universität wien

MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master’s Thesis

„On the dynamic formation of production networks: An empirical assessment“

verfasst von / submitted by

Maximilian Reiner, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 953
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Research in Economics and Finance
Betreut von / Supervisor:	Michael Pfarrhofer, PhD

Acknowledgements

First and foremost, I would like to thank my supervisor, Michael Pfarrhofer, for his support and guidance throughout my Master's thesis. His suggestions and outstanding feedback were crucial for the success of my work and significantly contributed to shaping my interests, certainly influencing my future professional life.

In addition, I want to give my deepest appreciation to my fellow students for the valuable discussions we had during our studies. Our frequent exchanges helped me to expand my perception of topics related to my thesis, making it a more complete and comprehensive body of work.

Finally, but most importantly, special thanks to my friends and family for their support during my studies. Without their encouragement, I would not have been able to complete this journey. Although it was surely not easy being with me for some time, they showed understanding and, beyond that, kept me motivated during particularly stressful periods.

Abstract

The production network is the structure on which microeconomic shocks propagate through firms or industries. It defines the input-ouput relations between economic agents and constitutes direct and indirect interdependencies. Recent theories link its structure to the macroeconomic effects of microeconomic shocks. Yet, the factors determining production networks are empirically less covered. Building upon the idea that the production network is the outcome of dependent individual decisions, the present paper proposes a novel approach toward identifying structural patterns in input-ouput relations. The econometric framework, in form of a stochastic actor-orientated model, captures the adoption of input linkages with the neighborhood of agents and, thereby, endogenously describes the production network. From another perspective, the production network directs the diffusion of production technologies through successive input choices. The used model class overcomes the limitations of conventional regression analyses related to path dependencies by simulating the latent network evolution between observations.

Using time-series data of the US production network, the analysis reveals that existing indirect horizontal and vertical relations significantly affect the formation of linkages. Technological complementarity and vertical integration are essential to explain the heterogeneity of industries regarding their importance as suppliers or for the indirect connectivity of others. In general, industries with strong high-level interdependencies are more likely to form relations. These results show that the evolution of the production network is partly endogenous and questions theories about the propagation of shocks assuming static network structures.

Kurzfassung

Das Produktionsnetzwerk stellt jene Struktur dar, über welche sich mikroökonomische Schocks von Unternehmen zu Unternehmen bzw. zwischen Branchen ausbreiten. Es definiert die Input-Output-Beziehungen zwischen Wirtschaftsakteuren und stellt direkte und indirekte Interdependenzen dar. In neueren Theorien wird seine Struktur mit den makroökonomischen Auswirkungen mikroökonomischer Schocks in Verbindung gebracht. Nichtsdestotrotz sind Faktoren, die Produktionsnetzwerke bestimmen, empirisch kaum untersucht. Ausgehend von der Idee, dass Produktionsnetzwerke das Ergebnis abhängiger individueller Entscheidungen sind, findet in dieser Arbeit ein neuartiger Ansatz zur Identifizierung struktureller Muster in Input-Output-Beziehungen Anwendung. Der ökonometrische Rahmen in Form eines stochastischen, akteursorientierten Modells assoziiert die Entstehung von Input-Beziehungen mit der Nachbarschaft von Akteuren und beschreibt damit das Produktionsnetzwerk durch sich selbst bzw. endogen. Aus einem anderen Blickwinkel betrachtet, beeinflusst das Produktionsnetzwerk die Diffusion von Produktionstechnologien durch aufeinanderfolgende Input-Entscheidungen. Die herangezogene Modellklasse überwindet die Einschränkungen herkömmlicher Regressionsanalysen im Zusammenhang mit Pfadabhängigkeiten, indem sie die latente Entwicklung des Netzwerks zwischen den Beobachtungen simuliert.

Unter Verwendung von Zeitreihendaten des US-Produktionsnetzwerks zeigt die Analyse, dass bestehende indirekte horizontale und vertikale Beziehungen die Bildung von Verflechtungen zwischen Branchen erheblich beeinflussen. Technologische Komplementarität und vertikale Integration sind wesentlich, um die Heterogenität von Branchen hinsichtlich ihrer Bedeutung als Zulieferer bzw. für die indirekte Konnektivität anderer zu erklären. Im Allgemeinen gehen Branchen mit starken indirekten Interdependenzen direkte Beziehungen mit größerer Wahrscheinlichkeit ein. Diese Ergebnisse zeigen, dass die Entwicklung des Produktionsnetzwerks teilweise endogen zu erklären ist und stellen damit Theorien über die Ausbreitung von Schocks, die von statischen Netzwerkstrukturen ausgehen, in Frage.

Contents

List of Tables

List of Figures

1. Introduction

Macroeconomic theory neglects the role of microeconomic shocks as explanatory factors of business cycle fluctuations for the most part. Dating back to the 'other Lucas critique', shocks at the level of firms and sectors were viewed as random disturbances averaging out in aggregate. Based on Lucas (1977), research focused primarily on macroeconomic sources for aggregate fluctuations like monetary policy. This view traces sectoral dynamics back to outside factors generating the observed comovement between producers externally. Only recently have disaggregated shocks received more attention as driving forces of aggregate macroeconomic dynamics (e.g., Acemoglu et al. 2012; Carvalho 2014; Gabaix 2011; Acemoglu et al. 2016).

In fact, recent events like the energy supply shock in Europe as a consequence of the Russian-Ukraine War highlight the importance of sectoral shocks for the economy as a whole. In this case, the relevance of cascading effects played a prominent role in the discussion about the driving factors of inflation. Generally, linkages between industries or firms form a network structure, namely the supply chain or production network. Thus, they build the foundation on which theories related to propagation mechanisms of firm-specific shocks rest. Moreover, there is some suggestive evidence that the explicit consideration of input-output relations between firms can bring the interplay of the micro and macro level within and between economies closer to observed patterns.[1]

In addition, disentangling aggregate and sectoral economic movement is challenging from an empirical perspective due to the inherent (time) dependence constituted by the production network. Consider again the energy supply shock mentioned above. Clearly, industries, directly and indirectly, demand energy – directly for their own production and indirectly via intermediary goods, which also require energy usage. However, the observable impact on a specific industry is the sum of the direct and indirect cost shock and, thus, not decomposable without considering the production network. Since most empirical and theoretical investigations treat the relationships defining the network as exogenous (see, e.g., Foerster et al. 2011; Barrot and Sauvagnat 2016), sectoral dynamics cannot be unequivocally attributed to the observational unit but may instead originate in network dynamics.

1. Among others, Carvalho (2014) provides some suggestive evidence that the structure of and the comovement within the production network relates to aggregate output dynamics.

Furthermore, sectoral and aggregate shocks as well as the emergence of new technologies can lead to the formation or revision of individual input-output relations due to changes in economic circumstances (including the production network itself) and via the described propagation mechanism to major adaptions of the whole network structure. Overall, such cascading link adaptions refer to network endogeneity. The term suggests that the network evolution is partly explained by the network's history and, accordingly, by past structures. For example, Carvalho and Voigtländer (2014) point out that firms appear to consider the relative supply chain positions of potential suppliers in their decision about input adoptions.

The widely debated role of production networks in macroeconomics together with the caveats in the present works, motivate research about the determinants of network formation within the scope of business cycle dynamics (see Acemoglu et al. 2012; Acemoglu et al. 2016) and endogenous growth theory (e.g., Bresnahan and Trajtenberg 1995). In fact, limited knowledge of network dynamics restricts the informativeness of analyses dealing with micro-macro relationships partly attributed to network effects. Understanding the network as the structural outcome of individual decisions conditional on the network state itself is an essential step for the literature on micro-based macroeconometric models. However, papers dealing with the endogeneity of production networks are primarily theoretical and generally scarce.

This paper fills this gap by proposing a novel econometric approach to uncover inter-dependencies of input-output relations and their effect on the evolution of production networks. More precisely, it borrows from the literature on social networks (e.g., friendship networks) and discusses how the idea behind the dynamic network model of Snijders (2001) reflects the described view of production networks. The application to the US production network at the sector level confirms the presence of network endogeneity. Thereby, estimates suggest that the embedment of sectors in the network affects their input-output relations underlining the dynamic character of network structures. Finally, this strengthens the relevance of the interplay between structural dynamics and short and long-run economic developments for empirical examinations.

1.1. Methodological Contributions

In the context of a model of strategic network formation, the subsequent examination aims to inspect and empirically explain the structure of the US production network at the sector level. Such models explicitly address the endogenous evolution of supply chains via structural network effects. These effects describe the formation or dissolution of input-output relations as a function of the current network structure. In analogy to social networks, an example of a structural effect is the 'friends-of-friends' effect. It refers to a commonly found pattern in social networks, namely that friendships often develop between people who share friends.[2] Among the different types of explanatory factors, the emphasis is on structural effects and the method used for their identification. Despite differences in approaches, implications for the network structure are contrastable with those obtained from conventional non-network models.

In order to assess the endogeneity of production networks, repeated observations of the network, as used here, are advantageous: Longitudinal network data contains the necessary time and cross-sectional variation to directly model network evolution. By conditioning on its initial state, it is no longer necessary to explain the entire structure. In particular, the production network resulted from many historical individual input-output decisions. These ministeps in the evolution of the network, in the wording of Snijders (2005), are not independent if, as assumed, industries take the existing network structure into account. For example, linkages may act as information channels or direct the search for new input goods. Yet, a single observation of the production network does not contain information about path dependencies between ministeps and thus requires researchers to model the entire network formation process. Thus, the analysis of network data often boils down to descriptive evidence. When utilizing longitudinal network data, however, the structural complexity of networks reduces, which eventually supports the identification of structural effects driving network dynamics.

