0.0549*** (0.001)	0.0311*** (0.0012)	0.0812*** (0.0011)	-0.002*** (6e-04)	9e-04 (5e-04)	-5e-04 (5e-04)	0.0012** (4e-04)
$A_{j,t-1}^{\mu}$	0.0087*** (0.0012)	0.0143*** (0.001)	-0.0018 (0.001)	-0.0083*** (9e-04)	-0.001* (4e-04)	1e-04 (4e-04)	-0.0021*** (5e-04)	-5e-04 (5e-04)
$A_{i,t-1}^{\tau}$	-0.0313*** (0.0028)	-0.0301*** (0.0025)	-0.0087** (0.0029)	0.0427*** (0.0026)	0.0246*** (0.0024)	0.0474*** (0.0022)	0.0262*** (0.0021)	0.0145*** (0.0017)
$A_{j,t-1}^{\tau}$	-0.0179*** (0.0028)	0.0049* (0.0023)	-0.0097*** (0.0024)	0.0108*** (0.0023)	-0.0015 (0.0011)	-0.0096*** (0.0012)	-0.0074*** (0.0015)	-0.0151*** (0.0015)
$D_{i,t-1}^{\mu,d}$	-0.007*** (8e-04)	0.0021** (8e-04)	0.0746*** (0.003)	0.2038*** (0.0028)	0.0018*** (4e-04)	4e-04 (4e-04)	-0.0153*** (0.0011)	-0.0172*** (0.001)
$D_{j,t-1}^{\mu,d}$	0.0011 (8e-04)	0.0023*** (7e-04)	0.0612*** (0.0028)	0.09*** (0.0026)	4e-04 (5e-04)	-0.0012** (4e-04)	-0.0025* (0.0011)	0.0205*** (0.001)
$D_{i,t-1}^{\tau,d}$	-0.01*** (0.0013)	-0.0047*** (0.0012)	0.009*** (0.001)	-0.0017 (0.001)	-1e-04 (4e-04)	4e-04 (4e-04)	-0.0065*** (4e-04)	0.0105*** (5e-04)
$D_{j,t-1}^{\tau,d}$	0.0089*** (0.0014)	8e-04 (0.0013)	0.0045*** (0.0011)	0.001 (0.001)	0.0057*** (6e-04)	0.0163*** (5e-04)	0.0065*** (5e-04)	0.0218*** (5e-04)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0239*** (0.0011)	-0.0188*** (0.001)	0.005*** (0.0011)	-0.001 (0.001)	0.0067*** (5e-04)	0.0024*** (5e-04)	0.0039*** (6e-04)	0.0044*** (5e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	0.0034** (0.0011)	-0.0172*** (0.001)	0.0212*** (0.0011)	-0.0101*** (0.001)	-2e-04 (6e-04)	-8e-04 (5e-04)	0.0012* (6e-04)	-0.0053*** (6e-04)
$Spill(A)_{i,t-1}^{\tau,d}$	0 (0.001)	0 (0.001)	0.0037*** (8e-04)	0.0047*** (7e-04)	-0.0047*** (3e-04)	-0.0044*** (3e-04)	-0.0012*** (2e-04)	0 (2e-04)
$Spill(A)_{j,t-1}^{\tau,d}$	0.003** (0.0011)	0.0031** (0.0011)	0.0012 (7e-04)	-0.001 (6e-04)	0.0011* (4e-04)	0.0046*** (4e-04)	0.002*** (3e-04)	0.0022*** (2e-04)
R^2	0.03	0.0718	0.0424	0.1118	0.0234	0.1128	0.0218	0.156
N	531934	527054	534610	524376	653733	658546	652979	659335
Sample avg.	0.2937	0.2993	0.2882	0.3048	0.1226	0.1277	0.1224	0.1279
Subset avg.	0.0678	0.0699	0.0622	0.0688	0.0395	0.0406	0.0352	0.0351

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). Data: 6-digit balanced panel.

Table A.8: Regression of link formation & breaking (6-digit).

A.4. Link formation at the intensive margin

Here, I present additional results on the determinants of link formation at the intensive margin for more aggregate data. Table A.9 and A.10 show the results for the 2- and 3-digit level and Table A.11 for the 6-digit level.

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	0.083 (0.096)	-0.4125*** (0.0971)	-0.0787 (0.0519)	-0.1634* (0.0826)	0.0146 (0.0604)	-0.0586 (0.0997)	-0.0266 (0.1121)	-0.0865 (0.0742)
$\Delta w_{ij,t-1}^{\tau,d}$	-0.0392 (0.0313)	-0.0264 (0.0135)	-0.023 (0.0233)	0.0174 (0.0115)	-0.0684 (0.0449)	-0.1245** (0.042)	0.0888 (0.0751)	-0.1107 (0.0694)
$\sigma_{ij,t-1}^{\mu,d}$	0.0171 (0.0093)	-0.0066 (0.006)	0.0112* (0.0047)	-0.0149* (0.0059)	0.0072 (0.0065)	0.0022 (0.0036)	0.0131 (0.0079)	-0.0114 (0.0064)
$\Delta \sigma_{ij,t-1}^{\mu,d}$	-0.002 (0.004)	-0.0027 (0.0033)	0.0036 (0.0026)	-0.0093* (0.0036)	0.0023 (0.0033)	-0.0014 (0.0028)	-0.0044 (0.005)	-0.0024 (0.0036)
$\sigma_{ij,t-1}^{\tau,d}$	6e-04 (0.0022)	0.0019 (0.0015)	0.0029 (0.0025)	4e-04 (7e-04)	-9e-04 (0.0037)	0.0023 (0.0029)	-0.0034 (0.0061)	0.0037 (0.0024)
$\Delta \sigma_{ij,t-1}^{\tau,d}$	-0.001 (0.0018)	5e-04 (0.0012)	0.0027 (0.0017)	7e-04 (0.0011)	-4e-04 (0.0018)	0.0034 (0.0025)	-0.0019 (0.0031)	0.0026 (0.0025)
$A_{i,t-1}^{\mu}$	0.0077* (0.0033)	-0.0056 (0.0033)	0.008 (0.0047)	-0.0116** (0.0036)	-0.0024 (0.003)	0.0054 (0.0028)	0.0059 (0.0047)	0.0082* (0.0034)
$A_{j,t-1}^{\mu}$	-0.0115** (0.0043)	0.0084** (0.003)	3e-04 (0.0044)	0.0093 (0.0053)	0.0029 (0.0055)	0.0026 (0.0034)	-0.0195*** (0.0054)	0.0137* (0.0067)
$A_{i,t-1}^{\tau}$	-0.0109 (0.0144)	0.0155 (0.0231)	-0.008 (0.0158)	-0.0147 (0.0107)	0.0292 (0.0156)	-0.0046 (0.0126)	0.0316 (0.0188)	-0.0214 (0.017)
$A_{j,t-1}^{\tau}$	-0.001 (0.0175)	0.0071 (0.0096)	0.0017 (0.0085)	-0.0054 (0.0087)	0.006 (0.0059)	0.002 (0.0046)	0.0285* (0.0111)	-7e-04 (0.0091)
$D_{i,t-1}^{\mu,d}$	0.0032 (0.003)	0.0025* (0.0012)	0.0059 (0.0032)	-8e-04 (0.0027)	0.0023 (0.0021)	0.0011 (7e-04)	0.0037 (0.003)	-0.0011 (0.002)
$D_{j,t-1}^{\mu,d}$	-0.0037 (0.0032)	0.0045 (0.0024)	3e-04 (0.0053)	-0.0038 (0.0047)	-0.0022 (0.0025)	-6e-04 (0.0016)	0.0012 (0.0065)	4e-04 (0.0061)
$D_{i,t-1}^{\tau,d}$	0.0014 (0.0037)	-0.0049 (0.003)	-0.0144*** (0.0041)	0.0016 (0.0026)	0.0017 (0.003)	0.0023 (0.0027)	0.0097* (0.0049)	0.0026 (0.0043)
$D_{j,t-1}^{\tau,d}$	0.002 (0.0046)	-0.0017 (0.0027)	0.0021 (0.0041)	-0.0016 (0.0032)	-0.0044 (0.0059)	-0.01 (0.0072)	-0.0439* (0.019)	0.039 (0.0202)
$Spill(A)_{i,t-1}^{\mu,d}$	0.0025 (0.0066)	0.0199*** (0.0058)	-0.0034 (0.0034)	0.0086* (0.004)	0.0055 (0.0066)	-0.0041 (0.0047)	-0.0022 (0.0045)	0.0022 (0.0046)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.0051 (0.0075)	0.0025 (0.0058)	-0.0081 (0.0042)	0.0064 (0.0052)	-8e-04 (0.006)	0.0026 (0.0044)	0.0139 (0.0088)	-0.0094 (0.0059)
$Spill(A)_{i,t-1}^{\tau,d}$	9e-04 (0.0026)	0.0024 (0.0014)	0.0027 (0.0059)	-5e-04 (0.0028)	3e-04 (0.0016)	-9e-04 (0.0022)	-0.0082 (0.0102)	-0.0024 (0.0072)
$Spill(A)_{j,t-1}^{\tau,d}$	0.0055* (0.0026)	-0.0024 (0.0014)	0.0034 (0.0054)	-0.0092 (0.0052)	0.0027 (0.0032)	-0.0031 (0.0033)	0.0067 (0.0087)	0.0121 (0.0072)
R^2	0.0904	0.2536	0.0723	0.1258	0.079	0.1339	0.1935	0.199
N	1016	1080	1023	1073	708	760	687	777
Average*	0.0192	-0.0181	0.0168	-0.0175	0.0166	-0.0153	0.0228	-0.02
Subset avg.**	0.0506	-0.0444	0.045	-0.045	0.0668	-0.0745	0.0892	-0.0856

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 '.' 1

The models are OLS regressions including industry-pair and time FE and using two-ways clustered SE estimating the magnitude of change conditional on $\Delta w_{ij,t}^{\alpha,d} \neq 0$. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta w_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.

Data: 2-digit balanced panel.

Table A.9: Regression of link formation & breaking at the intensive margin (2-digit).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	0.1373 (0.089)	-0.1843* (0.0734)	-0.0636 (0.0965)	-0.2596*** (0.0348)	-0.0015 (0.0295)	0.0297 (0.0259)	0.0746 (0.0638)	-0.0402 (0.0595)
$\Delta w_{ij,t-1}^{\tau,d}$	0.0064 (0.0135)	0.0106 (0.0098)	0.0013 (0.0052)	-0.007 (0.0043)	0.002 (0.0419)	-0.1574*** (0.0404)	0.0388 (0.0348)	-0.2143*** (0.0251)
$\sigma_{ij,t-1}^{\mu,d}$	0.0015 (0.0023)	0.0013 (0.0025)	0.0014 (0.002)	-0.0049*** (8e-04)	-0.0023 (0.0024)	-5e-04 (0.0017)	-1e-04 (0.0023)	-0.0023 (0.002)
$\Delta \sigma_{ij,t-1}^{\mu,d}$	6e-04 (0.0014)	2e-04 (0.001)	-3e-04 (8e-04)	-0.0011** (4e-04)	-0.0013 (0.0014)	-0.0011 (0.001)	-9e-04 (0.0012)	-4e-04 (0.001)
$\sigma_{ij,t-1}^{\tau,d}$	0.0011 (7e-04)	-2e-04 (4e-04)	-2e-04 (5e-04)	-6e-04 (4e-04)	-3e-04 (0.0018)	0.0029 (0.0018)	-0.0038 (0.0027)	0.0035 (0.002)
$\Delta \sigma_{ij,t-1}^{\tau,d}$	0.0012** (4e-04)	-2e-04 (2e-04)	2e-04 (3e-04)	-4e-04 (3e-04)	-0.001 (0.0013)	7e-04 (0.0011)	-0.0018 (0.0016)	0.0012 (0.0014)
$A_{i,t-1}^{\mu}$	0.0084*** (0.0013)	-0.0043*** (9e-04)	0.005*** (0.0012)	-0.0089*** (8e-04)	0.0017 (0.0016)	0.0012 (0.0018)	0.0016 (0.0018)	0.0017 (0.0012)
$A_{j,t-1}^{\mu}$	-0.0021* (0.0011)	0.0016** (6e-04)	-8e-04 (9e-04)	0.0015* (7e-04)	0.0022 (0.002)	-7e-04 (0.0016)	-0.005** (0.0019)	0.0027 (0.002)
$A_{i,t-1}^{\tau}$	-0.0106** (0.0036)	9e-04 (0.003)	-0.0071* (0.0033)	-0.0016 (0.002)	-0.0037 (0.0061)	-0.003 (0.004)	0.0151** (0.005)	-0.0104* (0.0043)
$A_{j,t-1}^{\tau}$	-0.0027 (0.0052)	0.0017 (0.0012)	0.0032 (0.0017)	0.0032 (0.0037)	5e-04 (0.0029)	-0.0034 (0.0022)	-3e-04 (0.0027)	-0.0016 (0.0018)
$D_{i,t-1}^{\mu,d}$	8e-04 (5e-04)	-3e-04 (2e-04)	-2e-04 (0.0013)	-0.0064*** (0.0013)	-9e-04 (6e-04)	-2e-04 (4e-04)	-0.0019 (0.0017)	-0.0049 (0.0026)
$D_{j,t-1}^{\mu,d}$	8e-04 (8e-04)	0.0038* (0.0018)	-0.0043 (0.0032)	0.0044 (0.0029)	3e-04 (7e-04)	-0.0013 (0.0012)	-0.0063* (0.0027)	0.0019 (0.0027)
$D_{i,t-1}^{\tau,d}$	-0.0031*** (9e-04)	6e-04 (7e-04)	-6e-04 (0.001)	0.0024** (8e-04)	6e-04 (0.0017)	9e-04 (0.0019)	-6e-04 (0.0024)	7e-04 (0.0019)
$D_{j,t-1}^{\tau,d}$	0.0017 (0.0017)	-4e-04 (8e-04)	8e-04 (0.0011)	-9e-04 (9e-04)	-0.0218*** (0.0054)	0.0312*** (0.0083)	-0.0413*** (0.0069)	0.0637*** (0.01)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0011 (0.0018)	-0.001 (0.0013)	0.0024** (8e-04)	-2e-04 (6e-04)	0.0033 (0.0017)	-0.0013 (0.0016)	-0.0019 (0.001)	-5e-04 (9e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.003 (0.002)	-0.0016 (0.0012)	-0.002* (8e-04)	8e-04 (5e-04)	1e-04 (0.002)	0.0022 (0.0014)	0.0025 (0.0013)	-0.0018 (0.0013)
$Spill(A)_{i,t-1}^{\tau,d}$	-7e-04 (4e-04)	9e-04** (3e-04)	7e-04** (2e-04)	3e-04 (2e-04)	9e-04 (0.001)	-0.0011 (0.0017)	0.0011 (0.0014)	-9e-04 (0.0013)
$Spill(A)_{j,t-1}^{\tau,d}$	0.0011* (5e-04)	1e-04 (3e-04)	-3e-04 (4e-04)	-1e-04 (4e-04)	-0.0048 (0.004)	-8e-04 (0.0018)	0.0032 (0.0025)	-0.0016 (0.002)
R^2	0.038	0.0751	0.0166	0.1587	0.0627	0.2168	0.0982	0.4008
N	9308	11201	9521	10988	4955	4967	4698	5225
Average*	0.0095	-0.008	0.0082	-0.0077	0.0149	-0.0147	0.0189	-0.0162
Subset avg.**	0.0322	-0.0277	0.0286	-0.0281	0.0542	-0.0525	0.0658	-0.0621

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are OLS regressions including industry-pair and time FE and using two-ways clustered SE estimating the magnitude of change conditional on $\Delta w_{ij,t}^{\alpha,d} \neq 0$. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta w_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.