Since repeated observations of networks are generally rare, models dealing with such data are less common in the literature. Furthermore, the path dependency of ministeps, as described above, requires, in general, the simulation of the latent network evolution between observations. A class of models having these features are the so-called stochastic actor-oriented (SOA) models for network dynamics introduced by Snijders (2001).[3] This framework explains the network evolution

2. For example, various studies examine the effect of social networks on labor market outcomes. Most prominently, Granovetter (1973) provides evidence for the 'strength of weak ties'. He describes that even loose (weak) relationships possess information about available jobs by giving access to further parts of the network. Irrespective of the strength of ties, indirect connectivity expands the amount of information and thereby supports the employment of workers or the formation of ties.

3. SOA models are also known as `SIENA` (Simulation Investigation for Empirical Network Analysis) models – named after the homonymous software suite implemented in `R`.

from the perspective of the network nodes (actors) themselves, in the present case industries, and hence through individual decisions. Compared to temporal exponential random graph (TERG) models first presented by Hanneke et al. (2010), SOA models explicitly specify the change of linkages as a latent process of time-dependent ministeps. Hence, SOA models focus on the micro perspective to describe the overall network structure. Leifeld and Cranmer (2019) provide a more detailed comparison of these prominent classes of temporal network models.

Unlike other approaches, modeling the evolution of the production network as dependent ministeps explicitly builds on endogenous variation. Hence, omitted variable bias is of little concern compared to non-network models. For example, consider the emergence of new technologies; the associated industries integrate into the production network, meaning that their introductions are accompanied by input choices of themselves. Due to the emerging character of new industries, these adoptions are arguably independent of the network structure and, thereby, second-best. Through technological spillover, such network alterations in terms of an expanding set of technologies, arising exogenously or endogenously, can entail changes of input-output relations. When using conventional models, cascading effects imply a bias because the individuals' outcomes (linkages) are not independent and thus vary across industries and time. Especially in the case of endogenous technological progress like recombinational innovations, namely novel combinations of existing inputs described by Weitzman (1998), propagation implies biased estimates.

Using SOA models to characterize production networks is novel. This model class is particularly suitable to test theoretical implications concerning the formation of input-output relations because its specification closely resembles the chain of reasoning, namely technological progress, delineated by theories. In this sense, structural effects directly relate to the documented aspects of propagation mechanisms and their implications for the supply chain (see Section 2).

1.2. Findings

In the context of SOA models and in terms of structural tendencies, the analysis focuses on two main aspects of tie formation. The data basis comprises the input-output benchmark tables between 1997 and 2012 of the US economy. After harmonization, the data represents a time-series of the intersectoral linkages between manufacturing industries.

First, the role of hub industries in the network evolution is subject to discussion. Earlier studies, like Carvalho (2014), show strong heterogeneity of input-output linkages across industries in the US production network. Moreover, most research (e.g., Oberfield 2018; Acemoglu et al. 2012) concludes that these asymmetries of suppliers are necessary for the transmission of microeconomic shocks to the macroeconomy. Yet, the presented empirical approach can only indirectly capture the formation of hubs, not least because it takes the initial network structure as given but because it describes the input relations of industries through their neighborhood. Directly, it addresses how already existing hubs shape the adaption of relations. In this regard, the characterization of industries as hubs, i.e., those supplying intermediates to relatively many others, is not significantly associated with any systematics in tie formations of and to themselves. In particular, there is no evidence that hubs per se are disproportionately attractive as suppliers. This finding undermines the hypothesis that the heterogeneity of industries with respect to their network position fosters network concentration or sparsity. Against the theory of Oberfield (2018), the data does not support the assumption that hubs are more efficient than others as measured by total factor productivity (TFP) or that particularly efficient industries are preferred suppliers. However, due to their central network position, hubs are crucial for the indirect connectivity of industries.

Second, the paper examines how existing higher-order (indirect) linkages affect the creation or termination of relations. For their representation, the production network provides the structure to construct a distance measure between industries. In particular, the associated metrics consider (unweighted) input relations and thus reflect a rudimentary measure of technological distance. For example, semiconductors are closer to electronic devices regarding their production technology than to handcrafted products. Relative to random goods, the usage of comparable goods may be more likely. The associated structural effects distinguish between horizontal and vertical similarity. As in Conley and Dupor (2003), the former captures the similarity of the production technologies and compares the intermediates of industry pairs. The latter refers to indirect linkages further up the supply chain. It follows Carvalho (2014) and builds upon the idea that industries may tend to adopt inputs already used by their suppliers. From a different perspective, the effect of the vertical distance measure identifies upstream search or tendencies toward vertical integration. Note that these statistics vary over industry pairs. In SOA models, individual agents incorporate the network structure in an ego-centric way, whereby these distance measures basically describe different forms of transitive relationships.

The results point towards the relevance of both effects for the formation of ties. Specifically, industries are more likely to adopt an input-output relation when industries are either close in terms of vertical or horizontal distance. Accordingly, the examination emphasizes their relevance for the aggregate level, namely the production network. While Carvalho (2014) recognizes the former, the latter tendency has not been documented yet. Nevertheless, vertical search explains the variation of individual input relations to a considerably larger extent than horizontal similarity. The average standardized effect ratio for individual linkages comparing ex-post tied and untied industry pairs is around 1:1.928, implying that upstream search is almost twice as important for assessing differentials in linkages than indirect horizontal relations.

However, the analysis shows that the US production network features structural time heterogeneity. The magnitude of structural effects substantially differs across periods, also indicated by descriptives. Despite this, the tendencies mentioned above are robust regarding their significance.

In conclusion, the US production network exhibits significant path dependency. Although by no means the only source affecting input-output relations, its evolution is partly described through endogeneity by associating past structures to the linkage formation at the micro level. Estimates clearly show the relevance of the neighborhood of industries for explaining their input relations.

1.3. Organization

The rest of the thesis is structured as follows. Section 2 gives an overview of the related literature. Section 3 describes how production networks relate to graphs as the formal counterpart of network data, followed by a detailed specification of SOA models and structural network effects. Section 4 documents facts about the structure of the US production network, which are subsequently incorporated in the estimated SOA models in Section 5.

2. Related Literature

2.1. Microeconomic Shocks and Aggregate Fluctuations

Long and Plosser (1983) recognized the role of microeconomic idiosyncratic shocks for aggregate fluctuations. A major part of the literature treats supply chains as exogenous and conditions on the network when assessing the propagation of idiosyncratic shocks. Due to a lack of granular panel network data, however, empirical studies on the relation of production networks and aggregate fluctuations are challenging.

Therefore, Foerster et al. (2011), Gabaix (2011), Atalay (2017) and others calibrate multi-industry models to provide evidence for the importance of sectoral shocks for business cycle fluctuations. They construct supply chains by representing the input-output relations of firms as a network. Moreover, Atalay (2017) observes that the production technology of US industries exhibits complementary inputs. Within their model, these complementarities generate dependency between micro- and macroeconomic dynamics.

Closely related, Acemoglu et al. (2012), Pesaran and Yang (2020) and Carvalho (2014) document stylized facts about the production network at the sector level as given by US input-output data and suggestively link its structure to aggregate dynamics. For example, they confirm the existence of dominant (hub) sectors, which supply many other sectors. Using a theoretical model, Acemoglu et al. (2012) assert that these supply-side asymmetries are crucial for establishing a causal link between sectoral shocks and aggregate fluctuations. In particular, hubs are highly interconnected and, thereby, possess the potential to directly and, by propagation, indirectly affect large parts of the economy.

Most research identifies asymmetric linkage formation as crucial for microeconomic network effects on aggregate dynamics. Yet, the subsequent analysis also connects to research on international economic comovement. Giovanni and Levchenko (2010), for example, views trade between firms of different countries as a factor contributing to comovement. In particular, trading connects national production networks and makes international spillover effects on a disaggregated level possible. In this sense, trade mediates microeconomic shocks (e.g., Dhyne et al. 2020), whereby this literature strand faces similar limitations to that elaborated above.

Moreover, Boehm et al. (2019), Carvalho et al. (2020) and Barrot and Sauvagnat (2016) exploit exogenous variation in industrial production induced by natural disasters in Japan and the US and thereby focus on how these shocks propagate through the production network and ultimately affect aggregate variables.

2.2. Network Formation

The literature on economic network formation mainly examines the diffusion of information between individual economic agents and technologies embedded in networks and its effects on the network structure (Conley and Udry 2010; Banerjee et al. 2013). Nonetheless, it constitutes the foundation for more complex models in which production networks are viewed as graphs and the sum of individual input decisions. From this point of view, the connection between research on the endogeneity of supply chains and social interactions within networks is evident.[4] More specifically, this paper connects the literature on strategic network formation in the context of social networks with those examining production networks. Related models all build upon the assumption that individuals adapt relationships to maximize welfare and are used, for example, to analyze outcomes driven by networks as is the case for labor market outcomes (see, e.g., Calvó-Armengol and Jackson 2004).

In fact, most work on the evolution of production networks identifies technological progress as the driving factor behind changing input-output relations. Technological progress, however, can be of various forms. For instance, Weitzman (1998) broadly describes innovations as the recombination of ideas and, in a narrow sense, of input goods. Guided by this definition, any change in the supply chain, either concerning linkages or firms, represents technological progress.

In that regard, Acemoglu and Azar (2020) examine the adoption of the production network at the level of firms to new technologies. According to their theory, single innovations amplify economic growth through substitution effects expressed in cascading network changes. In particular, with the arrival of new products, the input choice set of firms expands such that a different combination of inputs may lead to higher efficiency. These gains pass over to the consuming firms via price reductions and trigger the revision of their input linkages. Additionally, each decision imposes externalities on other firms through indirect relations given by the network. This mechanism, also described by Gualdi and Mandel (2018), aligns with the

4. See Jackson (2011) for a general introduction to social networks in the context of economics.

findings of Oberfield (2018), whose theoretical model describes the emergence of star suppliers caused by productivity differences. Note that, when holding the set of firms and thereby technological possibilities constant, the recombination of inputs is the only source of network variation. Finally, insights of Oberfield (2018) explain supply-side asymmetries outlined in Acemoglu et al. (2012) through network endogeneity and further underpin the importance of network disruptions for the economy. Regarding the theory proposed by Elliott et al. (2022), firms employ distinct inputs and strengthen relations to protect against the risk of firm failure. Thereby, their decisions result in network heterogeneity or asymmetric linkages.

Related to production networks and in contrast to Oberfield (2018), Carvalho and Voigtländer (2014) draw from the friendship model of Jackson and Rogers (2007) to explain the search of firms for new inputs along existing linkages. In particular, they depart from a fixed effects model and address an omitted variable bias by conditioning on the initial state of the production network. Nonetheless, their empirical approach potentially induces a bias because regression models disregard path dependencies of individual input decisions by construction (see Section 1 and 3.2 for a more detailed discussion). Still, by incorporating the initial network proximity between sectors, Carvalho and Voigtländer (2014) most closely resembles the idea behind structural network effects, which guide the present discussion.

3. Econometric Framework

3.1. Production Networks as Graphs

As pointed out by Acemoglu et al. (2012) and in the terminology of graph theory, the production network is a directed graph (or digraph) $\mathcal{G}$ described by a set of arcs (ties) comprised in the adjacency matrix $\boldsymbol{D} = (d_{ji})_{1 \leqslant i,j \leqslant N}$ and a set of nodes $\mathcal{N}$ with $|\mathcal{N}| = N$, i.e., $\mathcal{G} = (\boldsymbol{D}; \mathcal{N})$. More specifically, each tie d_{ji} represents a (directed) flow of goods produced by the upstream (preceding) economic agent $j \in \{1, \ldots, N\}$ to the input consuming one $i \in \{1, \ldots, N\}$. Consequently, the set of economic agents (e.g., firms) and their linkages entirely describe the production network.

Subsequently, it is assumed that $d_{ji} \in \{0, 1\}$ and $d_{ii} = 0$ (no self-relations) for all $i, j \in \{1, \ldots, N\}$. Hence, the analysis does not consider the intensity of relations and accordingly (valued) flows of goods. In other words, the continuous input choice of an economic agent collapses to a binary decision per feasible tie. The discussion, therefore, primarily refers to the informational character of ties as in Carvalho (2014) or rather to information about input-ouput relations between agents governed by the network. Technically, the imposed constraint shrinks the space of networks compared to the general case of weighted ties ($d_{ji} \in \mathbb{R}$) and, thereby, transforms the infinite network space into a finite one. Nevertheless, the number of feasible directed graphs is huge, i.e., $2^{N(N-1)}$, and increases disproportionally with N.

Figure 1a depicts a stylized example of a graph and, ultimately, of a production network as discussed here. Arrows indicate directed flows, while circles represent agents. In principle, reciprocal (bidirectional) relationships ($d_{jh} = d_{hj} = 1$) are possible, although generally not consistent with the view of production networks being vertically aligned. However, when considering levels of aggregation above the firm level such ties may result from pooling well-distinct goods and firms into categories.

For future reference, define the in- and outdegree of agent i as follows:

$$\deg_i^{in} := \sum_{j=1}^{N} d_{ji} = \boldsymbol{d}'_{\bullet i}\boldsymbol{\iota}_N, \qquad \deg_i^{out} := \sum_{j=1}^{N} d_{ij} = \boldsymbol{d}_{i\bullet}\boldsymbol{\iota}_N$$

with $\boldsymbol{\iota}_N$ being a $N \times 1$ vector of ones and $\boldsymbol{d}_{i\bullet}$ denoting the ith row of $\boldsymbol{D}$. In words, the indegree of i is equal to the number of (distinct) input goods used by agent i, whereas the outdegree of i is equal to the number of output relations. Hence, $\boldsymbol{d}_{i\bullet}$ summarizes the supply-side (outgoing ties) and $\boldsymbol{d}_{\bullet i}$ the demand-side (incoming ties) of agent i.

Consider the network shown in Figure 1a. Here, agent i directly uses the products of agents $\{j, k\}$ s.t. $\deg_j^{in} = |\{j, k\}| = 2$, whereas agent h supplies only agent j s.t. $\deg_h^{out} = 1$.

Extending the graph structure by a discrete-time dimension is straightforward. Let t denote time and assume a constant set of nodes, i.e., $\mathcal{N} = \mathcal{N}_t$ for all $t \in \{0, \ldots, T-1\}$. Moreover, summarize node characteristics in matrices $\boldsymbol{X}_t$ of dimension $N \times K$. Then, the augmented time-series graph is given by $\{\mathcal{G}_t\}_{t=0}^{T-1}$ with $\mathcal{G}_t = (\boldsymbol{D}_t; \mathcal{N}, \boldsymbol{X}_t)$.

3.2. Model Specification

The model specification below follows Snijders (2001) and defines a SOA model. However, explanations focus on the application to the production network as outlined above and the concrete implications of the model assumptions. Snijders (2001, 2005) present further details with respect to formal definitions and the statistical methodology.

In general, the model explains the dynamic evolution of networks from the nodes' perspective, i.e., network changes are due to the individual decisions of actors embedded in a network. Through a combination of endogenous (structural effects) and exogenous components, the underlying actor-driven stochastic process aims to reproduce the network structure by identifying regularities in tie changes. For simplicity, the subsequent definitions of the model ingredients disregard node-specific covariates.

From the graph panel data structure described in Section 3.1, the observed network states $\boldsymbol{D}_t$, $t \in \{0, \ldots, T-1\}$, are discrete. Yet, viewing network dynamics as the aggregate of sequential individual adaptions, so-called ministeps, gives rise to a continuous-time setting.

More precisely, at any time point $\tau \in [0, T-1]$, a specific agent i, subsequently referred to as the focal actor or ego, gets the opportunity to change one of its incoming ties $\boldsymbol{d}_{\bullet i,\tau}$. Agents associated with $\boldsymbol{d}_{\bullet i,\tau}$, namely the counterparts considered by agent i, are commonly referred to as alters. Thus, the resulting network $\boldsymbol{D}_{\dot{\tau}}$ at time $\dot{\tau} = \tau + d\tau$ differs at most by one relation $d_{ji,\dot{\tau}} = 1 - d_{ji,\tau}$ for some j. Figure 1 illustrates such a ministep. Note that only the first ministep in the first observed interval $[0, 1]$, $\tau = 0$, takes place conditional on $\boldsymbol{D}_0$. Observations at $t > 0$ are solely used for estimation. The framing of tie changes as decisions only emphasizes the agent-based character of the model, namely that actors actively send out ties.

Concerning production networks, industries are thought of as making input choices. Conceptionally, this relates to markets through which economic agents interact. They enter into negotiations and send nonrejectable offers (ties) representing input choices. As a consequence, industries cannot affect their supply-side or outgoing ties, i.e., they do not decide on the industries using their product as input. It is reasonable to assume that input factors are more important for the development of industries than their consumers' characteristics because inputs determine production technologies. In fact, input substitutions based on short-term considerations are not part of the analysis. Instead, it deals with sector level data and long-term trends. Hence, the model captures major technological adaptions. In the sense of recombinant growth, the choice of inputs is crucial for the innovation output of industries and hence their economic growth.

Note that this assumption does not restrict technological diffusion to downstream spillover effects because the direction of good flows does not limit the direction of information flows; rather, it defines that industries choose production technology. In other words, industries consider all up- and downstream industries as input suppliers, whereby the network represents their relative technological position. Yet, Bellamy et al. (2014), Hauknes and Knell (2009) and others attribute the value of supply chains for technological change to their informational character. In this respect, Forni and Paba (2003) examine the direction of technological spillover and, contrary to intuition, provide evidence for upstream spillover effects. As typically assumed in the literature on the network origins of economic growth (see, e.g., Acemoglu and Azar 2020; Saviotti and Pyka 2008), the set of production technologies comprises all possible combinations of goods, i.e., the choice set of industries is unrestricted.