Data: 3-digit balanced panel.

Table A.10: Regression of link formation & breaking at the intensive margin (3-digit).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	-0.0888* (0.0374)	-0.279*** (0.0224)	-0.052* (0.0244)	-0.2768*** (0.0131)	-0.0033 (0.0102)	0.0071 (0.0079)	-0.0189 (0.0134)	0.0251** (0.0094)
$\Delta w_{ij,t-1}^{\tau,d}$	-0.0018 (0.0028)	0.003 (0.0031)	-0.0046 (0.0043)	-9e-04 (0.0021)	0.1336*** (0.0219)	-0.191*** (0.0138)	0.1873*** (0.0317)	-0.2811*** (0.0139)
$\sigma_{ij,t-1}^{\mu,d}$	-1e-04 (2e-04)	-0.0014*** (2e-04)	3e-04 (3e-04)	-0.0022*** (2e-04)	9e-04** (3e-04)	6e-04*** (2e-04)	0.0017*** (4e-04)	-5e-04 (3e-04)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	-1e-04 (1e-04)	-4e-04*** (1e-04)	1e-04 (1e-04)	-9e-04*** (1e-04)	6e-04*** (1e-04)	-1e-04 (1e-04)	5e-04** (2e-04)	-2e-04 (1e-04)
$\sigma_{ij,t-1}^{\tau,d}$	7e-04*** (1e-04)	-6e-04*** (1e-04)	0 (1e-04)	-1e-04 (1e-04)	-1e-04 (2e-04)	1e-04 (2e-04)	5e-04 (4e-04)	-4e-04 (3e-04)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	3e-04*** (1e-04)	-1e-04 (1e-04)	-1e-04 (0)	0 (0)	1e-04 (1e-04)	0 (1e-04)	9e-04** (3e-04)	-4e-04* (2e-04)
$A_{i,t-1}^{\mu}$	0.002*** (1e-04)	-0.0015*** (1e-04)	0.0011*** (1e-04)	-0.0032*** (1e-04)	-7e-04*** (2e-04)	4e-04* (2e-04)	-0.0013*** (4e-04)	-1e-04 (2e-04)
$A_{j,t-1}^{\mu}$	-3e-04*** (1e-04)	5e-04*** (1e-04)	-1e-04 (1e-04)	4e-04*** (1e-04)	-1e-04 (2e-04)	2e-04 (2e-04)	-3e-04* (2e-04)	0 (1e-04)
$A_{i,t-1}^{\tau}$	-0.0026*** (4e-04)	9e-04*** (2e-04)	-0.0013*** (3e-04)	-0.0019*** (3e-04)	0.0024*** (5e-04)	-0.0011** (4e-04)	0.0089*** (8e-04)	-0.0025*** (5e-04)
$A_{j,t-1}^{\tau}$	0 (3e-04)	-4e-04** (1e-04)	7e-04*** (2e-04)	2e-04 (2e-04)	-6e-04** (2e-04)	5e-04* (2e-04)	-6e-04** (2e-04)	3e-04 (1e-04)
$D_{i,t-1}^{\mu,d}$	1e-04 (1e-04)	1e-04 (1e-04)	-0.0028*** (2e-04)	0.0028*** (2e-04)	-1e-04 (2e-04)	-3e-04 (2e-04)	0.0025*** (6e-04)	-2e-04 (4e-04)
$D_{j,t-1}^{\mu,d}$	7e-04*** (1e-04)	6e-04 (5e-04)	-0.0023*** (2e-04)	0.0014*** (2e-04)	4e-04* (2e-04)	0 (2e-04)	6e-04 (7e-04)	0.0017** (5e-04)
$D_{i,t-1}^{\tau,d}$	5e-04*** (1e-04)	-3e-04** (1e-04)	6e-04*** (1e-04)	-4e-04*** (1e-04)	4e-04 (3e-04)	9e-04** (3e-04)	0 (2e-04)	-5e-04*** (1e-04)
$D_{j,t-1}^{\tau,d}$	-5e-04*** (1e-04)	2e-04* (1e-04)	-2e-04* (1e-04)	0 (1e-04)	-0.0051*** (8e-04)	0.0121*** (0.0016)	-0.0079*** (9e-04)	0.022*** (0.0021)
$Spill(A)_{i,t-1}^{\mu,d}$	-1e-04 (2e-04)	0.001*** (2e-04)	3e-04 (2e-04)	5e-04*** (1e-04)	-4e-04 (3e-04)	0.0015*** (2e-04)	-7e-04* (3e-04)	0.0012*** (2e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	-4e-04* (2e-04)	7e-04*** (2e-04)	-1e-04 (2e-04)	4e-04*** (1e-04)	-4e-04* (2e-04)	-3e-04* (2e-04)	2e-04 (3e-04)	-3e-04* (2e-04)
$Spill(A)_{i,t-1}^{\tau,d}$	-5e-04*** (1e-04)	-1e-04 (1e-04)	1e-04* (1e-04)	1e-04** (0)	3e-04 (2e-04)	6e-04 (4e-04)	-3e-04 (5e-04)	-0.0012*** (4e-04)
$Spill(A)_{j,t-1}^{\tau,d}$	1e-04 (1e-04)	1e-04 (1e-04)	-1e-04 (1e-04)	-1e-04 (1e-04)	-2e-04 (3e-04)	0.0013*** (4e-04)	-3e-04 (9e-04)	6e-04 (7e-04)
R^2	0.0156	0.1479	0.0091	0.2004	0.0385	0.1979	0.0834	0.3708
N	231496	226616	234174	223940	100004	104817	99215	105571
Average*	0.0031	-0.0031	0.0029	-0.0032	0.0061	-0.0057	0.0068	-0.0062
Subset avg.**	-0.0111	0.0112	-0.0123	0.0109	-0.0219	0.0229	-0.0257	0.0284

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta w_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.
Data: 6-digit balanced panel.

Table A.11: Regression of link formation & breaking (6-digit).

Part II.

SUPPLEMENTARY MATERIAL

SI.1. Detailed information on data processing

This analysis builds on two distinct sources of data brought into a consistent form that enables the statistical analyses; this is (1) a series of time snapshots of the network of cross-industrial IO flows and patent citations and (2) a panel data set of aggregate statistical indicators at the industry level. Industries are identified by 6-digit NAICS codes (and later aggregated to the 4-digit level). The time snapshots cover 5-year intervals from 1976 to 2006 (1977 to 2007) for the patent (IO) data. Obtaining these data involved a series of steps of re-formatting and harmonization which are explained in detail below.

SI.1.1. Patent data

The patent data is downloaded from the [NBER](#) patent data project.²⁷ It covers all US utility patents from 1976 to 2006. Three files from this database are used:

1. `cite76_06.dta` which consists of links between individual patents identified by the patent number and granted between 1976 and 2006,
2. `pat76_06_assg.dta` which is a mapping from patent number to unique identifiers for the assignee of a patent. This file also contains meta-information about patents and assignees (e.g. time of application, technology classes, geo-information, etc.),
3. `dynass.dta` which is a dynamic, i.e. time dependent, mapping from assignees to so-called GVKEYs as used by Compustat as firm IDs. The assignment is dynamic because the GVKEY of a patent holder may change through mergers and acquisitions (M&A).

The citation raw data covers roughly 23.7 M citation links, 3.2 M patents and 224 K unique assignees. By far the largest share of assignees are US and non-US corporations

²⁷<https://sites.google.com/site/patentdataport/> [accessed on Jan 29, 2021]

Final data	
# citations	5,785,121
# unique patents	1,152,773
# different NAICS industries	526
Statistics at intermediate steps	
# unique patent assignees	12,649
# unique firms (GVKEYs)	7,160

The data covers patents with an application date between 1972 and 2006. Firms in the data are publicly traded US firms.

Table SI.1: Overview of patent data and processing

(next to individuals and research institutions holding a patent). For some of the patents, the assignee is not known which is due to the matching process. [Hall et al. \[2001\]](#) describe these data in detail. In addition to these data, I use a mapping from GVKEYs to industrial codes (SIC, NAICS) obtained from [CapitalIQ](#).²⁸ This enables to map patents to NAICS industry codes via the firm ID of the owner of the patent. I use concordance tables provided by the [US Census Bureau](#) to transform SIC and NAICS 2007 codes to NAICS 2002.²⁹ During the data processing, unique patents and citations between patents are first assigned to firms and then to industries. In the data, roughly 7 K firms hold more than 1.25 M patents and 6.77 M citations between patents. Incomplete coverage at some of the processing steps reduces that data which finally cover 5.79 M citation links between 1.15 M unique patents assigned to 526 different 6-digit NAICS sector (see [Table SI.1](#)).

SI.1.1.1. Overview

To synchronize the patent data with the IO data, I created five-year snapshots of the patent citation network where citations between individual patents are aggregated to citations between industries. The networks $W^{\tau,d}$ have the same structure as the IO matrices which show flows of goods between economic sectors. A positive entry $w_{ij,t}^{\tau,in}$ indicates that industry i filed patents in time t that make citations to patents owned by industry j . Time t is a time interval corresponding to the periods 1972-1976, 1977-1981, etc. The entries in the flow matrix are aggregates of all citations from i to j in t .

The time stamp of a patent is the application year which is used (instead of the grant year) because the application date is closer to the time when the cited technological

²⁸The compilation of the patent data was performed while I was PhD student at the University Paris 1 in 2018 through which I had access to the data.

²⁹<https://www.census.gov/eos/www/naics/concordances/concordances.html> [accessed on Jan 19, 2021]

knowledge was used [see [Hall et al., 2001](#)]. The application date is typically much closer to the date of invention and the time when the knowledge encoded in the citation was used. The time lag between application and grant year can be large. On average, it accounts for two years.

Some structural and conceptual concerns about patent data, its interpretation and time consistency need to be mentioned. Truncation by the end of horizon is an issue: Patents that were applied for, but not yet granted are not documented. The truncation also affects the number of citations because younger patents have a shorter lifespan in which they can be cited.

Moreover, citation practices have changed over time: The number of citations per patent is increasing over time: More recent patents cite more other patents than older ones which is partly due to the computerization of the patent file at USPTO in more recent years. Further, a legal patent reform in 1980 induced a structural break in the number of patents, the number of citations and reduced the time lag between application and grant year [[Gallini, 2002](#)].

To cope with these consistency problems, citation flows are normalized to shares and patent stocks are normalized to represent industrial hierarchies rather than absolute numbers. Moreover, the regressions include time and industry FE.

Conceptually, it should be noted that patent citations are often not made by the inventor himself but are added by patent examiners. This undermines the interpretation of a citation as explicit flow of knowledge. Here, this is of minor relevance the analyses describe aggregate patterns of innovation and technological relatedness but do not aim to identify individual predecessors of inventions.

SI.1.1.2. Processing steps in detail

Here, I explain the single processing steps. The steps are consistent with the R-code used for the data compilation and provided in [Hötte \[2021a\]](#).

Step 1: Starting point is the `cite76.06.dta` which is sequentially expanded. First, I added columns with the application year for the citing and cited patents taken from `pat76_06_ass.dta` to the columns of patent numbers of the citing and cited patent. The year of the citing patent indicates when the citation was made. The difference between year of the citing and the year of the cited patent is the citation lag. This was also used for plausibility checks.

Step 2: Next, the citation data is expanded by two columns with the IDs of the patent assignees (`pdpass`) for citing and cited patents. Some patents have more than

one assignee. In this case, the data entry is copied. This occurs for example if a patent is the result of collaboration of multiple firms.³⁰

- Step 3:** Columns with a mapping from the patent number to the GVKEY are added to the citation file. The result is a firm-level patent citation network. The firm-level year mapping allows to correct for M&A. `dynass.dta` provides data about the dynamic assignment of GVKEY, i.e. it indicates when a firm was merged, i.e. when its GVKEY changed. Hence, the mapping between GVKEYs and patent citations is time-dependent. Using the year of the citation (which equals the application year of the citing or cited patent), I make sure that the patent is mapped to the firm that own the patent in the year of application.
- Step 4:** GVKEYs are mapped to SIC and NAICS codes. Before this can be done, some reformatting of the GVKEY-industry mapping files is needed. I obtained two independent files from Compustat (CapitalIQ). One of the files provides a GVKEY-SIC mapping, the other a GVKEY-NAICS (2007) mapping.³¹ Both files have an incomplete coverage, i.e. not every GVKEY maps to an industry code. To increase the coverage, I merged both data sets. SIC and NAICS columns are added to the file with the citation links.
- Step 5:** A concordance table on the transformation of 1987 SIC codes to 2002 NAICS codes is used to convert all SIC codes into NAICS 2002 codes. The merge of two data sets has caused some inconsistencies across codes. In these cases, the GVKEY to SIC mapping is used as baseline because it has a larger coverage. The GVKEY-NAICS file is only used for those GVKEYs that are not yet mapped codes or when the SIC-NAICS mapping is ambiguous.
- Step 6:** The data is cleaned because only a subset of the very large supplemented citation file is needed. Only those citation links are used for which a mapping to NAICS codes is available for both the citing and cited patent. For the sake of

³⁰The procedure of accounting for multiple assignees is step-wise outlined in Step 2b in the R-files. Note that the double counting may affect the interpretation of patent counts as knowledge stock. This is especially true if multiple assignees have the same NAICS code. This effect is expected to be only of minor importance because: (1) This concerns only 2.5% of patents. (2) It could only be a concern if there are systematic cross-industry differences in the number of assignees per patent. (3) I make robustness checks in the analysis using the binary instead of weighted patent citation network.

³¹The consultant at CapitalIQ declared by mail that the NAICS codes are codes of the 2007 version but a formal documentation of the mapping and its sources is not available. The same is true for the SIC mapping.

documentation, I retrieve also an extended GVKEY-industry mapping. This can be used for consistency checks.³²

Step 7: The citations are split into time aggregates, i.e. into 5-year windows, using the application year of the citing patent as time stamp. This is the year when the knowledge encoded in a patent of j was used as input by i .³³ Aggregation means that all patents of industry i that cite patents from industry j within a given time window are summed up. The result is the flow of citations from i to j in t , i.e. $flow_{ij,t}^{\tau,in}$.

The aggregation is roughly consistent with the quinquennial IO data which is available for 2007, 2002, 1997, 1992, etc. The patent data set ends in 2006. To exploit the full data, I use the time windows 2002-2006, 1997-2001, 1992-1996, etc. The windows are proxies for the knowledge used within a five year period. The data are not discounted. Also citations prior to 1976 are available which allows to compute the aggregate of the first time window.