As the network dynamically evolves during simulation, it is unclear which tie agent i will modify a priori; and accordingly results as the optimal decision. In particular,

$$\boldsymbol{D}_{\dot{\tau}} = \arg\max_{\boldsymbol{D}' \in \mathcal{D}} u_i(\boldsymbol{D}', \boldsymbol{D}_\tau; \boldsymbol{\beta})$$

whereby $u_i(\cdot)$ is the so-called objective function, $\boldsymbol{\beta}$ a vector of parameters and $\mathcal{D}$ denotes the set of feasible networks relative to $\boldsymbol{D}_\tau$. Hence, $\mathcal{D}$ comprises networks that may result from a ministep of agent i conditional on the previous state fully known by all actors. This specification implicitly assumes myopic agents since they do not consider future network changes caused by their decisions. Moreover, $u_i(\cdot)$ permits (path) dependencies and cascading effects through $\mathcal{D}$. In its simplest form, the objective function is the same for the creation and dissolution of linkages and is equal to the gain in terms of the evaluation function $f_i(\cdot)$:

$$u_i(\boldsymbol{D'}, \boldsymbol{D}_\tau; \boldsymbol{\beta}) = f_i(\boldsymbol{D'}; \boldsymbol{\beta}) - f_i(\boldsymbol{D}_\tau; \boldsymbol{\beta}) + \epsilon_i$$

with ϵ_i being an i.i.d. random variable following a standard Gumbel distribution.[5] This error term captures the non-systematic component of the agent's preferences. Put simply, $f_i(\cdot)$ is a linear combination of K endogenous and exogenous effects $s_{ik}(\boldsymbol{D})$ weighted by coefficients $\beta_1, \ldots, \beta_K$. A more detailed discussion of these effects and the evaluation function is provided below. Analogously to the random utility model of McFadden (1974), these assumptions on $u_i(\cdot)$ or ϵ_i imply

$$\mathrm{Pr}_i(\boldsymbol{D}_{\dot{\tau}} | \boldsymbol{\beta}; \boldsymbol{D}_\tau) = \frac{\exp(u_i(\boldsymbol{D}_{\dot{\tau}}, \boldsymbol{D}_\tau; \boldsymbol{\beta}))}{\sum_{\boldsymbol{D'} \in \mathcal{D}} \exp(u_i(\boldsymbol{D'}, \boldsymbol{D}_\tau; \boldsymbol{\beta}))} \tag{1}$$

comparable to the form of a multinomial logistic regression model. Intuitively, the probability that agent i changes the network from $\boldsymbol{D}_\tau$ to $\boldsymbol{D}_{\dot{\tau}}$ exponentially increases with the gain associated with its decision.

Although there is a methodological nexus to random utility theory, $u_i(\cdot)$ cannot be interpreted as a utility function. Technically, the objective function solely measures the relative fit of network adjustments given the information provided by the observed and augmented (simulated) data.

Besides the evaluation function, the rate functions $r_{it}(\cdot)$, $t \in \{1, \ldots, T-1\}$, are an integral part of the model. Their sum over agents defines the parameter of an exponential distribution and therefore affects the time between two arbitrary link adaptions per observed time interval. Indirectly, they specify the probabilities with which agents get the opportunity to decide on a tie and therefore determine the speed of the network evolution. Similar to the evaluation function, these probabilities are either constant or vary between agents through dependencies on the network state and covariates. In its simplest form and subsequently assumed, $r_{it}(\cdot) = \theta_t$ implying that $\boldsymbol{\theta}$ is a $T - 1 \times 1$ vector of rate parameters, one for each period, which are independent of the network structure.

5. Because of its favorable properties, this specification constitutes a widely used discrete choice model, namely the multinomial logit model. In particular, consider n options with utility $v_i + \eta_i$ and assume a standard Gumbel distribution for η_i. Then option j is chosen with probability

$$\frac{\exp(v_j)}{\exp(v_1) + \ldots + \exp(v_n)}.$$

See Guadagni and Little (1983) for further details.

As a direct consequence of viewing time as a continuous variable, the model has to simulate the latent network evolution between observations. Essentially, it approximates the true dynamics via a continuous-time Markov process. Informally, the probability that agent i adapts $d_{ji,\dot{\tau}} = 1 - d_{ji,\tau}$, i.e., the transition probability of the continuous-time Markov process, is composed of two events: (1) agent i gets the opportunity to decide on a tie (rate function) and (2) alters $d_{ji,\tau}$ (evaluation function). Hence, the Markov process combines the main model ingredients and thereby fully determines $\boldsymbol{D}$.

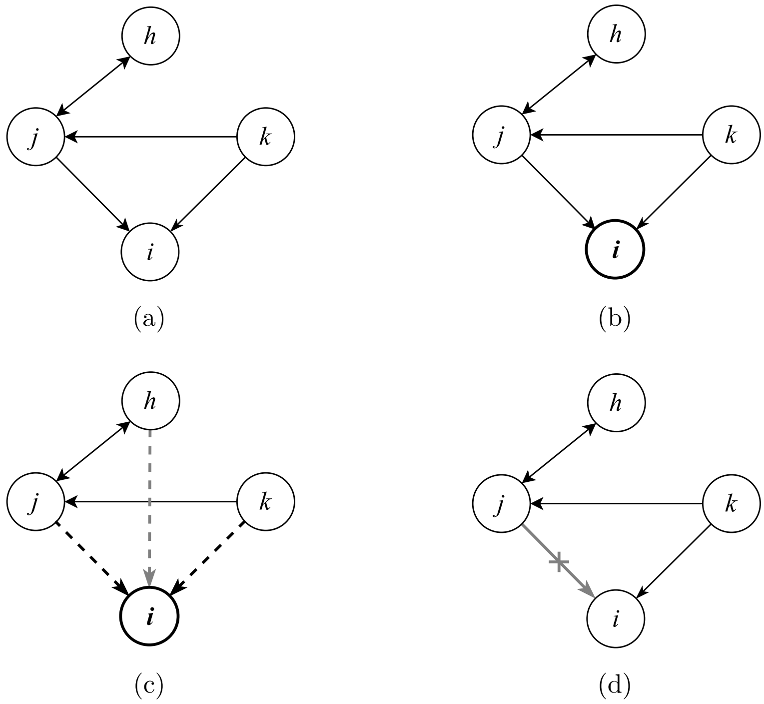


Figure 1.: Schematic network evolution

3.2.1. Endogenous and Exogenous Effects

As discussed, actors base their decision on different types of explanatory factors. These comprise ego and alter covariates and interactions thereof, as well as structural effects featuring the informational character of the network state. The former correspond to node-specific time-(in)variant covariates $\boldsymbol{x}_{it}$, and are, in their nature, equivalent to confounders used in conventional econometric models, while the latter capture the network state $\boldsymbol{D}$ represented by the set of ties. Formally, for each agent i the evaluation function is given by

$$f_i(\boldsymbol{D}; \boldsymbol{\beta}) = \sum_{k=1}^{K} \beta_k s_{ik}(\boldsymbol{D}).$$

From Eq. (1) follows that the parameters $\boldsymbol{\beta}$ are, in interpretation, identical to those entering logistic regressions, namely as (log-)odds. In particular,

$$\frac{\Pr_i(d_{ji}=1)}{\Pr_i(d_{ji}=0)} = \exp\left(\sum_{k=1}^{K} \beta_k s_{ik}(\boldsymbol{D})\right). \tag{2}$$

However, to have a clear idea of how effects enter the model, consider the spread of industrial robots as an illustrative example behind network dynamics. As a tool of automatization, they increased the productivity of manufacturers and paved the way for the development of key technologies like microelectronics. However, the dissemination process of new technologies is gradual and consists of many individual input decisions. In particular, each firm takes multiple factors into account. These relate to their characteristics (ego covariates) and, importantly, to that of alters. Naturally, the gains of automatization may differ for firms producing, for example, electronic components or handcrafted goods, respectively.

Moreover, the supply chain provides information about evolving technologies and spatial proximity of potential input goods and, ultimately, serves as the mean of technological diffusion. In what follows, relevant endogenous effects related to production networks are discussed.

Consider the communication technology behind computers. With the shrinkage of semiconductors driven by industrial robots, it diffused widely. Although the production of many electronic devices has required semiconductors before, computers significantly increased the connectivity of most. For example, televisions and domestic appliances have become ‘smart’. On the contrary, industries producing less similar goods, like bicycles and handcrafted goods, have either never adopted computers or, if so, with an extensive time lag. This pattern illustrated in Figure 2b refers to structural equivalence with respect to inputs (SE), namely technological similarity, and in a broader sense, horizontal search. In Figure 2b, the focal actor i (bold) evaluates the creation of a tie to actor j, i.e., $d_{ji,\tau} = 0 \rightarrow 1$ (dashed). In terms of structural equivalence, they have n input suppliers, $k_1, \ldots, k_n$, in common. Conley and Dupor (2003) use a conceptually similar metric to quantify the closeness of industries and establish a relationship between productivity comovement and technological similarity.

Formally, the effect definition is

$$s_i^{\mathrm{SE}}(\boldsymbol{D}) = \sum_{j=1}^{N} d_{ji} \sum_{\substack{k=1 \\ k\neq i,j}}^{N} \left(b_0 - |d_{ki} - d_{kj}|\right)$$

where b_0 is a constant. Let $d_{ji,\tau} = 0 \to 1$ define the transition of state $\boldsymbol{D}_\tau$ to $\boldsymbol{D}_{\dot{\tau}}$ whereby $d_{hi,\tau} = d_{hi,\dot{\tau}}$ for all $h \neq j$ and $d_{ji,\dot{\tau}} = 1 - d_{ji,\tau} = 1$. In addition, let actors i and j differ at exactly $|\mathcal{A}|$ incoming ties, i.e., $d_{ki,\tau} = d_{kj,\tau}$ for all $k \notin \mathcal{A}$ and $d_{ki,\tau} \neq d_{kj,\tau}$ for all $k \in \mathcal{A}$, s.t.

$$\Delta s_i^{\mathrm{SE}}(d_{ji,\tau} = 0 \to 1) = s_i^{\mathrm{SE}}(\boldsymbol{D}_{\dot{\tau}}) - s_i^{\mathrm{SE}}(\boldsymbol{D}_\tau) = -|\mathcal{A}|.$$

Ceteris paribus, namely $\Delta s_{ik}(d_{ji,\tau} = 0 \to 1) = 0$ for all structural effects except the structural equivalence effect, the expected gain $\mathbb{E}\, u_i(\cdot)$ of this transition is

$$f_i(\boldsymbol{D}_{\dot{\tau}}; \boldsymbol{\beta}) - f_i(\boldsymbol{D}_\tau; \boldsymbol{\beta}) = \beta_{\mathrm{SE}} \Delta s_i^{\mathrm{SE}}(d_{ji,\tau} = 0 \to 1) = -\beta_{\mathrm{SE}}|\mathcal{A}|.$$

Accordingly, β_{SE} is the marginal expected loss associated with a tie creation to an actor differing at one incoming tie ($|\mathcal{A}| = 1$).