Moreover, in this step, final data formatting issues are done. Some NAICS codes are only available at aggregate 3-, 4-, or 5-digit level. This concerns roughly 12% of the links. These are mapped to more granulate 6-digit codes using an uniform-splitting rule, i.e. dividing the number of citations by the number of sub-sectors and allocating the citation codes uniformly.

The output of this step is a sample of 3-column lists with column (1) showing the NAICS-6-digit ID of the citing sector i , (2) the ID of the cited j , and (3) the number of citations from i to j which is equal to $flow_{ij,t}^{\tau,in}$ for each time window.

Step 8: For each t , I create NAICS \times NAICS matrices with citation counts $flow_{ij,t}^{\tau,in}$ as entries.

Step 9: In addition to the citation network, I create aggregate stocks of patents $A_{i,t}^{\tau}$ for each period. I use the citation-file that is linked to unique assignees created in step 2. The number of patents per assignee per year is counted. These are mapped to GVKEYs and NAICS. These stocks are used in the panel data analysis of industries.

³²From this data, also an IPC-NAICS mapping can be constructed which might be interesting to compare with other concordances, similarly as [Lybbert and Zolas \[2014\]](#), [Dorner and Harhoff \[2018\]](#). But this is beyond the scope of this project.

³³Alternatively, it would be also possible to use the application year of the cited patent as time stamp focusing on the time of knowledge production. An implicit assumption of the aggregation procedure by year of citing or cited patent (in contrast to a simultaneous approach) are fix sector boundaries.

SI.1.1.3. Technical and conceptual issues of the data processing

General format The data files are large. For the patent-year mapping and the patent-pdpass mapping, the working format is based on dimensions of the `cite76_06.dta` input file. The patent-pdpass mapping will change the dimensions when rows are copied to account for multiple assignees of one patent (see step 2). This format will then also be used for the pdpass-GVKEY mapping. It becomes a very large file with more than 23 M rows, each representing a citation link.

How to deal with dynass? Due to M&A, firms that have an GVKEY entry were acquired by other firms and have thereafter a new identity indicated by another GVKEY. I make corrections and use the GVKEY that was valid at the date of patent application (step 3).

How to deal with inconsistencies across the SIC and NAICS to industry mappings? I have two files that provide a GVKEY-industry mapping. The first table is a mapping from GVKEYs to SIC. The second is a mapping from GVKEYs to NAICS. The first is more comprehensive and primarily used in the analysis. I added a mapping from SIC to NAICS. The second file which is a direct mapping from GVKEYs to NAICS 2007 is only used to fill gaps when no SIC is available but NAICS is.

Further, there are inconsistencies between the SIC to NAICS conversion and the direct NAICS codes. Based on a random sampling manual checking of the data, I concluded that these inconsistencies are negligible. The inspected inconsistencies occurred at a very high-resolution level (5-digit or more). The inconsistencies may arise from changes over time in the assignment of GVKEYs to industries that is done by CapitalIQ, or from updates in the classification schemes. I checked some of these updates manually, but could not confirm that the latter is the reason. Hence, most likely that has something to do with the assignment of GVKEYs (securities) to industries.

SI.1.2. Input-output data

The IO data is constructed by the composition and harmonization of the historical benchmark tables provided by [Bureau of Economic Analysis \(BEA\)](#).³⁴ Since 1947, BEA publishes IO tables at the detailed industry level every 5 years. The data is collected

³⁴<https://www.bea.gov/industry/historical-benchmark-input-output-tables> [accessed on Dec 21, 2020]

in BEA’s quinquennial Survey of Current Business. A detailed manual on BEA’s IO data is provided by [Horowitz and Planting \[2006\]](#)

SI.1.2.1. Overview

The raw data shows monetary transactions between industries. It covers also final demand sectors and public services. For this project, I use tables from 1977-2007. I made a series of conversions and processing steps to harmonize the data. Over time, industrial classification systems and technical methods of data processing, formatting and saving have changed. The earliest tables are only available in text format that was manually edited to make it readable for statistical software. A further challenge arises from changes in the classification system, most pronounced in the conversion from SIC to NAICS.

The final data structure is a series of quadratic matrices for each period that show the monetary transactions between industries in NAICS 2002 codes. The data is also used to create a panel of industry level indicators, i.e. outputs, inputs, and growth rates.

SI.1.2.2. Processing steps in detail

Here, I explain single steps of data processing. Again, the steps are consistent with the R-code provided in the data publication [Hötte \[2021a\]](#).

Step 1: For each period, the IO data are downloaded separately. Some manual harmonization and data conversions were made to obtain machine-readable, harmonized data tables. For example, the very old data is only available in text format which is not ready to be read by statistical software. The more recent table are Excel files with many macros and text-explanations. All tables were reformatted individually. The scripts are available in the data publication. After this step, all tables have a uniform format which is a long 3-column table with column (1) as producer ID, column (2) as user ID, and column (3) indicating the monetary value of the goods that flow from producer to user.

Step 2: I created large quadratic matrices with rows as producers and columns as users. The entries $flow_{i,j,t}^{out,\mu}$ are flows of goods from i to j . Hence, column-wise reading indicates the composition of inputs used by sector j and row-wise reading indicates the composition of customer industries to which industry i delivers.

Step 3: This is an intermediate step. All concordance and IO-to-industry conversion tables have to be harmonized. Again, some of the data are not machine readable. Moreover, all codes need to be harmonized to obtain a mapping from IO codes for each period to 2002-NAICS codes. Some of the IO codes map to multiple industries. In this step, tables were created where each row indicates an IO code and all NAICS codes to which the IO maps.

Step 4: NAICS-based IO tables were harmonized and consistency checks were done. For example, I tested whether the differences in the tables e.g. regarding the sector coverage are negligible. Some normalizations of IO-flows to input (output) shares were made through division by row (column) sum. The full 6-digit list is used as row and column names.

Not every time snapshot has a full sector coverage. This is a result of reclassification issues, obsolescence and introduction of new sectors. For example, some of the finely granulate computer industries were not yet existing in 1977. For these cases, empty vectors are included to present missing sectors to ensure that matrices have same dimensionality.

Additional steps of harmonization are done. Rows represent the range of inputs that is used, columns represent customers. After this step, NAICS 6-digit data on IO flows, sector weights (row and column sums), input shares (measured in percentage points), 6-digit distance matrix computed by the input-share dissimilarity are obtained.

Step 5: Harmonization of quadratic NAICS 2002 matrices. The matrices are 1179×1179 matrices of 6-digit industries. Empty rows and columns are included for industries that are not producing in some t , for example if an industry was not yet existing or disappeared over time.

Step 6: For each t , I create NAICS \times NAICS matrices with flows of goods $flow_{ij,t}^{\mu,in}$ as entries.

SI.1.2.3. Technical and conceptual issues

General remarks about IO codes, NAICS and SIC The original IO data in early years uses IO codes which are an internal metric of the accounting system used to construct social accounting matrices (SAM). These codes are converted into industry codes (SIC and NAICS). The classification system has changed over time. Fortunately,

the IO-codes in the raw IO tables are largely consistent across time. I converted the accounting codes into SIC and from SIC into NAICS or directly into NAICS if such mapping is available.

How to make a decision about the set of economic sectors to be considered?

The accounting matrices include also dummy industries like private household industries (not same as personal consumption expenditures), government industries, and special positions (e.g. *Non-comparable imports*, *Scrap*, *Rest of world* adjustments, inventory valuation adjustments). These positions are required to ensure completeness in the calculation of GDP which is one of the original purposes of the data sets) [cf. [Horrowitz and Planting, 2006](#), Chap. 4].

As a pragmatic rule, all final demand sectors were kept that can be mapped to NAICS codes. An output link to final demand as customer reveals information about the technological capabilities of the producer. As an input link, final demand is not relevant because it will not appear as a production input.

For the main analyses presented in this paper, the subset of industries is further reduced because industries were only included if having a non-zero patent stock $A_{i,t}^r$ and goods output $A_{i,t}^u$ in all periods. This excludes the majority of final demand positions.

Another concern is the consistency of the data across time. This is partly addressed by normalizations and the focus of analysis on relative industry differences instead of quantitative cross-time comparisons. In the panel data analyses, I control for FE and make checks using clustered standard errors.

Note that some of the rows and columns are empty for some periods at the 6-digit level. This is a result of the harmonization procedure to uniform NAICS codes. This may happen if an industry disappears or a new industry emerges. Often, the emergence (disappearance) of an industry is associated with a split (merge) of pre-existing industries. This problem arises more often for final demand and service industries. I cope with this problem by the use of more aggregate data and robustness checks.

How to deal with accounting codes that are mapped into multiple SIC sectors?

Some of the accounting codes are associated with multiple SIC sectors, i.e. multiple industries have been aggregated into one accounting position. Information about the strengths of links to each of these these subsectors is missing. For reasons of simplification, I assume that the accounting position is equally related to all of them. The strengths of single links is weighted uniformly by the number of sectors. For example, the IO code 020401 (“Fruits”) is linked to 9 SIC sectors (0171, 0172, 0174,

0175, *0179, *019, *0219, *0259, *029). The links are weighted by factor 1/9.

How to deal with inconsistencies across time in changing classification systems? The accounting codes of 1977 and 1982 data are mapped to SIC 1987. All mappings from accounting positions to SIC are based on the 1987 data after having ensured that the accounting codes are consistent across time. Also the vast majority of IO-to-SIC-mappings is consistent in 1977 and 1992 data. For 1977 some minor deviations exist but these are largely explainable by adjustments in the SIC system between 1977 to 1987. Some of the old SIC industries do not exist any longer. A reconstruction is practically not feasible with reasonable effort given that the value added of higher precision is negligibly small if existing at all. The 1977 IO-SIC mapping is only used when 1987-data is not available.

In the 2002 NAICS file, some IO codes are mapped to a very high number of sub-sectors. This is for example the case for aggregate positions such as retail and wholesale trade and construction. I kept them in the mapping. It should be noted that an accounting position that has a link to more than hundred 6-digit NAICS industries is not necessarily meaningful. I cope with this problem by a series of robustness checks using only a subset of the data, higher levels of aggregation and rounding of IO links that fall below a certain threshold.

The more recent versions of the classification systems are more detailed. I used equal weights when one coarse industry mapped to several more detailed industry when using another (typically more recent) classification system. Hence, the transaction volume is equally distributed across sub-sectors.

Which NAICS version to use? I use NAICS 2002 codes. These codes have a direct mapping to SIC 1987 codes.

SI.1.3. Combined datasets

To study the evolution of industries and link formation processes, I constructed a panel data set covering NAICS industries at the detailed level and the time horizon 1976-2006. For the main analysis, I use a balanced panel of 4-digit level data but the data are available at the 6-digit level, too. The network raw data is used to compile measures for the industry size, connectivity and spillover effects. For the purpose of data exploration, I also compiled a series of network centrality measures; some of them are available in the data publication [[Hötte, 2021a](#)].

For further test, two types of external data are additionally included in for robustness checks. The first is data on trade indicators, the second are industry characteristics provided by the US Bureau of Labor and Statistics (BLS). These additional data is not used as "baseline data" in the main analyses because it significantly reduces the number of industries (and in case of the BLS data time periods) covered and truncates the data to sectors that mainly belong to the manufacturing sector.

First, trade and import competition have been extensively studied as drivers of industrial restructuring during the past decades [e.g. Bernard et al., 2006]. To control for this impact, import competition indices provided by Bernard et al. [2006] were included in robustness checks of the regression analyses.³⁵

It is available for the time horizon from 1972-2007. The data until 1987 uses 4-digit SIC codes for the sector classification. For later periods, 6-digit NAICS 2002 codes have been used which is already consistent with the network data compiled before.

The SIC codes were transformed into NAICS 2002. Bernard et al. [2006] provided an own concordance table which is largely consistent with the concordance provided by the US Census Bureau used before (see SI.1.1). To be consistent with the steps above, I use the US Census Bureau concordance for the transformation. This enables me to map the SIC codes to 486 different 6-digit NAICS industries.

For additional robustness checks, a second data set on industry characteristics is used. It reduces the sector and time coverage considerably. Only 123 industries are covered and the data is only available for the periods after 1987. The data is downloaded from the BLS.³⁶ This data contains detailed information about multi-factor productivity, employment, wages, capital shares etc. at the detailed 6-digit level for a subset of manufacturing industries.

The data is available year-wise. To make the format consistent, I compiled 5-year averages (e.g. time window from 1988-1992, 1993-1996, etc.) and growth rates based on the these averages. I also include the single-year entry for 1987, but control for time FE to capture this inconsistency. Creating a balanced panel, i.e. removing all industries without patents, reduces the data size further.

These indicators are used (1) as control variables in the regression analysis, (2) multi-factor productivity is also used as dependent variable. The findings are either statistically not significant (mainly due to the reduced coverage), but qualitatively largely consistent with the main analysis.

³⁵The data were downloaded from https://sompks4.github.io/sub_data.html and <https://faculty.som.yale.edu/peterschott/international-trade-data/>.

³⁶<https://www.bls.gov/mfp/>

SI.1.4. Regression analyses

SI.1.4.1. Data pre-processing for regression analyses

The data is pre-processed before the regressions are run.

First, the data is log-linearized using the formula $\text{sign}(x) \cdot \log(1 + |x|)$ where $|x|$ is the absolute of x and $\text{sign}(x)$ returns its sign. Log-linearization is used to cope with skewness and to make the data more robust to outliers. The formula using the absolute value and adding one allows to make the logarithm of data ranging $\in [-1, 1]$ comparable with data > 1 or < -1 . Multiplying the term by the sign of x keeps its original sign.

Second, the data is scaled before used in the regressions using the z-score normalization $\frac{x - \text{mean}(x)}{\text{std}(x)}$ where $\text{mean}(x)$ returns the mean of x and $\text{std}(x)$ its standard deviation. Scaling makes the coefficients of the regression comparable.

SI.1.4.2. Selection of explanatory variables

Network-based indicators can be compiled in many different ways. For example, the raw stock of patents or a citation weighted measures as a proxy to control for the quality of a patent can be used. It is possible to use upstream or downstream similarities or combined measures. For the similarity, different measures (e.g. Euclidean, Canberra, Cosine) are available. Moreover, it is possible to include different time lags for the spillover variable as it was done by other authors [e.g. [Acemoglu et al., 2016](#)].

I applied a step-wise procedure to select the final subset of variables that is included in the regression analyses. First, on the basis of tests for the explanatory power and conceptual considerations, some pre-selections are made. I find that up- and downstream similarity measures and network indicators are highly correlated such that multicollinearity could arise if using both of them. I restrict the analysis to the measures computed on the basis of the cosine similarity because it is one of the most common measures used in the statistical analysis of networks (see also [3.3](#)). Simple regression tests and robustness checks confirm this choice.

In the final analysis, I use only one-period time lagged data. Other authors argued that the inclusion of time lags may smooth highly volatile patterns of patenting activity [e.g. [Acemoglu et al., 2016](#)]. This issue is less problematic in this analysis because 5-year averages are used. Moreover, including longer time lags can only be computed for more recent periods which reduces the sample size. I made some simple regression tests including longer lags but could not find any indication that this contributes to the explanatory power.