However, Carvalho and Voigtländer (2014) describe a more intuitive form of search, namely vertical search. It expresses screening along existing linkages. According to their hypothesis, firms direct their search for new inputs upstream. Under the assumption of search costs, firms utilize the information provided by the network and identify potential suppliers as those producing inputs for their upstream neighbors. Hence, existing demand linkages establish an indirect connection to new suppliers via mediator firms. This line of argument resembles the process of technological dissemination as the adoption of enabling technology. For example, some medical devices rely on industrial robots and microelectronic components. However, producing the latter was made possible by advancements in automation. Arguably, the miniaturization of medical products was facilitated by more precise assembly robots – not only associated with the emergence of microelectronics but also with the production of medical products itself.

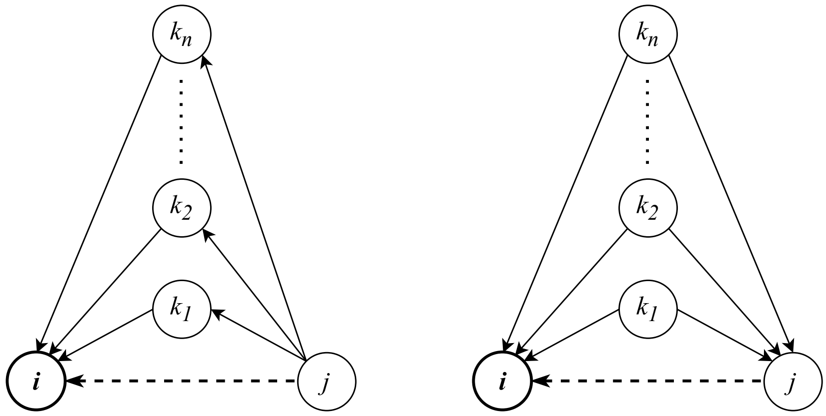
As Figure 2a depicts, vertical search represents a tendency to transitive closure (TC). The associated structural effect is given in its extensive form by

$$s_i^{\mathrm{int.TC}}(\boldsymbol{D}) = \sum_{j=1}^{N} d_{ji} \max_k (d_{ki} d_{jk})$$

and as intensive specification

$$s_i^{\mathrm{ext.TC}}(\boldsymbol{D}) = d_{ji} \sum_k d_{ki} d_{jk}.$$

Note that $s_i^{\mathrm{ext.TC}}(\boldsymbol{D}_{\dot{\tau}})\,|_{d_{ji,\dot{\tau}}=1} = \Delta s_i^{\mathrm{ext.TC}}(d_{ji,\tau} = 0 \to 1)$.



(a) Transitive closure (b) Structural equivalence

Figure 2.: Structural effects

Finally, the theoretical model of Oberfield (2018) demonstrates the endogenous emergence of star suppliers under certain assumptions. Empirically, outdegree popularity (OP) is indicative of such suppliers because it expresses the tendency for firms to adopt general-purpose technologies like automation.[6] The network structure helps firms to identify these technologies by providing information about the degree of product usage, namely product popularity. It is defined as

$$s_i^{\mathrm{OP}}(\boldsymbol{D}) = \sum_{j=1}^{N} d_{ji} \sum_{k=1}^{N} d_{jk}.$$

Appendix A.3 provides further effect specifications.[7]

3.2.2. Estimation

Originally, Snijders (2001) proposed to estimate $\boldsymbol{\beta}$ via Method of Moments (MoM).[8] In principle, the expected values of the population statistics $Z_{tk} = \sum_{i=1}^{N} s_{ik}(\boldsymbol{D}_{\tau_t})$, namely z_{tk}, summed over all $T-1$ time intervals and their sample equivalents $\hat{z}_{tk}$ form a system of K equations. Its solution defines the moment estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$. Since $\hat{\boldsymbol{\beta}}$ cannot be calculated explicitly, Snijders (2001) uses stochastic approximation methods for estimation. Note that estimation requires relating simulated network states with observations. In this sense, let the simulation of ministeps during a period

6. For a discussion on the role of general-purpose technologies for technological progress and growth see Bresnahan and Trajtenberg (1995).

7. For a comprehensive list of various exogenous and endogenous effects see the manual of the `RSiena` package (Ripley et al. 2023).

8. Other applicable estimation procedures comprise Maximum Likelihood estimation (Snijders et al. 2010b) and Bayesian inference (Koskinen and Snijders 2007).

t start at τ_{t-1} and end at τ_t such that $\boldsymbol{D}_{\tau_t}$ denotes the network state associated with the observation at time t. Note that, expect for $t = 0$, this does not imply $\boldsymbol{D}_t = \boldsymbol{D}_{\tau_t}$.

Following the MoM, for each structural effect, a statistic Z_{tk} must be chosen such that the expected values of Z_{tk} are sensitive with respect to changes in the associated parameter β_k. In other words, $\boldsymbol{\beta}$ characterizes $z_{tk} = \mathbb{E}_{\boldsymbol{\beta}}[Z_{tk}|\boldsymbol{D}_{\tau_{t-1}}]$, whereby the statistics Z_{tk} incorporate the average or aggregate structural effect over actors. Intuitively, larger values of β_k represent a tendency for all actors to form ties to maximize their value of the structural effect s_{ik}. Hence, Z_{tk} is expected to increase with β_k, which illustrates the relevance of the statistics for the parameters. Assuming time-invariant parameters and utilizing all sample periods, Snijders (2001) defines the MoM estimator as the solution to the equations

$$\sum_{t=1}^{T-1} \mathbb{E}_{\boldsymbol{\beta}}[Z_{tk}|\boldsymbol{D}_{\tau_{t-1}}] = \sum_{t=1}^{T-1} \hat{z}_{tk}, \quad k = 1, \ldots, K. \tag{3}$$

Note that this discussion leaves out the estimation of the $T-1$ rate parameters summarized in $\boldsymbol{\theta}$. However, the construction of their estimator follows a similar logic.

Since the population moment z_{tk} is subject to the simulation of the network evolution described above, it depends on $\boldsymbol{\beta}$. Formally, this feature is made implicit by the conditional expectation in Eq. (3). As a consequence, the estimator is not available in closed-form. Estimation has to be carried out conditional on a (previous) estimate $\hat{\boldsymbol{\beta}}_n$.

Snijders (2001) propose a stochastic approximation procedure in the spirit of Robbins and Monro (1951) to recover properties of the population moments z_{tk} and solve the MoM equation system. The so-called Robbins–Monro algorithm is an iterative method to solve root-finding problems dealing with expected values like that described by Eq. (3).

In step $n+1$, it draws a random sample $z_{tk,n}$ from the distribution of Z_{tk} conditional on $\hat{\boldsymbol{\beta}}_n$ to approximate the left side of Eq. (3). Together with the outcome of the previous step $\hat{\boldsymbol{\beta}}_n$ and the observed statistics z_{tk}, the algorithm uses these draws to construct $\hat{\boldsymbol{\beta}}_{n+1}$. In its simplest form, the iteration step is given by

$$\hat{\boldsymbol{\beta}}_{n+1} = \hat{\boldsymbol{\beta}}_n - a_n \sum_{t=1}^{T-1} \left(\begin{bmatrix} z_{t1,n} \\ z_{t2,n} \\ \vdots \\ z_{tK,n} \end{bmatrix} - \begin{bmatrix} \hat{z}_{t1} \\ \hat{z}_{t2} \\ \vdots \\ \hat{z}_{tK} \end{bmatrix} \right)$$

with a_n being a sequence of positive step sizes. Under certain conditions outlined by Blum (1954), $\hat{\boldsymbol{\beta}}_n$ convergences to the solution of Eq. (3), $\hat{\boldsymbol{\beta}}$, and coincides with the estimates obtained via the MoM.

In simple terms, the estimation algorithm updates the parameter vector conditional on its previous draw and the fit of a simulated network time-series relative to the observed one. On the other hand, the continuous-time Markov process outlined above simulates the latent network evolution conditional on the coefficients.

4. Data and Descriptive Analysis

4.1. Data Sources

The underlying network data stems from the US Bureau of Economic Analysis (BEA) and consists of the make and use tables. The combination of those tables is the empirical equivalent of an adjacency matrix $\boldsymbol{D}$. In particular, they define the US production network's structure at the industry level. Make tables are matrices showing the producer value of product transactions by commodity and the producing industry within a specific year; use tables describe the usage of intermediates by industry and commodity. In their most granular versions, more commonly known as the benchmark input-output (IO) tables, the BEA publishes them every five years. However, due to various redefinitions and the introduction of the North American Industry Classification System (NAICS) and its revisions, these tables do not form consistent longitudinal network data, $\mathcal{N} = \mathcal{N}_t$ for all $t \in \{0, \ldots, T-1\}$, and therefore require harmonization. Appendix A.1 presents further details concerning the data preparation process. Note that the industry classification system used to construct the IO tables is not the NAICS.

In addition, the NBER-CES Manufacturing Industry Database provides data about manufacturing industries based on the NAICS. This data set includes yearly observations of industry characteristics like total factor productivity (TFP) or capital expenditures. Due to different industry classification systems, all covariates are aggregated and calculated using the assignment tables provided by the BEA. Since the network and industry data are not of the same sampling frequency, additional transformations ensure the alignment of the time-series. If not stated otherwise, industry-specific covariates are 5-year means; IO tables are on an end-of-period basis.

All results base on the IO benchmark tables for 1997, 2002, 2007 and 2012 and the induced subgraph over the manufacturing industries, namely sectors associated with NAICS codes starting with 33, for which additional (time-varying) covariates are available.