Further reductions of the set of variables were made on the basis of simple correlation checks (see Fig. A.3). For example, I found the citation-weighted patent stock to be highly correlated with the raw patent stock.³⁷ Some network measures are highly correlated. For example, the strength is highly correlated with the PageRank. Whenever two variables are highly correlated, I dropped one of them from the final set of potential regression variables making the choice on the basis of conceptual considerations.

After this pre-selection of potential explanatory variables, a second data-driven approach is applied to reduce the complexity of the regression approach. This is motivated by two reasons. First, it reduces the number of the degrees of freedom and makes the analysis more robust (known as bias-variance trade-off [cf. Bishop, 2007]). Second, it facilitates the interpretation.

To identify the final regression equation, I apply a data-mining procedure. In particular, I am using a mixture of a step-wise selection procedure based on the minimal Bayesian Information Criterion (BIC) using the R-function `stepAIC()` [Rigby and Stasinopoulos, 2005]. This procedure takes a regression equation with a high number of explanatory variables as input and returns a reduced version containing the controls with the highest explanatory power only. To specify the input equation for this procedure, I made pairwise correlation tests to avoid multicollinearity issues and removed highly correlated pairs keeping the variable that is economically more meaningful and/ or performed better in regression exercises with reduced form models.

Finally, I reduced the set of variables to those that were consistently significant across different data sets, mutually sufficiently uncorrelated, easy to interpret and in line with theory.

SI.2. Additional statistics and tables

SI.2.1. Descriptive statistics

Looking at manufacturing sectors only (see SI.2 and Table SI.2), the difference between layer μ and τ is less pronounced. This is in line with previous findings that have shown that manufacturing industries are above-average patent-intensive [Blank and Kappos, 2012].

Table SI.3 offers a more aggregate ranking at the 3-digit level. A rise of chemical manufacturing, scientific and technical services is observable that came at the cost of

³⁷This might be different when looking at data more disaggregate in time and individual (e.g. firm-level data).

	Input-output			Patent			Combined		
	1981	1991	2006	1981	1991	2006	1981	1991	2006
<i>Upstream network</i>									
Density	0.15	0.15	0.18	0.10	0.13	0.15	0.03	0.04	0.05
Avg. degree	11.48	11.05	13.19	7.43	9.61	11.36	2.25	2.85	3.61
Avg. strength	0.65	0.65	0.58	0.59	0.60	0.65	0.02	0.02	0.02
Avg. weight	1.74	1.71	1.75	3.17	2.40	1.99	24.21	18.58	18.00
Reciprocity	0.24	0.22	0.56	0.27	0.30	0.27	0.14	0.12	0.30
Diameter	2.00	2.00	2.00	4.00	3.00	3.00	5.00	3.00	3.00
Assort. by degree	-0.10	-0.10	-0.18	-0.14	-0.20	-0.17	-0.07	-0.10	-0.20
<i>Downstream network</i>									
Density	0.17	0.16	0.18	0.09	0.11	0.13	0.03	0.03	0.04
Avg. degree	12.73	12.00	13.29	6.49	8.47	9.84	1.93	2.40	2.95
Avg. strength	0.57	0.58	0.53	0.60	0.64	0.67	0.01	0.01	0.02
Avg. weight	1.74	1.71	1.75	3.19	2.44	1.99	23.05	19.64	17.69
Reciprocity	0.31	0.27	0.54	0.20	0.22	0.19	0.18	0.16	0.22
Diameter	2.00	2.00	2.00	4.00	3.00	3.00	5.00	3.00	3.00
Assort. by degree	-0.10	-0.10	-0.18	-0.14	-0.20	-0.17	-0.07	-0.10	-0.20

Notes: These statistics are computed on the weighted upstream networks. For an introduction to network characteristics see [Jackson \[2008\]](#), [Csardi and Nepusz \[2006\]](#). "Assort." is the abbreviation for assortativity. Data: 4-digit balanced panel, manufacturing industries only.

Table SI.2: Aggregate network statistics over time (manufacturing).

declining market shares of the food manufacturing and petroleum and coal product industries. The process of structural change from a high relevance of primary and secondary sectors towards an increasing importance of the service industry can be also seen at the 2-digit level (Table [SI.4](#)).

<i>Top-10 industries by Aggregate output ($A_{i,t}^{\mu}$)</i>								
1981			1991			2006		
1	Scient. & Tech. Srv.	541 4.95	Scient. & Tech. Srv.	541 6.98	Scient. & Tech. Srv.	541 6.39		
2	Petroleum & Coal Prod.	324 3.66	Chemical Manufacturing	325 3.66	Real Estate	531 4.22		
3	Utilities	221 3.58	Real Estate	531 2.97	Chemical Manufacturing	325 3.18		
4	Chemical Manufacturing	325 3.41	Utilities	221 2.71	Mngm. Comp. & Enterpr.	551 2.91		
5	Food Manufacturing	311 3.08	Fabricated Metal Prod.	332 2.49	Transport. Equ. Mnft.	336 2.87		
6	Fabricated Metal Prod.	332 3.02	Food Manufacturing	311 2.48	Admin. & Support Serv.	561 2.53		
7	Primary Metal Manufactur.	331 3.01	Transport. Equ. Mnft.	336 2.48	Wholes., Durable Goods	423 2.46		
8	Transport. Equ. Mnft.	336 2.61	Computer & Electr. Prod.	334 2.19	Food Manufacturing	311 2.34		
9	Spec. Trade Contractors	238 2.06	Primary Metal Manufactur.	331 1.96	Insurance Carriers & Act.	524 2.11		
10	Real Estate	531 1.96	Paper Manufacturing	322 1.90	Secur. & Financ. Investm.s	523 2.06		
Quartiles:								
0.2225, 0.68, 1.3275			0.2625, 0.665, 1.305			0.2125, 0.515, 1.4375		

<i>Top 10 industries by Patent stock ($A_{i,t}^{\tau}$)</i>								
1981			1991			2006		
1	Computer & Electr. Prod.	334 16.75	Computer & Electr. Prod.	334 24.58	Computer & Electr. Prod.	334 30.39		
2	Chemical Manufacturing	325 11.94	Chemical Manufacturing	325 9.27	Chemical Manufacturing	325 6.13		
3	Transport. Equ. Mnft.	336 6.13	Transport Equ. Mnft.	336 5.22	Transport. Equ. Mnft.	336 5.26		
4	Machinery Manufacturing	333 4.62	Scient. & Tech. Srv.	541 3.00	Scient. & Tech. Srv.	541 3.87		
5	Petroleum & Coal Prod.	324 3.66	Machinery Manufacturing	333 2.80	Telecommunications	517 3.41		
6	Fabricated Metal Prod.	332 2.55	Telecommunications	517 2.59	Machinery Manufacturing	333 2.49		
7	Scient. & Tech. Srv.	541 2.54	Petroleum & Coal Prod.	324 2.56	Fabricated Metal Prod.	332 2.48		
8	Electr. Equ., Appl. &	335 2.34	Fabricated Metal Prod.	332 2.52	Electr. Equ., Appl. &	335 1.97		
9	Telecommunications	517 2.02	Electr. Equ., Appl. &	335 2.41	Transit & Passng. Transp.	485 1.38		
10	Paper Manufacturing	322 1.47	Paper Manufacturing	322 1.46	Paper Manufacturing	322 0.98		
Quartiles:								
0.02, 0.12, 0.67			0.02, 0.145, 0.51			0.02, 0.15, 0.54		

Notes: Each period shows a block of three columns indicating (1) the industry name, (2) the 3-digit NAICS code, (3) the normalized industry size $A_{i,t}^{\alpha}$. The average size normalized size equals one. Data: 3-digit level balanced panel.

Table SI.3: Ranking of the Top-10 industries by industry size ($A_{i,t}^{\alpha}$) over time.

<i>Top-10 industries by Aggregate output ($A_{i,t}^{\mu}$)</i>								
1981			1991			2006		
1	Mnft. (metal & machin.)	33 3.95	Mnft. (metal & machin.)	33 3.53	Mnft. (metal & machin.)	33 3.27		
2	Manufact. (nat. product)	32 3.88	Manufact. (nat. product)	32 3.44	Manufact. (nat. product)	32 2.89		
3	Manufact. (food proc.)	31 1.59	Scient. & Tech. Srv.	54 2.05	Finance & Insurance	52 2.01		
4	Scient. & Tech. Srv.	54 1.45	Manufact. (food proc.)	31 1.27	Scient. & Tech. Srv.	54 1.91		
5	Agr., Forest., Fish. & Hunt.	11 1.36	Real Est., Rental & Leas.	53 1.26	Real Est., Rental & Leas.	53 1.69		
6	Construction	23 1.14	Finance & Insurance	52 1.23	Wholesale Trade	42 1.19		
7	Mining	21 1.08	Agr., Forest., Fish. & Hunt.	11 1.05	Manufact. (food proc.)	31 0.98		
8	Utilities	22 1.05	Construction	23 0.93	Information	51 0.97		
9	Transp. & Warehous.	48 0.85	Transp. & Warehous.	48 0.91	Mngm. Comp. & Enterpr.	55 0.87		
10	Real Est., Rental & Leas.	53 0.83	Wholesale Trade	42 0.83	Waste Mngm. Srv.	56 0.85		
Quartiles:								
0.4275, 0.825, 1.195			0.48, 0.8, 1.23			0.43, 0.8, 1.19		

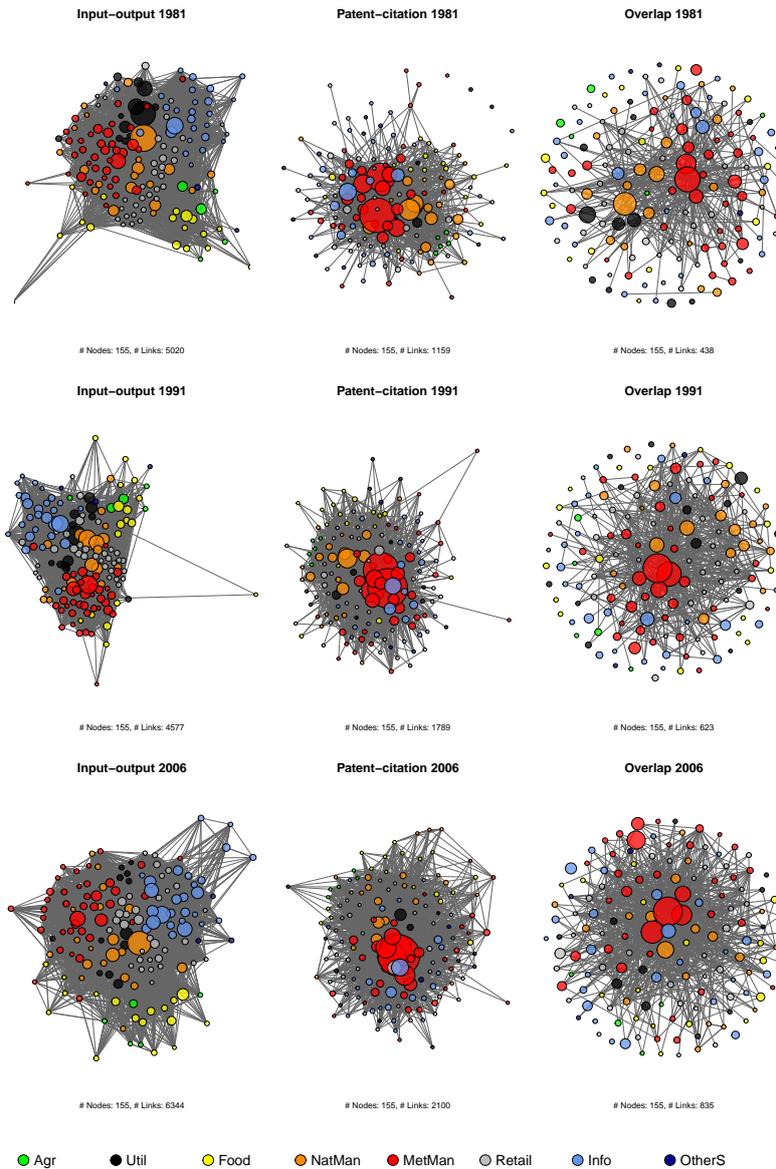
<i>Top 10 industries by Patent stock ($A_{i,t}^{\tau}$)</i>								
1981			1991			2006		
1	Mnft. (metal & machin.)	33 10.53	Mnft. (metal & machin.)	33 12.25	Mnft. (metal & machin.)	33 13.69		
2	Manufact. (nat. product)	32 5.85	Manufact. (nat. product)	32 4.46	Manufact. (nat. product)	32 2.68		
3	Information	51 1.23	Information	51 1.30	Information	51 1.68		
4	Scient. & Tech. Srv.	54 0.80	Scient. & Tech. Srv.	54 0.95	Scient. & Tech. Srv.	54 1.22		
5	Mining	21 0.75	Mining	21 0.52	Transp. & Warehous.	48 0.45		
6	Finance & Insurance	52 0.44	Finance & Insurance	52 0.39	Mining	21 0.34		
7	Wholesale Trade	42 0.42	Manufact. (food proc.)	31 0.33	Finance & Insurance	52 0.31		
8	Manufact. (food proc.)	31 0.33	Transp. & Warehous.	48 0.26	Manufact. (food proc.)	31 0.22		
9	Agr., Forest., Fish. & Hunt.	11 0.16	Wholesale Trade	42 0.23	Wholesale Trade	42 0.15		
10	Retail Trade	44 0.15	Retail Trade	44 0.12	Real Est., Rental & Leas.	53 0.08		
Quartiles:								
0.035, 0.15, 0.595			0.03, 0.23, 0.52			0.0125, 0.115, 0.4225		

Notes: Each period shows a block of three columns indicating (1) the industry name, (2) the 3-digit NAICS code, (3) the normalized industry size $A_{i,t}^{\alpha}$. The average size normalized size equals one. Data: 3-digit level balanced panel.

Table SI.4: Ranking of the Top-10 industries by industry size ($A_{i,t}^{\alpha}$) over time.

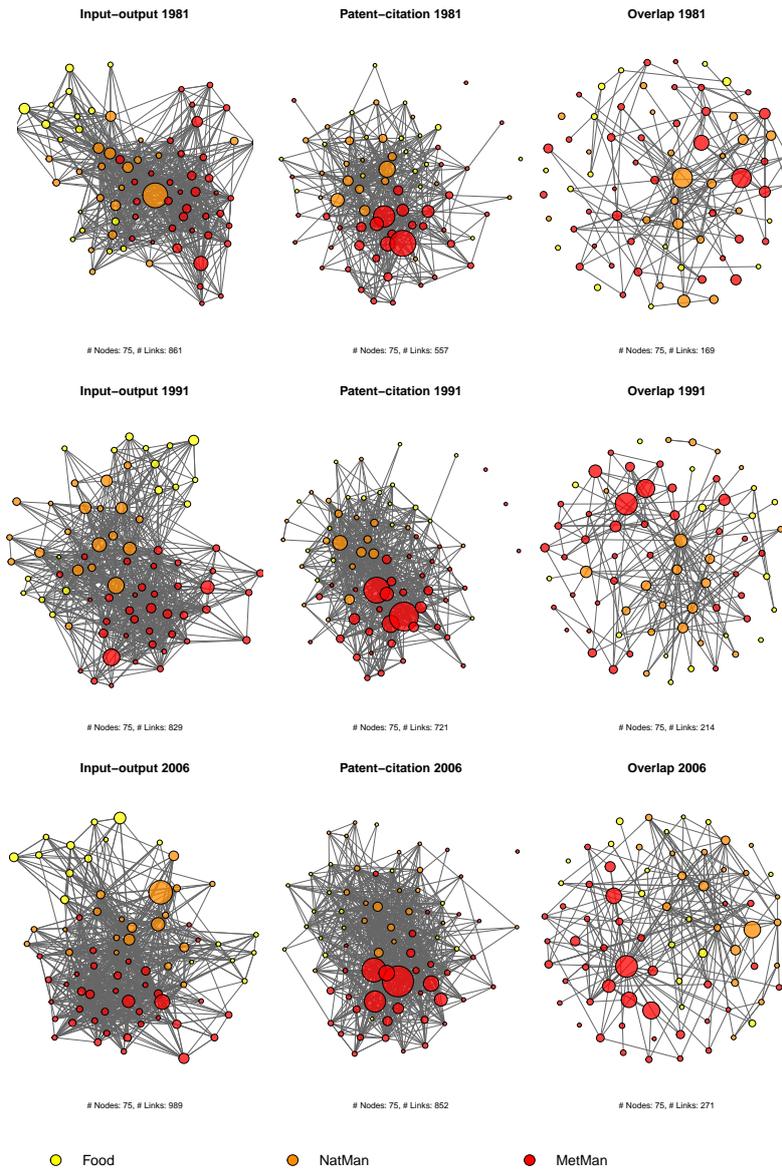
SI.2.2. Network plots

On the following pages, additional network plots at different levels of aggregation and different types of data subsets are provided.



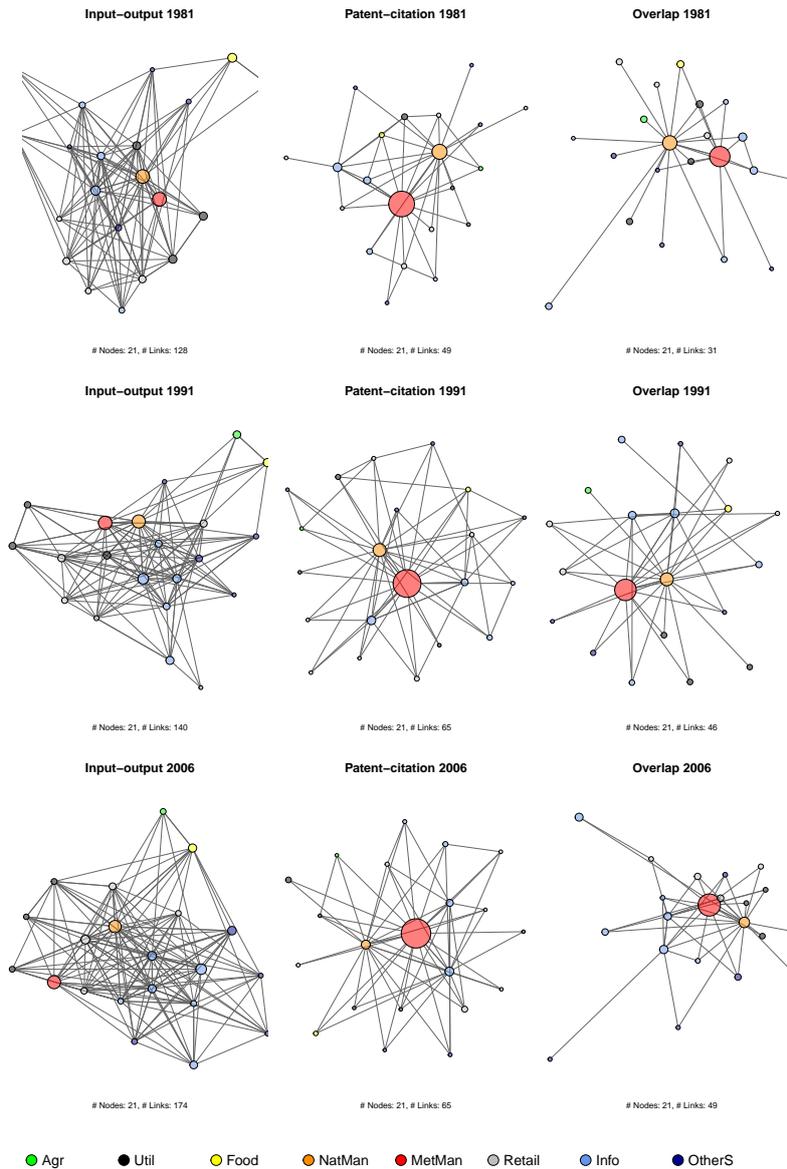
Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only out-going links are considered. Plots of the downstream network are available in [SI.2](#). Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Agr is Agriculture (1); Util are Mining and Utilities (2); Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33); Retail are Trade, Retail and Transportation sectors (4); Info are Information, Financial, Management and Administrative Services (5); OtherS are Other services and Public Sector (6-9). Data: Balanced panel of 4-digit industries.

Figure SI.1: Network plots at different periods for 4-digit level data (downstream).



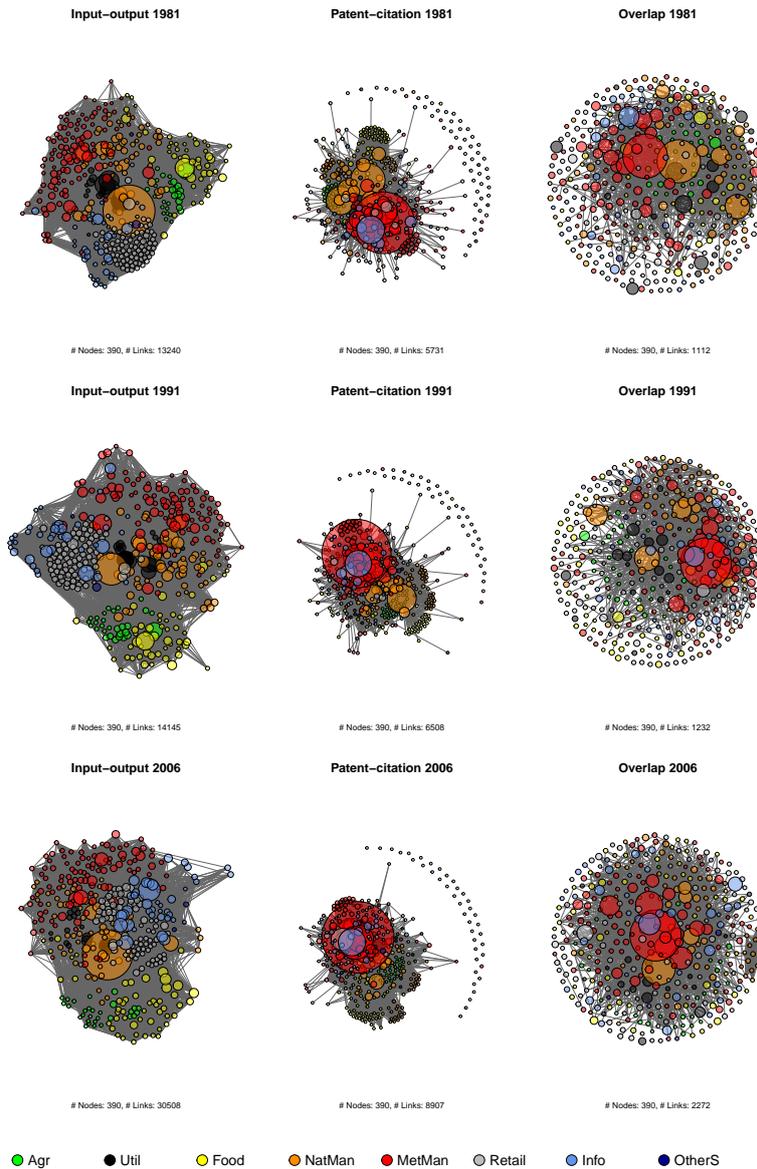
Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only in-going links are considered. Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33). Data: Balanced panel of the subset of manufacturing 4-digit industries.

Figure SI.2: Network plots at different periods for 4-digit level data (manufacturing).



Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only in-going links are considered. Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Agr is Agriculture (1); Util are Mining and Utilities (2); Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33); Retail are Trade, Retail and Transportation sectors (4); Info are Information, Financial, Management and Administrative Services (5); OtherS are Other services and Public Sector (6-9). Data: Balanced panel of 2-digit industries.

Figure SI.3: Network plots at different periods for 2-digit level data.



Notes: The overlap network shows nodes as being connected if they are connected on both layers. Only in-going links are considered. Self-citations and within-sector IO flows are not shown. The colors indicate broad industrial categories classified by the first (and second) digit level (indicated in parenthesis) where Agr is Agriculture (1); Util are Mining and Utilities (2); Food are Food processing and Textile industries (31); NatMan is Non-metal manufacturing (32); MetMan is Metallic and Machinery manufacturing (33); Retail are Trade, Retail and Transportation sectors (4); Info are Information, Financial, Management and Administrative Services (5); OtherS are Other services and Public Sector (6-9). Data: Balanced panel of 6-digit industries.

Figure SI.4: Network plots at different periods for 6-digit level data.

SI.2.3. Additional regression results

SI.2.3.1. Industrial hierarchies

The Tables [SI.5-SI.7](#) show the results of regression analyses explaining the evolution of sector hierarchies for different subsets of data and levels of aggregation. Table [SI.5](#) shows the results including trade variables. Trade variables capture industry-level import pressure and its evolution which is one explanation for industrial restructuring that is prominently discussed in the literature [e.g. [Bernard et al., 2006](#)]. Trade variables include import penetration $IP_{i,t}$ and import penetration from poor countries $IP5_{i,t}$ (defined by <5% of US per capita GDP) and the share of low-wage country imports in total industry imports $LWC_{i,t}$. The data has been taken from [Bernard et al. \[2006\]](#). The trade data is not available for all industries which reduces the sample size significantly.

Table [SI.6](#) shows the results using data on all 4-digit industries, i.e. including industries that have zero accounts for the patent stock or output in some t . Table [SI.7](#) show the results at the 3-digit level.

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
with trade		with trade		with trade		with trade		
$A_{i,t-1}^\mu$	0.2444*** (0.0192)	0.3027* (0.1373)	0.2376*** (0.0207)	0.3529* (0.1553)	-0.0168 (0.0144)	0.0083 (0.0153)	-0.0055 (0.0148)	-4e-04 (0.0156)
$A_{i,t-1}^\tau$	0.0821 (0.0594)	0.0304 (0.0924)	0.0165 (0.0588)	-0.0482 (0.0846)	0.1236** (0.0387)	0.6768*** (0.0542)	0.0587 (0.033)	0.6821*** (0.0497)
$D_{i,t-1}^{\mu,d}$	0.023 (0.0173)	0.1062** (0.0343)	-0.043*** (0.0129)	-0.0938** (0.0288)	0.0116 (0.0117)	-0.0018 (0.0081)	0.0058 (0.0124)	-0.0126* (0.006)
$D_{i,t-1}^{\tau,d}$	-0.0021 (0.0336)	0.0395 (0.0332)	0.0374 (0.0233)	-7e-04 (0.03)	-0.1352* (0.0529)	0.0175 (0.0148)	0.1319** (0.0406)	0.036** (0.0137)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0628*** (0.0123)	-0.0301 (0.0162)	0.0132 (0.0185)	0.0287 (0.0198)	-0.0158 (0.0137)	0.0109 (0.0066)	0.0158 (0.014)	-6e-04 (0.012)
$Spill(A)_{i,t-1}^{\tau,d}$	8e-04 (0.0118)	-0.0055 (0.0171)	1e-04 (0.0119)	0.0427* (0.0187)	5e-04 (0.0238)	0.0035 (0.0059)	-0.0426* (0.0181)	-0.006 (0.0063)
$IP5_{i,t-1}$	-0.0021 (0.0188)	0.0359* (0.0177)	-0.0025 (0.018)	0.0261 (0.0151)	-0.0182 (0.0126)	-0.0025 (0.0053)	-0.0155 (0.0119)	0 (0.0048)
$IP_{i,t-1}$	0.0284 (0.0147)	-0.022 (0.0145)	0.0284* (0.014)	-0.0097 (0.0132)	0.0177* (0.0085)	-0.003 (0.004)	0.0116 (0.0097)	-0.0029 (0.0038)
$LWC_{i,t-1}$	0.0201 (0.0148)	-0.0446* (0.0198)	0.0267 (0.0159)	-0.0251 (0.0138)	-0.0038 (0.012)	-0.0038 (0.0061)	-0.0093 (0.0124)	-0.0016 (0.006)
R^2	0.4607	0.2107	0.4455	0.2454	0.058	0.6814	0.0825	0.6877
N	458	458	458	458	458	458	458	458
Average	-0.0215	0.6169	-0.0215	0.6169	-0.0475	0.5696	-0.0475	0.5696

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Notes: The model is estimated as linear OLS model including time and industry FE and using two-ways clustered SE. The explanatory variables have been selected in a step-wise selection procedure searching for the highest explanatory power measured by the BIC (see text for further detail). Regressions are run separately for spillovers and network-based variables computed on the basis of upstream ($d = in$) and downstream ($d = out$) links. Data: 4-digit, balanced panel of subset of industries for which trade data is available.

Table SI.5: Regression results explaining the evolution of industrial hierarchies (4-digit including trade).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
$A_{i,t-1}^\mu$	0.1906*** (0.0214)	0.5126*** (0.0342)	0.1841*** (0.0217)	0.5035*** (0.0345)	0.0019 (0.0105)	1e-04 (0.0048)	0.0075 (0.0102)	-5e-04 (0.0048)
$A_{i,t-1}^\tau$	0.0088 (0.0324)	0 (0.0428)	-0.0188 (0.0347)	-0.0261 (0.0405)	0.0592* (0.0231)	0.6705*** (0.0353)	0.0164 (0.0239)	0.6693*** (0.0343)
$D_{i,t-1}^{\mu,d}$	0.1058*** (0.0141)	0.1128*** (0.0142)	-0.0262* (0.0132)	-0.0405*** (0.0122)	0.0097 (0.0086)	0.0017 (0.0028)	-0.0015 (0.0085)	-0.0016 (0.0017)
$D_{i,t-1}^{\tau,d}$	-0.0111 (0.0269)	0.0099 (0.0275)	-0.009 (0.0272)	0.0262 (0.0264)	-0.2422*** (0.0374)	-0.0011 (0.0059)	0.2438*** (0.0437)	0.0322*** (0.007)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.035* (0.0147)	-0.0248* (0.0119)	0.118*** (0.0125)	0.1098*** (0.0135)	0.0062 (0.009)	0.0019 (0.0021)	-0.0047 (0.0115)	0.0019 (0.0024)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.0242 (0.0129)	0.0184 (0.0135)	-0.0075 (0.0142)	0.009 (0.0142)	0.1902*** (0.0337)	0.0057 (0.0031)	-0.1084*** (0.0291)	-0.0041 (0.0027)
R^2	0.2768	0.2747	0.2805	0.2785	0.1022	0.6687	0.0663	0.6723
N	1902	1902	1902	1902	1902	1902	1902	1902
Average	0.0095	0.5335	0.0095	0.5335	-0.0194	0.2901	-0.0194	0.2901

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Notes: The model is estimated as linear OLS model including time and industry FE and using two-ways clustered SE. The explanatory variables have been selected in a step-wise selection procedure searching for the highest explanatory power measured by the BIC (see text for further detail). Regressions are run separately for spillovers and network-based variables computed on the basis of upstream ($d = in$) and downstream ($d = out$) links. Data: 4-digit, unbalanced panel.