4.2. Descriptive Analysis

The inclusion of structural effects in SOA models is either justified by theoretical models and arguments or descriptive evidence. While the former are literature-based, the latter result from a descriptive network analysis. In this regard, statistics summarize specific aspects of the network structure and, together with industry characteristics, help to uncover tendencies of IO relations. The time-variation of these statistics provides insights about patterns in linkage adaptions and, appropriately, the systematic evolution of the US production network.

Although the elaborated econometric model is industry-centric, statistics related to the entire network structure in Table 1 illustrate the evolution of the production network. First, a trend towards sparsity of $\boldsymbol{D}$ (low density) is evident, implying that the complexity of industries in terms of the number of distinct input goods decreases. However, this interpretation requires precaution because of unobservable substitution effects. In particular, the data does not contain information about within-industry relations. Little changes in self-relations $\text{diag}(\boldsymbol{D})$ mitigate these concerns.

Second, the unweighted mean distance, measured as the average number of edges along the shortest paths between network nodes, is nearly constant over time. The same holds for the longest distance in the network, namely the diameter. The difference between them is small, such that the longest supply chain comprises roughly 2.5 times as many stages of intermediary production as the average. Based on the mean distance being approximately proportional to $\log(N)$, the production network is, concerning the connectivity of industries, considered 'small'. In addition, the global clustering coefficient of the production network is, on average, roughly 50%. Accordingly, the observed networks contain almost half of all possible (undirected) industry triangles. The subgraph induced by industries $\{i, j, k\}$ in Figure 1a constitutes an example of such an industry triangle or transitive relation.

These two properties together characterize what is commonly called a small-world network. Specifically, while most industries do not directly supply to many others, they are part of their supply chain. Hence, the connectivity of industries is primarily indirect and given by the supply of inputs for intermediates. The discussion below points to several other properties of the US production network often associated with small-world networks.

Compared to their unweighted equivalence, the weighted mean distance and diameter are subject to high time variation. Although the econometric network model cannot handle the formation of weighted ties, these weighted statistics provide suggestive evidence for the network evolution regarding intensities. However, weighted statistics resemble short and long-term developments. Since weights are good flows measured in real producer values, weights vary, for example, with the business cycle and represent short-term economic developments. On the other hand, long-term structural changes capture input decisions related to technological adaptions targeted by unweighted statistics. This implies that differences between the weighted and unweighted connectivity measures are for the most part explainable by the variation in real output across industries.

Either way, Table 1 shows the weighted mean distance between industries, according to which the network followed a strong tendency towards concentration until 2007 and divergence afterward. Yet, the observed trend reversal, not pronounced in the unweighted mean distance, may be attributed to asymmetric shocks within the manufacturing sector during the Great Recession.

Conceptually, the embedment of industries in the network gives information about the flows of goods. In the broadest sense, industries interact through markets by buying inputs from and selling produced goods to each other. Thereby, IO relations are the observed market outcomes. This view motivates the distinction between the industries' demand-side and supply-side represented by the set of in- and outgoing ties.

4.2.1. Supply-side

As argued, the econometric model estimates effects related to the formation of input linkages. Nevertheless, this does not restrict the ego in its evaluation of potential input suppliers. Structural effects summarizing the supply-side of alters are explicitly necessary to capture properties of small-world networks.

In this regard, Table 1 indicates that industries are of different importance for the connectivity of the production network. The existence of central industries suggests that a few industries act as mediators by reducing the distance between industries. Formally, the greatest distance from these nodes to others is minimal; intuitively, they are mediators. These hub-like industries indirectly connect network nodes by producing widely used goods and by definition build upon general-purpose (key) technologies. From 2002-2012, mediators make up approximately 11% of all industries included in the sample.

The empirical cumulative distribution function (eCDF) of outdegrees shown in Figure 3 underpins the view of mediators as hubs. It resembles the CDF of the Pareto distribution, a power-law probability distribution. However, the distribution

Table 1.: Summary statistics

		1997	2002	2007	2012
# Industries		204	204	204	204
# IO relations		13690	8805	7639	7571
Tie turnover		-	+1022; -5907	+95; -1261	+55; -123
Panel A *Density*					
Density		0.331	0.213	0.184	0.183
Transitivity		0.674	0.531	0.494	0.492
Reciprocity		0.399	0.284	0.258	0.259
Panel B *Connectivity*					
Mean distance	unweighted	1.800	2.086	2.192	2.184
	weighted	0.072	0.045	0.018	0.027
Diameter	unweighted	5	5	6	5
	weighted	1.959	2.076	0.499	1.044
# Centers		12	23	23	23

Note: The table shows yearly summary statistics of the harmonized US production network. These comprise the number of industries N, their relations $\sum_i \sum_{j \neq i} d_{ji}$, and the number of formed and terminated relations relative to the previous year.
Panel A presents density statistics at the graph level. The density of a graph refers to the ratio of the number of realized linkages and the number of all possible linkages, $(\sum_i \sum_{j \neq i} d_{ji})/[N(N-1)]$. The transitivity statistic or clustering coefficient is defined as the ratio of closed industry triangles and the maximal count thereof. Note that this statistic does not consider tie directions. Reciprocity is the share of mutual links, $\sum_i \sum_{j \neq i} d_{ji} d_{ij}$, relative to the number of connected industry pairs.
Panel B reports statistics related to the connectivity of industries. They build upon a distance measure between industry pairs. In its unweighted version, the length of a (directed) path coincides with the number of industries on it. The average distance of shortest paths or geodesics, $\delta(i,j)$, between all 2-tuples of industries is called the mean distance of a graph. The greatest geodesic among all pairs, $\max_{i,j} \delta(i,j)$, defines the diameter. Similarly, the centers of a graph are industries for which the greatest geodesic is minimal, $\{i | \max_j \delta(i,j) = \min_i \max_j \delta(i,j)\}$. Here, output relations characterize centers, i.e., the relevant distance measure considers the shortest path from central industries to others (downstream geodesic).

of outdegrees is positively skewed and indicates asymmetries regarding the centrality of industries. While most industries do not directly supply to many others, a relatively small proportion, referred to as hubs, does. Note that outdegrees are not indicative of node centrality in terms of mediation but for hubs. The reason for this is that centrality formally pertains to high-order or indirect relations, whereas the outdegree distribution reflects direct linkages.

The presence of mediators and hubs in production networks is a well-established fact frequently incorporated in theoretical models. For example, Acemoglu et al. (2012) draw conclusions about the rate at which aggregate volatility induced by sectoral shocks decays, conditional on the parameterization of the outdegree distribution. According to them, the persistence of aggregate fluctuations increases with the role industries have as direct and indirect input suppliers.

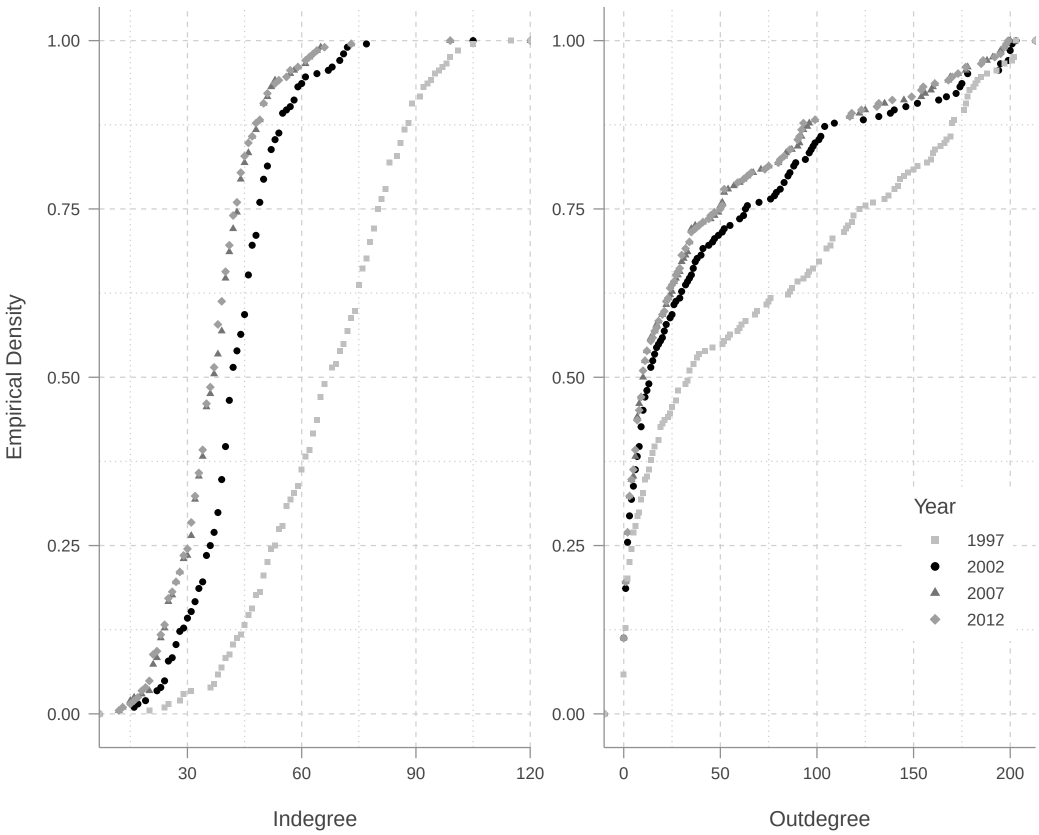


Figure 3.: Empirical distribution of in- and outdegrees

4.2.2. Demand-side

Cross-sectional variation in indegrees reflects, inter alia, differences in the production technology and product characteristics. Along these lines, the indegree distribution with its eCDF depicted in Figure 3 represents the heterogeneity of industries. Compared to the outdegree distribution, it is less skewed.