Table SI.6: Regression results explaining the evolution of industrial hierarchies (4-digit unbalanced panel).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\mu)$	$A_{i,t}^\mu$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$	$gr(A_{i,t}^\tau)$	$A_{i,t}^\tau$
	upstream		downstream		upstream		downstream	
$A_{i,t-1}^\mu$	0.204*** (0.031)	0.5143*** (0.0469)	0.1855*** (0.0306)	0.5282*** (0.0416)	0.0036 (0.0145)	0.0107 (0.0073)	0.0059 (0.0133)	0.01 (0.0077)
$A_{i,t-1}^\tau$	-0.0349 (0.0608)	-0.0064 (0.057)	-0.1281* (0.0581)	-0.0483 (0.0455)	0.106* (0.0515)	0.6935*** (0.0612)	0.0055 (0.0392)	0.6934*** (0.0566)
$D_{i,t-1}^{\mu,d}$	0.0164 (0.0169)	0.0331** (0.0108)	-0.0694*** (0.0136)	-0.0428** (0.014)	-0.0068 (0.0121)	-6e-04 (0.0055)	0.0069 (0.008)	0.0021 (0.0019)
$D_{i,t-1}^{\tau,d}$	-0.0082 (0.0417)	0.0675* (0.0279)	-0.0246 (0.0451)	-0.0042 (0.0403)	-0.1251*** (0.036)	0.0171. (0.0091)	0.2296*** (0.0449)	0.0448*** (0.0127)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0964*** (0.0169)	-0.0289 (0.0199)	-0.0055 (0.0157)	0.0238 (0.028)	-0.0056 (0.0091)	0.0046 (0.0045)	-0.0125 (0.0141)	0.0082. (0.0046)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.0189 (0.0176)	0.0059 (0.0156)	0.0138 (0.0201)	0.0206 (0.0162)	0.0606* (0.0252)	0.0042 (0.0044)	-0.0083 (0.0237)	-0.0028 (0.0036)
R^2	0.3043	0.2761	0.2884	0.2953	0.072	0.7192	0.1081	0.7274
N	396	396	396	396	396	396	396	396
Average	0.0058	0.5738	0.0058	0.5738	-0.0334	0.3498	-0.0334	0.3498

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Notes: The model is estimated as linear OLS model including time and industry FE and using two-ways clustered SE. The explanatory variables have been selected in a step-wise selection procedure searching for the highest explanatory power measured by the BIC (see text for further detail). Regressions are run separately for spillovers and network-based variables computed on the basis of upstream ($d = in$) and downstream ($d = out$) links. Data: 3-digit, balanced panel.

Table SI.7: Regression results explaining the evolution of industrial hierarchies (3-digit balanced panel).

SI.2.3.2. Link formation at the extensive margin

Here, additional regression results on link formation processes are provided. Table SI.8 shows the results for the subset of manufacturing sectors at the 4-digit level of aggregation. Table SI.9 shows the results for all 4-digit industries including those with missing data for some periods in time. Table SI.10 shows the results including trade data which reduces the subset of industries.

	market layer				Patent-layer			
	(1) $Pr(\Delta w_{ij,t}^{\mu} \gg 0)$	(2) $Pr(\Delta w_{ij,t}^{\mu} \ll 0)$	(3) $Pr(\Delta w_{ij,t}^{\tau} \gg 0)$	(4) $Pr(\Delta w_{ij,t}^{\tau} \ll 0)$	(5) $Pr(\Delta w_{ij,t}^{\tau} \gg 0)$	(6) $Pr(\Delta w_{ij,t}^{\tau} \ll 0)$	(7) $Pr(\Delta w_{ij,t}^{\alpha} \gg 0)$	(8) $Pr(\Delta w_{ij,t}^{\alpha} \ll 0)$
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	-0.0018 (0.0101)	-0.0207*** (0.0068)	3e-04 (0.0096)	-0.0213** (0.0066)	0.0039 (0.0084)	-0.0099 (0.0081)	-0.0059 (0.0081)	-0.0014 (0.0075)
$1(\Delta w_{ij,t-1}^{\mu,d} \gg 0)$	0.0749*** (0.0108)	0.088*** (0.007)	0.0897*** (0.0101)	0.0565*** (0.0068)	2e-04 (0.0043)	-0.0021 (0.0043)	-0.0048 (0.0045)	0.0044 (0.0046)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	-0.0068 (0.0118)	-0.0112 (0.0096)	-0.0013 (0.0111)	-0.0113 (0.0094)	0.0732*** (0.0089)	-0.1304*** (0.0102)	0.0736*** (0.0087)	-0.1229*** (0.0098)
$1(\Delta w_{ij,t-1}^{\tau,d} \gg 0)$	1e-04 (0.0084)	-0.0054 (0.0069)	-0.0117 (0.0079)	-0.0019 (0.0063)	0.0944*** (0.0095)	0.1274*** (0.0083)	0.1081*** (0.0101)	0.1261*** (0.0082)
$\sigma_{ij,t-1}^{\mu,d}$	0.007 (0.0148)	0.078*** (0.0131)	-0.0394*** (0.0113)	0.0758*** (0.0092)	0.0073 (0.0085)	-0.0196* (0.0083)	-0.0047 (0.0056)	-0.0114* (0.0054)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	0.0068 (0.006)	0.0371*** (0.0054)	-0.0065 (0.0055)	0.0278*** (0.0044)	0.0056 (0.0037)	-0.0031 (0.0039)	-0.0024 (0.0029)	-0.0045 (0.0031)
$\sigma_{ij,t-1}^{\tau,d}$	0.014* (0.007)	-0.0052 (0.0059)	0.0165* (0.0078)	-0.0044 (0.0069)	0.013* (0.0051)	0.0021 (0.0047)	-0.0041 (0.0054)	0.0105* (0.0051)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	0.0061 (0.0039)	-0.0053 (0.0035)	0.003 (0.0046)	-0.0025 (0.0038)	0.004 (0.0029)	7e-04 (0.0027)	-0.0017 (0.0032)	0.0081** (0.0029)
$A_{i,t-1}^{\mu}$	0.0049 (0.0104)	0.035*** (0.0081)	-0.0019 (0.0117)	0.1237*** (0.0098)	-0.0024 (0.0076)	0.0042 (0.0071)	0.0114 (0.0076)	0.0013 (0.0073)
$A_{j,t-1}^{\mu}$	-0.0186 (0.0112)	0.0086 (0.0087)	-0.0054 (0.009)	-0.0298*** (0.0079)	0.0051 (0.0054)	0.004 (0.0061)	0.002 (0.0061)	0.0043 (0.0063)
$A_{i,t-1}^{\tau}$	-0.1129*** (0.0195)	-0.0539*** (0.0162)	0.0798*** (0.019)	0.0628*** (0.0154)	-0.0025 (0.0192)	0.0806*** (0.0184)	0.0883*** (0.0186)	0.0621*** (0.0168)
$A_{j,t-1}^{\tau}$	-0.0305 (0.021)	-0.0283 (0.0145)	-0.052** (0.0167)	-0.012 (0.014)	-0.0192 (0.0113)	-0.0249* (0.0113)	-0.0164 (0.0104)	-0.0302* (0.012)
$D_{i,t-1}^{\mu,d}$	0.0549*** (0.0051)	0.0094* (0.0046)	0.1638*** (0.0154)	-0.0196 (0.0126)	-0.0012 (0.0042)	0.0089* (0.0039)	0.0055 (0.0098)	0.004 (0.0102)
$D_{j,t-1}^{\mu,d}$	0.0322*** (0.0052)	0.0191*** (0.0039)	0.0689*** (0.0176)	0.1153*** (0.0144)	7e-04 (0.0043)	-1e-04 (0.0039)	0.0044 (0.0118)	-0.0166 (0.0119)
$D_{i,t-1}^{\tau,d}$	-0.01 (0.0076)	0.0028 (0.0063)	-0.0371*** (0.0075)	-0.0074 (0.0058)	0.008 (0.0043)	0.0028 (0.0039)	-0.0113*** (0.0032)	0.0038 (0.0033)
$D_{j,t-1}^{\tau,d}$	0.0033 (0.0075)	-0.0199*** (0.0054)	-0.0071 (0.0078)	0.0044 (0.007)	0.0223*** (0.0053)	0.03*** (0.0045)	0.0113* (0.0049)	0.0311*** (0.0046)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0328*** (0.0084)	-0.0067 (0.0076)	0.0109 (0.0062)	0.0219*** (0.0049)	-0.005 (0.0061)	-0.0082 (0.0061)	0.0049 (0.004)	9e-04 (0.004)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.0041 (0.0088)	-0.0227*** (0.0077)	0.0023 (0.0062)	-0.0117* (0.0057)	-0.0042 (0.0069)	-0.0124 (0.0064)	0.0048 (0.0047)	0.0022 (0.0048)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.008 (0.0052)	-0.0017 (0.0042)	0.0038 (0.0044)	0.0048 (0.0039)	-0.008*** (0.0024)	-0.0044 (0.0023)	-0.0028 (0.002)	-0.0048* (0.0021)
$Spill(A)_{j,t-1}^{\tau,d}$	-0.011 (0.0056)	0.0081 (0.005)	0.0044 (0.0046)	-0.0054 (0.0039)	-0.003 (0.0044)	0.0115** (0.0041)	0.0044 (0.0026)	-7e-04 (0.0025)
R^2	0.0415	0.0376	0.0362	0.0565	0.0185	0.1049	0.0272	0.0931
N	16159	19161	16212	19108	19591	19725	19400	19911
Sample avg.	0.3268	0.4063	0.3324	0.4007	0.2886	0.2628	0.2833	0.2683
Subset avg.	0.0865	0.096	0.0898	0.098	0.0859	0.0815	0.0829	0.0827

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). The sample average is the probability of link formation or breaking in the full sample, while the subset average is the frequency of significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$. Data: 4-digit balanced panel subset of manufacturing.

Table SI.8: Regression of link formation & breaking (manufacturing).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Pr(\Delta w_{ij,t}^{\mu} \gg 0)$	$Pr(\Delta w_{ij,t}^{\mu} \ll 0)$	$Pr(\Delta w_{ij,t}^{\mu} \gg 0)$	$Pr(\Delta w_{ij,t}^{\mu} \ll 0)$	$Pr(\Delta w_{ij,t}^{\tau} \gg 0)$	$Pr(\Delta w_{ij,t}^{\tau} \ll 0)$	$Pr(\Delta w_{ij,t}^{\tau} \gg 0)$	$Pr(\Delta w_{ij,t}^{\tau} \ll 0)$
	upstream		downstream		upstream		downstream	
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	-0.0165*** (0.0028)	-0.0412*** (0.0023)	-0.0406*** (0.0026)	-0.0173*** (0.0021)	-4e-04 (8e-04)	-7e-04 (8e-04)	-0.0017 (9e-04)	-0.0017* (8e-04)
$1(\Delta w_{ij,t-1}^{\mu,d} \gg 0)$	0.0764*** (0.0023)	0.0985*** (0.0018)	0.0681*** (0.0023)	0.058*** (0.0018)	-7e-04 (4e-04)	-6e-04 (3e-04)	-7e-04 (4e-04)	-6e-04 (3e-04)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	-0.0141* (0.0066)	-0.0091 (0.0057)	-0.0115 (0.0062)	-0.0213*** (0.0054)	0.0763*** (0.005)	-0.1946*** (0.0051)	0.0759*** (0.0048)	-0.2262*** (0.0052)
$1(\Delta w_{ij,t-1}^{\tau,d} \gg 0)$	-0.0035 (0.0051)	-0.0082 (0.0044)	0.0016 (0.0048)	-0.0036 (0.0042)	0.159*** (0.0063)	0.229*** (0.0049)	0.2007*** (0.0063)	0.2194*** (0.0048)
$\sigma_{ij,t-1}^{\mu,d}$	0.0044 (0.0026)	0.0316*** (0.0024)	-0.0043 (0.0024)	0.0484*** (0.0022)	0 (5e-04)	1e-04 (4e-04)	-0.001* (4e-04)	-0.0014*** (3e-04)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	0.0142*** (0.0013)	0.0225*** (0.0012)	0.0039** (0.0013)	0.0285*** (0.0011)	5e-04 (2e-04)	0 (2e-04)	0 (2e-04)	-4e-04** (2e-04)
$\sigma_{ij,t-1}^{\tau,d}$	0.002 (0.0016)	0.0025 (0.0015)	0.0263*** (0.0017)	0.0109*** (0.0015)	0.0077*** (6e-04)	-0.0014** (5e-04)	0.0109*** (6e-04)	-0.0011* (5e-04)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	0.0011 (9e-04)	0.0011 (9e-04)	0.0048*** (9e-04)	0.0015 (8e-04)	0.0022*** (3e-04)	-8e-04** (3e-04)	0.002*** (3e-04)	6e-04** (2e-04)
$A_{i,t-1}^{\mu}$	0.0919*** (0.0019)	0.0757*** (0.0019)	0.0564*** (0.0023)	0.1677*** (0.002)	-0.0014** (5e-04)	-0.0018*** (4e-04)	-0.0013** (5e-04)	-6e-04 (4e-04)
$A_{j,t-1}^{\mu}$	-0.0028 (0.0021)	0.015*** (0.0019)	0.0138*** (0.0016)	-0.0062*** (0.0015)	0.0012* (5e-04)	0.001* (4e-04)	-5e-04 (5e-04)	0.0013** (4e-04)
$A_{i,t-1}^{\tau}$	-0.0592*** (0.0053)	-0.0149** (0.0046)	-0.0628*** (0.005)	-0.0376*** (0.0046)	0.0093*** (0.0025)	0.0188*** (0.0021)	0.03*** (0.0025)	0.0122*** (0.0019)
$A_{j,t-1}^{\tau}$	-0.032*** (0.005)	-0.0166*** (0.0041)	-0.02*** (0.0042)	-0.0012 (0.0037)	-0.0011 (0.0015)	-0.001 (0.0016)	-0.003 (0.0015)	-6e-04 (0.0018)
$D_{i,t-1}^{\mu,d}$	-0.0086*** (0.0012)	0.0114*** (0.0013)	0.0148*** (0.0014)	0.0051*** (8e-04)	-0.001*** (2e-04)	-1e-04 (2e-04)	4e-04 (2e-04)	-1e-04 (2e-04)
$D_{j,t-1}^{\mu,d}$	0.0111*** (0.0014)	0.039*** (0.0013)	0.0201*** (0.0019)	0.0418*** (0.0021)	4e-04 (3e-04)	0.0011*** (3e-04)	4e-04 (2e-04)	4e-04* (2e-04)
$D_{i,t-1}^{\tau,d}$	0.031*** (0.0033)	0.0039 (0.0031)	0.03*** (0.003)	0.0141*** (0.0027)	0.0019** (7e-04)	-0.0063*** (6e-04)	-0.0047*** (5e-04)	-0.0057*** (5e-04)
$D_{j,t-1}^{\tau,d}$	0.0089* (0.0036)	-8e-04 (0.0032)	0.0105*** (0.003)	0.01*** (0.0028)	0.0174*** (9e-04)	0.0158*** (6e-04)	0.016*** (9e-04)	0.0144*** (6e-04)
$Spill(A)_{i,t-1}^{\mu,d}$	0.0104*** (0.0015)	0.008*** (0.0016)	0.0239*** (0.0014)	-0.0118*** (9e-04)	0.0018*** (3e-04)	0.0013*** (3e-04)	3e-04 (2e-04)	7e-04** (2e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	0.0069*** (0.0016)	-0.002 (0.0013)	-0.0163*** (0.0015)	-0.0081*** (0.0014)	-7e-04 (4e-04)	-0.0019*** (4e-04)	4e-04 (3e-04)	-3e-04 (2e-04)
$Spill(A)_{i,t-1}^{\tau,d}$	0.0195*** (0.0019)	0.0109*** (0.0017)	0.01*** (0.0018)	0.0023 (0.0017)	-0.0032*** (3e-04)	-0.0021*** (2e-04)	-0.0033*** (2e-04)	-0.0022*** (2e-04)
$Spill(A)_{j,t-1}^{\tau,d}$	0.0077*** (0.0021)	0.007*** (0.0019)	7e-04 (0.0018)	-3e-04 (0.0016)	-9e-04 (5e-04)	0.0016*** (3e-04)	4e-04 (4e-04)	0.0021*** (3e-04)
R^2	0.0346	0.0567	0.0329	0.0865	0.0246	0.2367	0.0379	0.2468
N	294815	324588	299715	319684	474725	474400	473999	475125
Sample avg.	0.3544	0.408	0.3631	0.3993	0.0513	0.0457	0.0501	0.0469
Subset avg.	0.0818	0.091	0.0817	0.0869	0.0264	0.023	0.0242	0.0219

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). Data: 4-digit unbalanced panel.

Table SI.9: Regression of link formation & breaking (unbalanced).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$Pr(\Delta w_{ij,t}^{\mu} \geq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \leq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \geq 0)$	$Pr(\Delta w_{ij,t}^{\mu} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \geq 0)$	$Pr(\Delta w_{ij,t}^{\tau} \leq 0)$
	upstream		downstream		upstream		downstream	
$1(\Delta w_{ij,t-1}^{\mu,d} = 0)$	-0.0177 (0.0099)	-0.0269*** (0.007)	-0.0033 (0.0095)	-0.0192** (0.0068)	0.0043 (0.0084)	-0.0119 (0.0079)	-0.0042 (0.0079)	0.0012 (0.0074)
$1(\Delta w_{ij,t-1}^{\mu,d} \geq 0)$	0.121*** (0.0106)	0.1083*** (0.0067)	0.0934*** (0.0102)	0.066*** (0.0062)	-0.0065 (0.0042)	-0.003 (0.004)	-0.0038 (0.0042)	-0.0037 (0.0042)
$1(\Delta w_{ij,t-1}^{\tau,d} = 0)$	-0.0166 (0.0117)	0.0029 (0.0097)	-0.0048 (0.0112)	-0.0139 (0.009)	0.0862*** (0.0091)	-0.1557*** (0.0107)	0.0768*** (0.009)	-0.1489*** (0.0102)
$1(\Delta w_{ij,t-1}^{\tau,d} \geq 0)$	-0.004 (0.0083)	0.001 (0.0067)	0.0034 (0.0081)	-0.0015 (0.0063)	0.1079*** (0.0099)	0.1337*** (0.0086)	0.1218*** (0.0104)	0.1422*** (0.0085)
$\sigma_{ij,t-1}^{\mu,d}$	-0.0075 (0.0157)	0.0738*** (0.0132)	-0.0196 (0.0175)	0.1269*** (0.0137)	0.0033 (0.0094)	-0.0121 (0.0085)	-0.0082 (0.008)	-0.0011 (0.0085)
$\Delta\sigma_{ij,t-1}^{\mu,d}$	-0.0083 (0.008)	0.0378*** (0.0071)	0.0076 (0.0071)	0.0443*** (0.0054)	4e-04 (0.005)	9e-04 (0.0047)	-0.003 (0.0035)	0.002 (0.0037)
$\sigma_{ij,t-1}^{\tau,d}$	0.0119 (0.007)	0.0019 (0.0062)	0.0103 (0.0083)	0.0014 (0.0068)	0.0131* (0.0051)	0 (0.0049)	-0.0045 (0.0055)	0.0036 (0.0052)
$\Delta\sigma_{ij,t-1}^{\tau,d}$	0.0065 (0.0043)	-0.002 (0.004)	-8e-04 (0.0052)	-0.0044 (0.0041)	0.0024 (0.0031)	0.0029 (0.0029)	-0.001 (0.0034)	0.0071* (0.0032)
$A_{i,t-1}^{\mu}$	0.0105 (0.0093)	0.0306*** (0.0075)	-9e-04 (0.0114)	0.1068*** (0.0089)	-0.0068 (0.0066)	-5e-04 (0.0057)	-0.0075 (0.0072)	-0.0019 (0.0068)
$A_{j,t-1}^{\mu}$	-0.0113 (0.0101)	0.0063 (0.0085)	-0.0139 (0.0079)	-0.0129 (0.0072)	0.0042 (0.0062)	-0.0018 (0.006)	-0.0063 (0.0059)	0.0053 (0.0061)
$A_{i,t-1}^{\tau}$	-0.0611*** (0.0177)	-0.0383** (0.0142)	0.0874*** (0.0178)	0.0249 (0.0133)	-0.0078 (0.0161)	0.0511*** (0.0144)	0.0813*** (0.0163)	0.0509*** (0.014)
$A_{j,t-1}^{\tau}$	0.0054 (0.0179)	-0.0157 (0.0138)	-0.0419** (0.0141)	-0.0059 (0.012)	-0.0109 (0.011)	-0.0251* (0.0109)	-0.0165 (0.0096)	-0.0148 (0.0109)
$D_{i,t-1}^{\mu,d}$	0.0215*** (0.0057)	-0.0034 (0.0049)	0.1609*** (0.0251)	-0.0812*** (0.0191)	-0.0036 (0.0041)	0.0039 (0.0039)	0.0195 (0.0159)	0.0409* (0.0167)
$D_{j,t-1}^{\mu,d}$	0.0121* (0.0053)	0.0016 (0.0043)	0.074** (0.0256)	0.1514*** (0.0221)	0.0054 (0.005)	-0.001 (0.0047)	0.0073 (0.0183)	-0.0302 (0.0181)
$D_{i,t-1}^{\tau,d}$	0.0183 (0.0105)	0.0256** (0.0086)	-0.0431*** (0.0093)	-9e-04 (0.0074)	0.0116 (0.0065)	-0.0062 (0.0053)	-0.0174*** (0.0038)	-0.0046 (0.0041)
$D_{j,t-1}^{\tau,d}$	0.0114 (0.012)	-0.0126 (0.0085)	5e-04 (0.01)	6e-04 (0.0083)	0.0206** (0.0069)	0.0391*** (0.0062)	0.0152* (0.0068)	0.0372*** (0.0051)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0229** (0.0081)	-0.0181** (0.0065)	0.0182* (0.0071)	0.0198*** (0.0056)	0.0053 (0.0052)	-0.0082 (0.0048)	0.0094* (0.0042)	7e-04 (0.0044)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.0035 (0.0083)	-0.0112 (0.0068)	0.0163* (0.0074)	-0.0248*** (0.0063)	-0.0061 (0.006)	-0.0087 (0.0054)	0.0011 (0.0045)	-0.0052 (0.0048)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.0164*** (0.0045)	-0.0144*** (0.0042)	0.0058 (0.0041)	0.0063 (0.0038)	-0.0094*** (0.0021)	-0.0048* (0.0021)	-0.002 (0.0018)	-0.0029 (0.0018)
$Spill(A)_{j,t-1}^{\tau,d}$	-0.0088 (0.0055)	-0.0061 (0.0045)	0.0036 (0.0042)	-0.0014 (0.0033)	-7e-04 (0.0041)	0.0057 (0.0032)	0.0047* (0.0023)	0.0016 (0.0021)
$IP_{i,t-1}$	0.0064* (0.0031)	-0.0036 (0.0019)	-0.0133*** (0.003)	-0.0041* (0.0019)	0.0027 (0.0022)	0.0023 (0.002)	-0.0063** (0.0022)	0.0023 (0.002)
$IP_{j,t-1}$	-0.0097* (0.0038)	0.0018 (0.0023)	0.0068* (0.0033)	-0.003 (0.0026)	-0.002 (0.0022)	0.0049* (0.0021)	6e-04 (0.0024)	0.0038 (0.0023)
$LWC_{i,t-1}$	-0.0053 (0.0045)	0.01** (0.0039)	0.0053 (0.0037)	-0.0118*** (0.0034)	-0.0028 (0.0039)	0.0026 (0.0038)	-0.0032 (0.0038)	-0.0019 (0.0033)
$LWC_{j,t-1}$	0.011** (0.0042)	-0.0031 (0.0039)	0.0026 (0.0046)	0.0039 (0.0037)	-0.0021 (0.0029)	-5e-04 (0.0031)	-0.0072 (0.0039)	-0.0041 (0.0033)
R^2	0.0352	0.0404	0.0309	0.0566	0.0204	0.112	0.0297	0.1125
N	17578	21370	17604	21344	22281	22349	22190	22437
Sample avg.	0.3895	0.4367	0.3989	0.4273	0.1857	0.168	0.1816	0.1721
Subset avg.	0.087	0.0966	0.0839	0.0887	0.0638	0.0575	0.0587	0.0542

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are logistic regressions including industry-pair and time FE and using two-ways clustered SE. It estimates the probability that up- and downstream links form or break. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). The sample average is the probability of link formation or breaking in the full sample, while the subset average is the frequency of significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$. Data: 4-digit balanced panel including trade.

Table SI.10: Regression of link formation & breaking (trade subset).

SI.2.3.3. Link formation at the intensive margin

Here, additional regression results on link formation processes at the intensive margin are provided. Table [SI.11](#) shows the results for the subset of manufacturing sectors at the 4-digit level of aggregation. Table [SI.12](#) shows the results for all 4-digit industries including those with missing data for some periods in time. Table [SI.13](#) shows the results including trade data which reduces the subset of industries.

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	-0.0146 (0.0819)	-0.0827 (0.0613)	-0.1883** (0.0664)	0.0304 (0.0866)	-0.0228 (0.0213)	0.0034 (0.0139)	0.0398. (0.0218)	0.007 (0.0158)
$\Delta w_{ij,t-1}^{\tau,d}$	0.0209 (0.0136)	-0.0072 (0.0082)	-0.0076 (0.0093)	-0.0062 (0.0132)	0.0799 (0.0532)	-0.1466*** (0.0265)	0.0986. (0.0555)	-0.2156*** (0.0283)
$\sigma_{ij,t-1}^{\mu,d}$	0.0047* (0.0024)	-0.0015 (0.0023)	-0.0017 (0.002)	-5e-04 (0.0015)	-0.001 (0.001)	0.0013 (8e-04)	0.0016* (8e-04)	-4e-04 (7e-04)
$\Delta \sigma_{ij,t-1}^{\mu,d}$	0.0014 (0.0011)	-0.0011 (7e-04)	-7e-04 (0.0011)	-0.0015** (5e-04)	-2e-04 (4e-04)	5e-04 (3e-04)	-2e-04 (4e-04)	-1e-04 (4e-04)
$\sigma_{ij,t-1}^{\tau,d}$	6e-04 (0.0014)	-7e-04 (8e-04)	0.0032* (0.0016)	-0.0019* (8e-04)	5e-04 (9e-04)	-0.001 (8e-04)	-0.0012 (0.0013)	-0.0037** (0.0013)
$\Delta \sigma_{ij,t-1}^{\tau,d}$	-4e-04 (8e-04)	-3e-04 (4e-04)	0.0013 (9e-04)	-6e-04 (5e-04)	6e-04 (6e-04)	-5e-04 (6e-04)	-2e-04 (8e-04)	-0.0023** (7e-04)
$A_{i,t-1}^{\mu}$	0.0047** (0.0015)	-0.002** (7e-04)	-0.0028** (0.001)	-0.0077*** (0.001)	1e-04 (9e-04)	7e-04 (8e-04)	0.0021 (0.0018)	-7e-04 (0.0014)
$A_{j,t-1}^{\mu}$	-0.0022* (0.0011)	0.0011 (9e-04)	-5e-04 (8e-04)	0.0031*** (9e-04)	0.001* (5e-04)	3e-04 (6e-04)	2e-04 (8e-04)	3e-04 (9e-04)
$A_{i,t-1}^{\tau}$	-0.0069* (0.0031)	0.0016 (0.0015)	0.0043. (0.0022)	-0.0018 (0.0016)	9e-04 (0.0019)	-0.003* (0.0014)	0.0066** (0.0025)	-0.0069*** (0.002)
$A_{j,t-1}^{\tau}$	0.0028 (0.0029)	-0.0042** (0.0016)	0.006*** (0.0016)	2e-04 (0.0016)	0.001 (0.001)	0.001 (0.0011)	-6e-04 (8e-04)	-4e-04 (6e-04)
$D_{i,t-1}^{\mu,d}$	-2e-04 (7e-04)	-8e-04 (5e-04)	0.0094* (0.004)	0.0083*** (0.0022)	6e-04 (4e-04)	-0.0011* (5e-04)	0.0034 (0.0026)	-0.0015 (0.0013)
$D_{j,t-1}^{\mu,d}$	-3e-04 (0.0014)	0.0095*** (0.0025)	-0.0049 (0.0042)	0.0035 (0.0027)	-0.0012** (4e-04)	-3e-04 (4e-04)	-0.001 (0.0024)	0.0025 (0.002)
$D_{i,t-1}^{\tau,d}$	0.0076*** (0.0016)	-0.001. (6e-04)	2e-04 (0.0014)	-0.0013. (8e-04)	4e-04 (8e-04)	-5e-04 (8e-04)	-0.0026** (9e-04)	-6e-04 (7e-04)
$D_{j,t-1}^{\tau,d}$	9e-04 (0.0016)	2e-04 (5e-04)	1e-04 (0.0011)	-4e-04 (8e-04)	-0.0118*** (0.0029)	0.0144*** (0.0028)	-0.0093* (0.0045)	0.0313*** (0.0067)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0024 (0.0019)	-1e-04 (0.001)	0.0019 (0.0014)	-3e-04 (0.001)	9e-04 (8e-04)	-4e-04 (8e-04)	4e-04 (6e-04)	-8e-04 (5e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	-0.0025 (0.002)	-5e-04 (0.0014)	-3e-04 (0.001)	8e-04 (7e-04)	-1e-04 (9e-04)	-0.0015 (0.0011)	6e-04 (0.001)	-3e-04 (8e-04)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.001 (0.001)	1e-04 (5e-04)	0.0023* (0.0011)	-0.0015* (7e-04)	-0.0015. (9e-04)	0.0025* (0.001)	4e-04 (0.001)	0.0021* (0.001)
$Spill(A)_{j,t-1}^{\tau,d}$	-8e-04 (0.0012)	-7e-04 (7e-04)	-0.0013 (9e-04)	6e-04 (5e-04)	-0.0016 (0.0012)	-0.0031* (0.0013)	0.0011 (0.0011)	2e-04 (0.0013)
R^2	0.0197	0.0317	0.0442	0.0277	0.0411	0.1832	0.0401	0.3285
N	8589	11591	8642	11538	8025	8159	7839	8350
Average*	0.0103	-0.0076	0.0095	-0.008	0.0103	-0.01	0.0118	-0.011
Subset avg.**	0.0318	-0.0276	0.0284	-0.0287	0.0379	-0.0385	0.0435	-0.0429

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are OLS regressions including industry-pair and time FE and using two-ways clustered SE estimating the magnitude of change conditional on $\Delta w_{ij,t}^{\alpha,d} \neq 0$. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta w_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.