To identify structural effects related to industry complexity, a closer look at the characteristics and linkages of neighboring industries is instructive. The network position of second-order reveals details along the supply chain and, therefore, closely resembles the notion of structural effects. For this purpose, Figure 4 plots the indegree of industries against the average in- and outdegree of upstream-neighbors.

Among other things, the (qualitative) input choice of industries and their direct suppliers do not show any systematic relationship.[9] Industries of different complexity in terms of indegree do not differ with respect to the complexity of upstream industries. This indicates that input decisions are uncorrelated with the technology of direct suppliers, further supported by the assortativity coefficients of the indegree in Table A1 being close to zero. Generally, measures of assortativity quantify the homophyly of networks. It, therefore, captures the tendency of nodes to tie with others similar to themselves. As done here, assortativity can be expressed by correlation coefficients. Although not examined for higher orders, there is no evident reason to assume any structural effects involving industries further down the supply chain.

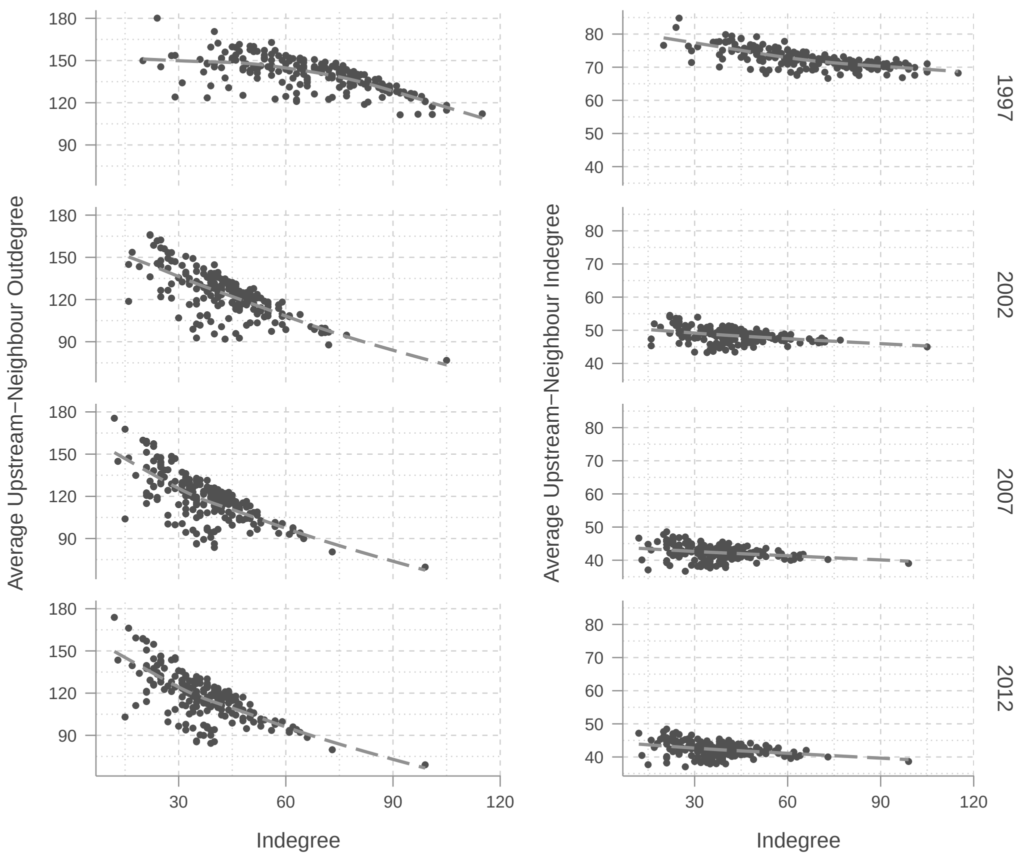


Figure 4.: Second-order network positions and indegrees

Moreover, Figure 4 suggests a negative correlation between industry complexity and the outdegree of direct suppliers. In particular, industries using fewer inputs demand a disproportionately high share of them from hubs. The negative upstream

9. Similar to outdegree popularity, the corresponding structural effect within SOA models, referred to as indegree popularity, describes a dependency between the indegree of alter and ego.

assortativity of in- and outdegrees points to the different importance hubs have as direct suppliers. Note that this fact does not contradict the mediating function of hubs defined through indirect connectivity. Yet, averaging over neighbors can drive this asymmetry due to the small number of hub industries. The fact that a similar relationship is observable when consulting the weighted average of outdegrees reduces this concern.

At the same time, hub industries are not distinguishable from others regarding complexity (see Figure A1). Hence, the input decision of central suppliers seems to be unaffected by their role as mediators.

Especially for network data, descriptives are not appropriate to reveal structural effects. The endogeneity of the production network combined with its granularity makes it particularly difficult to uncover systematic tendencies in input decisions and provide a reasonable explanation for the underlying mechanisms. The estimates in Section 5 seek to do precisely that.

5. The Empirics of Endogenous Network Formation

The discussions in Section 3.2 and 4.2 provide a set of hypotheses about the formation of production networks. Within the context of SOA models, they conform to specific structural effects and, therefore, can be tested through conventional statistical methods. Without loss of validity, the following interprets estimates as if obtained from multinomial logit regressions, namely as (log-)odds, by applying the steps delineated in Section 3.2.1 in an exemplary fashion. For the comprehensibility of this procedure and the interested reader, Table A2 gives definitions of all structural effects relevant here.

Before going into effects incorporating alter characteristics, consider solely ego-related effects first. They are comparable to explanatory variables of regression models. Such effects relate to ego characteristics but also include effects inferred from the network structure as the indegree density effect ($\deg_i^{in}$). In some sense, they serve as intercepts and capture tendencies in input choices for which the input-producing industry (alter) does not feature any IO relations or characteristics making it particularly attractive as a supplier. Broadly speaking, specific ties are not explainable by multilateral structural effects but through the different natures of industries. Nevertheless, this does not imply a perfect model fit in terms of the indegree distribution or any other metric. For example, the negative coefficients of the indegree density effect reported in Table 2 highlight the high sparsity of the US production network. Ceteris paribus, industries value additional incoming ties negatively and tend to have low complexity.

Besides this, the reciprocity effect is another basic effect representing mutual linkages. As pointed out above, within production networks, its interpretation is ambiguous. In fact, the estimated reciprocity effect is less pronounced than in social networks, which reflects the typical vertical structure of supply chains.

All other structural effects in Table 2 relate in some way to the alter. They are of particular interest because of their informative value concerning the systematic formation of networks beyond that obtainable from standard regression analyses.

The descriptives in Section 4.2 delineate substantial differences in the production network over the years. Although the observed tie variation is necessary for estimation, it is doubtful whether the production network indeed exhibits those changes or the harmonization of the IO tables induces them. This is especially true for the period 1997-2002. Therefore, estimates have to be, in any case, contrasted with those resulting from smaller sampling periods (see Table A4) or incorporating heterogeneous time effects (see column 3 in Table 2). Yet, the period 2007-2012 is unconcerned by this issue because statistics of both years are based on the identical industry classification system. In the context of SOA models, the estimated rate parameters in Panel B of Table 2 underline concerns regarding data consistency. They point out that the frequency of tie changes are multiples higher for the first and second period.

In addition to the reported effects, the main model specification comprises additional structural effects and time-varying covariates. Their inclusion is either motivated by descriptive statistics, like the in-outdegree assortativity effect, or concerns regarding the goodness of fit. Following Snijders et al. (2010a), the inclusion of effects follows a data-driven approach starting from a parsimonious model. Besides the outdegree distribution and the triad census, the most relevant target statistic is the indegree distribution.[10] These auxiliary statistics reflect the model fit when compared with their observed values and use one-step ahead network simulations. Determined by the input decisions of industries, the indegree distribution directly shows the network (or aggregate) implications of the micro level.

5.1. Degree-related Effects and Hubs

As concluded here and documented by others, the US production network contains hub and mediator industries. Utilizing the concept of structural effects, SOA models provide the possibility to examine their endogenous evolution. Note that this part of the analysis does not uncover the mechanism behind their emergence but rather points to the idiosyncrasies of existing hubs for other industries. For a discussion of effects inducing heterogeneity in IO relations see the subsequent sections.

First, Oberfield (2018) describes heterogeneity across entrepreneurs as the driving force behind star suppliers. In particular, he argues that productivity differences lead to market concentration in intermediates and, thereby, to the formation of hubs. If this mechanism applies, industries with high TFP attract ties. In other words, the TFP of the alter enters the consideration of industries positively.

10. The triad census counts the number of triads per type in a graph. Each triad type consists of a set of three nodes and a distinct set of six directed ties at most. For digraphs, all subsets of ties together define 16 types.