Data: 4-digit balanced panel of manufacturing.

Table SI.11: Regression of link formation & breaking at the intensive margin (manufacturing).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	0.1758*** (0.0424)	-0.1987*** (0.0358)	0.0074 (0.0482)	-0.2199*** (0.0194)	-0.0236 (0.0145)	0.0011 (0.0319)	0.0027 (0.0318)	0.0018 (0.038)
$\Delta w_{ij,t-1}^{\tau,d}$	0.0027 (0.0022)	-9e-04 (0.0016)	-7e-04 (0.0022)	-0.001 (0.0014)	0.1455*** (0.0303)	-0.1881*** (0.017)	0.1226*** (0.0309)	-0.2218*** (0.0184)
$\sigma_{ij,t-1}^{\mu,d}$	3e-04 (2e-04)	-5e-04** (2e-04)	4e-04* (2e-04)	-0.0012*** (1e-04)	4e-04 (0.0011)	3e-04 (9e-04)	0.001 (8e-04)	-8e-04 (7e-04)
$\Delta \sigma_{ij,t-1}^{\mu,d}$	2e-04** (1e-04)	-4e-04*** (1e-04)	0 (1e-04)	-5e-04*** (0)	0 (4e-04)	1e-04 (4e-04)	-3e-04 (4e-04)	-1e-04 (4e-04)
$\sigma_{ij,t-1}^{\tau,d}$	1e-04 (1e-04)	-2e-04** (1e-04)	1e-04 (1e-04)	-2e-04*** (1e-04)	0 (4e-04)	-1e-04 (4e-04)	1e-04 (8e-04)	-0.0016* (6e-04)
$\Delta \sigma_{ij,t-1}^{\tau,d}$	0 (1e-04)	0 (0)	1e-04 (0)	-1e-04* (0)	2e-04 (3e-04)	0 (2e-04)	1e-04 (4e-04)	-8e-04** (3e-04)
$A_{i,t-1}^{\mu}$	0.0022*** (2e-04)	-0.0012*** (1e-04)	0.0015*** (1e-04)	-0.0035*** (1e-04)	-8e-04 (5e-04)	6e-04 (5e-04)	0.0012* (6e-04)	-3e-04 (5e-04)
$A_{j,t-1}^{\mu}$	-3e-04*** (1e-04)	1e-04 (1e-04)	-5e-04*** (1e-04)	3e-04*** (1e-04)	9e-04 (6e-04)	-9e-04 (5e-04)	-2e-04 (5e-04)	6e-04 (5e-04)
$A_{i,t-1}^{\tau}$	-0.0031*** (4e-04)	8e-04*** (2e-04)	-7e-04* (3e-04)	3e-04 (3e-04)	2e-04 (0.0011)	0.0016 (9e-04)	0.0093*** (0.0014)	-0.0043*** (0.0011)
$A_{j,t-1}^{\tau}$	2e-04 (4e-04)	-1e-04 (2e-04)	4e-04 (2e-04)	1e-04 (2e-04)	0.0026*** (7e-04)	-0.0035*** (6e-04)	-5e-04 (5e-04)	-0.0011** (4e-04)
$D_{i,t-1}^{\mu,d}$	-3e-04*** (1e-04)	1e-04** (0)	-7e-04*** (1e-04)	0.003*** (4e-04)	-1e-04 (5e-04)	-3e-04 (4e-04)	6e-04 (5e-04)	3e-04 (5e-04)
$D_{j,t-1}^{\mu,d}$	4e-04*** (1e-04)	0.0064*** (0.0013)	-2e-04 (1e-04)	0.0029*** (6e-04)	4e-04 (0.001)	0 (0.0011)	-0.0178 (0.0103)	0.0066* (0.0033)
$D_{i,t-1}^{\tau,d}$	0.0014*** (2e-04)	-5e-04*** (1e-04)	6e-04*** (1e-04)	-8e-04*** (1e-04)	-6e-04 (0.0014)	-0.0017 (0.0013)	-0.0017* (8e-04)	-3e-04 (7e-04)
$D_{j,t-1}^{\tau,d}$	-2e-04 (2e-04)	5e-04*** (1e-04)	-2e-04 (2e-04)	0 (1e-04)	-0.0365*** (0.0044)	0.072*** (0.0079)	-0.0354*** (0.0043)	0.0637*** (0.006)
$Spill(A)_{i,t-1}^{\mu,d}$	-2e-04 (1e-04)	1e-04 (2e-04)	6e-04*** (1e-04)	0 (1e-04)	3e-04 (0.001)	0.0011 (8e-04)	-0.0013* (6e-04)	-9e-04 (5e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	-4e-04** (1e-04)	7e-04** (2e-04)	-2e-04* (1e-04)	2e-04*** (0)	0 (9e-04)	0.0011 (7e-04)	0.0028 (0.0018)	-0.0015 (0.0015)
$Spill(A)_{i,t-1}^{\tau,d}$	-2e-04* (1e-04)	1e-04 (1e-04)	1e-04 (1e-04)	1e-04 (1e-04)	9e-04 (9e-04)	0.0025 (0.0013)	-3e-04 (0.0014)	0.0045* (0.0021)
$Spill(A)_{j,t-1}^{\tau,d}$	0 (1e-04)	1e-04 (1e-04)	-2e-04* (1e-04)	1e-04 (1e-04)	-0.0078** (0.0028)	0.0034 (0.0021)	-0.0044 (0.0042)	0.0064 (0.0042)
R^2	0.0222	0.0772	0.0069	0.1314	0.0806	0.2983	0.0785	0.3498
N	176272	206045	181176	201145	26460	26135	25735	26861
Average*	0.003	-0.0025	0.0028	-0.0027	0.014	-0.013	0.0154	-0.0135
Subset avg.**	0.0107	-0.0091	0.0102	-0.0099	0.0321	-0.0316	0.0365	-0.0347

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are OLS regressions including industry-pair and time FE and using two-ways clustered SE estimating the magnitude of change conditional on $\Delta_{ij,t}^{\alpha,d} \neq 0$. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.

Data: 4-digit unbalanced panel.

Table SI.12: Regression of link formation & breaking at the intensive margin (unbalanced).

	market layer				Patent-layer			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} > 0$	$\Delta w_{ij,t}^{\mu}$ if $\Delta w_{ij,t}^{\mu} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} > 0$	$\Delta w_{ij,t}^{\tau}$ if $\Delta w_{ij,t}^{\tau} < 0$
	upstream		downstream		upstream		downstream	
$\Delta w_{ij,t-1}^{\mu,d}$	0.1806** (0.0631)	-0.0758 (0.0878)	-0.2139* (0.1008)	-0.0071 (0.0596)	-0.0267 (0.0193)	0.0082 (0.0186)	0.0346 (0.0254)	0.0147 (0.016)
$\Delta w_{ij,t-1}^{\tau,d}$	0.0045 (0.0087)	-0.014* (0.0066)	-0.0026 (0.0066)	-0.0046 (0.0058)	0.0975* (0.0448)	-0.1554*** (0.0286)	0.1642* (0.071)	-0.2238*** (0.0312)
$\sigma_{ij,t-1}^{\mu,d}$	0.0019 (0.0019)	0.0015 (0.0021)	-0.0011 (0.0017)	-0.0024 (0.0018)	-0.0024 (0.0026)	-1e-04 (8e-04)	0.0014 (0.001)	2e-04 (9e-04)
$\Delta \sigma_{ij,t-1}^{\mu,d}$	-6e-04 (0.001)	-5e-04 (9e-04)	-6e-04 (0.0012)	-0.0019*** (4e-04)	-4e-04 (7e-04)	0 (4e-04)	5e-04 (4e-04)	1e-04 (4e-04)
$\sigma_{ij,t-1}^{\tau,d}$	0.0024* (0.0011)	-5e-04 (6e-04)	0.0028* (0.0012)	-0.0012* (5e-04)	-1e-04 (0.001)	0 (0.001)	-0.001 (0.0012)	-0.0012 (0.0012)
$\Delta \sigma_{ij,t-1}^{\tau,d}$	0.0013* (6e-04)	-4e-04 (4e-04)	0.0013 (7e-04)	-5e-04 (3e-04)	3e-04 (7e-04)	-1e-04 (6e-04)	0 (8e-04)	-7e-04 (7e-04)
$A_{i,t-1}^{\mu}$	0.001 (9e-04)	-0.0012* (5e-04)	-0.0023** (8e-04)	-0.0032*** (6e-04)	-1e-04 (8e-04)	0.0013 (7e-04)	0.0017 (0.0012)	3e-04 (9e-04)
$A_{j,t-1}^{\mu}$	-0.0027*** (8e-04)	0.0012 (8e-04)	5e-04 (8e-04)	0.0015** (5e-04)	4e-04 (5e-04)	7e-04 (6e-04)	-4e-04 (6e-04)	6e-04 (9e-04)
$A_{i,t-1}^{\tau}$	-0.0013 (0.0017)	6e-04 (0.001)	0.003 (0.0017)	-2e-04 (8e-04)	0.001 (0.0014)	6e-04 (0.0013)	0.007*** (0.0019)	-0.0028* (0.0014)
$A_{j,t-1}^{\tau}$	0.0014 (0.0019)	-0.0034*** (0.0011)	0.0028* (0.0013)	-3e-04 (9e-04)	9e-04 (0.001)	-0.0017 (0.0011)	-7e-04 (9e-04)	4e-04 (6e-04)
$D_{i,t-1}^{\mu,d}$	-8e-04 (8e-04)	-0.001 (8e-04)	0.0109** (0.0035)	0.0014 (0.002)	8e-04 (6e-04)	-0.0014* (6e-04)	0.0027 (0.0028)	-0.0014 (0.0017)
$D_{j,t-1}^{\mu,d}$	3e-04 (0.001)	-3e-04 (0.0025)	-0.0089* (0.0041)	0.0011 (0.0021)	-6e-04 (7e-04)	-6e-04 (5e-04)	0 (0.0034)	0.0047* (0.0022)
$D_{i,t-1}^{\tau,d}$	0.0059*** (0.0012)	-0.0018* (7e-04)	-0.0013 (9e-04)	7e-04 (5e-04)	6e-04 (9e-04)	-0.0014 (0.0013)	-0.004*** (0.0011)	1e-04 (0.001)
$D_{j,t-1}^{\tau,d}$	1e-04 (0.0014)	0 (6e-04)	0.0012 (0.001)	5e-04 (6e-04)	-0.0147*** (0.0037)	0.024*** (0.0044)	-0.0119* (0.005)	0.0287*** (0.0075)
$Spill(A)_{i,t-1}^{\mu,d}$	-0.0012 (0.001)	-4e-04 (9e-04)	0.0027* (0.0012)	0.0012 (6e-04)	1e-04 (8e-04)	-1e-04 (7e-04)	0.001 (7e-04)	-0.0011 (6e-04)
$Spill(A)_{j,t-1}^{\mu,d}$	-9e-04 (0.0012)	5e-04 (8e-04)	0 (9e-04)	4e-04 (5e-04)	0.0013 (0.0012)	-5e-04 (8e-04)	5e-04 (9e-04)	-0.0013 (9e-04)
$Spill(A)_{i,t-1}^{\tau,d}$	-0.0012 (7e-04)	4e-04 (3e-04)	0.0012 (8e-04)	-8e-04* (3e-04)	-6e-04 (0.0011)	0.0022 (0.0012)	-5e-04 (7e-04)	2e-04 (9e-04)
$Spill(A)_{j,t-1}^{\tau,d}$	-0.0019*** (7e-04)	2e-04 (3e-04)	-0.001 (5e-04)	2e-04 (3e-04)	-6e-04 (0.0017)	-8e-04 (0.0016)	6e-04 (0.0012)	-0.0015 (0.0011)
$IP_{i,t-1}$	0.0026*** (6e-04)	9e-04*** (2e-04)	-0.0012* (6e-04)	-3e-04 (2e-04)	-1e-04 (2e-04)	1e-04 (2e-04)	-5e-04 (2e-04)	1e-04 (2e-04)
$IP_{j,t-1}$	-2e-04 (7e-04)	0 (2e-04)	7e-04 (6e-04)	2e-04 (3e-04)	-2e-04 (6e-04)	-4e-04 (3e-04)	3e-04 (7e-04)	1e-04 (3e-04)
$LWC_{i,t-1}$	-0.0033*** (7e-04)	-1e-04 (3e-04)	-0.001* (4e-04)	0.0015*** (4e-04)	-1e-04 (5e-04)	3e-04 (7e-04)	-4e-04 (4e-04)	-2e-04 (3e-04)
$LWC_{j,t-1}$	6e-04 (5e-04)	0.001 (5e-04)	-5e-04 (8e-04)	-6e-04* (2e-04)	2e-04 (5e-04)	2e-04 (3e-04)	1e-04 (5e-04)	-3e-04 (3e-04)
R^2	0.0528	0.0167	0.0504	0.0247	0.0393	0.2217	0.0922	0.285
N	9384	13176	9410	13150	8405	8473	8317	8564
Average*	0.0068	-0.0053	0.0062	-0.0051	0.0092	-0.0088	0.0107	-0.0097
Subset avg.**	0.0183	-0.0158	0.0183	-0.0182	0.0336	-0.0336	0.0418	-0.0413

Significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

The models are OLS regressions including industry-pair and time FE and using two-ways clustered SE estimating the magnitude of change conditional on $\Delta w_{ij,t}^{\alpha,d} \neq 0$. Superscript indices $d = out, in$ of explanatory variables indicate that upstream links were used in column (1), (2), (5), (6) and downstream links in column (3), (4), (7), (8). *This is the average $\Delta w_{ij,t}^{\alpha,d}$ of the full sample but filtered by sign, i.e. < 0 or > 0 . **This is the average change in weights in the subset where significant changes in $w_{ij,t}^{\alpha,d}$ defined by $|\Delta w_{ij,t}^{\alpha,d}| > c_t^{\alpha,d}$ occurred.

Data: 4-digit balanced panel accounting for trade.

Table SI.13: Regression of link formation & breaking at the intensive margin (trade subset).