Table 2.: SOA model results: Baseline specifications

$\log \frac{\Pr(d_{ji}=1)}{\Pr(d_{ji}=0)}$	(1)	(2)	(3)
Panel A *Evaluation function*			
Indegree density	-2.986*** (0.116)	-2.654*** (0.125)	-2.162*** (0.133)
Reciprocity	0.343*** (0.045)	0.327*** (0.047)	0.306*** (0.046)
Transitive ties	0.951*** (0.084)	0.913*** (0.100)	0.671*** (0.123)
Transitive closure	0.033*** (0.002)	0.028*** (0.003)	-0.006 (0.011)
Structural equivalence	0.003* (0.002)	0.010*** (0.002)	0.034*** (0.005)
$\mathbb{1}$(1997-2002) × Transitive closure			0.030 (0.019)
$\mathbb{1}$(1997-2002) × Transitive ties			0.436 (0.311)
$\mathbb{1}$(1997-2002) × Structural equivalence			-0.022** (0.009)
Outdegree popularity	0.005*** (0.001)	0.003*** (0.001)	0.002* (0.001)
TFP alter		-1.044*** (0.153)	-1.137*** (0.154)
Covariates	No	Yes	Yes
Panel B *Rate function*			
1997-2002	58.034 (1.123)	54.704 (0.973)	55.067 (1.235)
2002-2007	7.201 (0.209)	7.197 (0.205)	6.844 (0.192)
2007-2012	0.731 (0.061)	0.733 (0.060)	0.728 (0.061)

Note: The table reports the estimated coefficients $\hat{\boldsymbol{\beta}}$ in Panel A for different SOA model specifications. Panel B shows estimated rate parameters $\hat{\boldsymbol{\theta}}$ and indicates the observation periods used for estimation. Coefficients are marginal effects on the log-odds of forming a tie. Industry covariates comprise TFP, total real value added, total employment, and real capital stock as ego, alter, and similarity effects. Other included structural effects are the 3-cycles, outdegree activity, and out-indegree assortativity effect.
Standard errors are given in parentheses. Based on two-sided t-tests, stars indicate the significance of estimates, i.e., $^{***}p < 0.01$, $^{**}p < 0.05$ and $^{*}p < 0.1$. Overall maximum convergence ratios are equal to 0.142, 0.213 and 0.155, respectively. Values below 0.25 indicate good convergence of the stochastic approximation method of Snijders (2001) with respect to the MoM equations (see Ripley et al. 2023). Models were estimated with the `RSiena` package.

Since the sign of the parameter of the corresponding structural effect in Tables 2 and A4 flips significantly between specifications, results are not robust and do not permit any general conclusions in that regard. Baseline specifications show a significant and opposing effect, whereas estimates of restricted samples tend to be positive but less precise. Particularly striking is that the effect turns significantly positive when excluding the first period and when considering only the last period, namely 2007-2012, becomes negative again. This suggests that the production network features systematic differences between periods with respect to the attractiveness of TFP and the period 2002-2007. The most robust estimates regarding the data (see Table A4), namely the period 2007-2012, do not provide evidence in favor of the theory that the productivity of upstream suppliers guides input choices either; rather hubs seem to show no systematic deviation from others with respect to TFP (see Figure A2). Either way, the relevance for explaining tie heterogeneity is relatively moderate because the differences of TFP across industries are low with an interquartile range of 0.077. Little variation in TFP weakens the effect when comparing alter industries and adversely affects its identification. Nevertheless, it cannot be ruled out that the described mechanism prevails at more disaggregated levels or longer horizons.

Second, consider the outdegree popularity effect already mentioned. The effect itself measures the number of industries supplied by the alter industry. Equally, it indicates how frequently a specific good is used as an intermediate and directly defines to what extent the alter is a hub. Note that outdegree popularity is an intensive measure to characterize the relevance of hubs for input choices. During estimation, the algorithm evaluates the effect for all possible tie adaptions and compares potential suppliers (alters). Hence, it expresses relative outdegrees. Either way, a positive outdegree popularity effect indicates that industries with high outdegrees produce especially attractive intermediates. When considering inputs as the main determinants of production technologies, this implies that hubs supply general-purpose technologies in the form of widely used goods. Given the process of market concentration of Oberfield (2018) is true, popularity refers to a follow-up effect: Since a decreasing number of suppliers facilitate input substitutions, the remaining ones may acquire additional buyers and strengthen their central network position. Anyway, results in Table 2 cannot provide unequivocal evidence for such tendencies. Comparing estimates with those in Table A4, it appears that the significance of the outdegree-popularity effect in the baseline specifications is mainly driven by tie changes during the period 1997-2002, which constitutes the period most severely affected by the harmonization of the data.

5.2. Structural Equivalence and Horizontal Search

In the spirit of Conley and Dupor (2003), industries can be considered complementary when they use similar inputs. In terms of production technology, a metric following this idea captures technological similarity or closeness. The corresponding network effect, namely structural equivalence, measures the (relative) number of suppliers two industries have in common. As stated above, it reflects the complementarity of technologies and is one possibility to assess the horizontal distance between industries. In general, multiple aspects can describe complementarity. However, the distance measure used here explicitly builds upon the embedment of industries in production networks. Note that structural equivalence only captures the intensity of indirect linkages of distance two. In other words, industries connected via multiple successive mediators are not considered.

Alternatively, a tendency of IO relations between similar industries, expressed in a positive coefficient, is interpretable as horizontal searching. According to the hypothesis of Carvalho and Voigtländer (2014), firms may utilize existing ties to find potential inputs. Thereby, they determine their choice set and the new state of supply chains endogenously. Carvalho and Voigtländer (2014) show that the network implications of this mechanism carry over to the sector level allowing them to bring the underlying theoretical framework to data. Concerning the network structure, specific transitive relationships, like those depicted in Figure 2, hint toward such tendencies of input decisions. However, they do not examine horizontal search but focus on input diffusion along the vertical dimension or, in network terms, on transitive closure. Yet, it is unclear why this reasoning would not apply to the case of horizontal relations.

Indeed, results in Table 2 confirm horizontal searching. Again, incorporating the period 1997-2002 leads to less pronounced effects. However, all specifications show a significant and positive estimate of the structural equivalence effect. Consider an industry i and a choice set comprising inputs j and h. Further, suppose industries i and j use the same production technology or inputs, whereas i and h differ at one incoming tie. When considering the specification in column 3, the tuple of industries (i, j) is $\exp(0.034) = 1.035$ times as likely to build an IO relation, i.e., $d_{ji} = 0 \rightarrow 1$, than (i, h).

Since this effect expresses intensities, even small coefficients can imply huge impacts on the linkages of individual industries. For 2007, 2-tuples of tied (unrelated) industries use, on average, the same 25 (20) input goods. This implies, ceteris paribus, that the odds of forming a tie to an ex-post supplier is $\exp(0.034 \times 5) = 1.185$ times the odds of using inputs from an otherwise unrelated industry.

Overall, the US production network shows clear tendencies of horizontal input diffusion – a so far undocumented fact in the emerging literature on production networks. Therefore, it is hardly possible to link this finding to existing theoretical models. Horizontal searching behavior may be an explanation for this pattern.

5.3. Transitivity and Vertical Search

Besides horizontal search, firms may engage in vertical (upstream) search as discussed in Carvalho and Voigtländer (2014). It represents a more natural form of search because it closely resembles the notion of input substitution or vertical integration (see, e.g., Lafontaine and Slade 2007) instead of complementarity.

Regarding production networks, backward transitive relations between industries indicate vertical search. In particular, the transitive closure effect measures the indirect vertical distance of industries by the number of intermediate goods. In its extensive form, the effect, referred to as the transitive ties effect, reduces to a binary variable indicating if industries are connected via at least one mediator. Compared to the structural equivalence effect, the mediator supplies to the ego and demands from the alter, i.e., alter industries produce goods used in the production of the ego's current suppliers. Hence, vertical relations of higher distances feature good flows of identical directions. A significant and positive estimate of one of these effect forms implies that industries systematically adapt ties to upstream industries. In other words, indirect vertical IO relations make industries particularly attractive.

Tables 2 and A4 present evidence thereof. Although the intensive effect is not robust with respect to sample restrictions, the extensive effect is highly significant across all specifications. In terms of the specification incorporating time dummies, the presence of vertical mediators fosters almost a factor two increase of the odds of forming a tie. In 2007, connected industry tuples in the sense of the transitive ties effect make up around 33% of the sample. It is also worth mentioning that the effect is much more prominent for the period 2007-2012, for which even the intensive measure significantly affects tie formation (see column 3 in Table A4).

A combined measure of the transitive closure and ties effects is the backward geometrically weighted edgewise shared partners (GWESP) effect with parameter α.[11] For $\alpha = 0$, the GWESP effect corresponds to the transitive ties effect; and

11. Initially Snijders et al. (2006) developed GWESP effects for ERGMs and adapted them for SOA models. Variants differ by the direction of ties defining the indirect connectivity of nodes.

Table 3.: Relative effect importance

		Structural equivalence			
		untied	tied		
Transitive ties	untied	1.000; 1.100	1.000; 1.113	1.000	1.316
	tied	1.956; 1.111	1.956; 1.118	1.956	
		1.102	1.118		
		1.102			

Note: The table indicates the relevance of the transitive ties and structural equivalence effects for ex-post tied and untied pairs of industries in the sense of these effects. It reports exponentiated and standardized average effects for industry pairs grouped by distinct forms of indirect and direct relations. Calculations rest on the point estimates in column 3 of Table 2. As reference serves the network structure of 2002, such that groups and effects reflect tie changes during the period 2002-2007.
Since the transitive ties effect is a binary indicator, standardization is superfluous. For the structural equivalence effect, coefficients are divided by the standard deviation of the observed structural effects over tied and/or untied industry pairs and periods. Hence, coefficients represent conditional (table center) and unconditional (outer table cells) effects of a one standard deviation increase in the equivalence measure and/or the existence of a relationship in the sense of the transitive ties effect on the odds of forming a tie. For the former, the table shows both effects separately, whereby the right value in each table cell is the equivalence effect.

for $\alpha \to \infty$, it approaches the transitive closure effect. However, the main results are principally invariant against this alternative effect definition (see Table A3 and columns 3, 4 in Table A4). Note that its construction complicates a meaningful and direct interpretation as the definitions in Table A2 show.

Considering their different definitions as extensive and intensive measures and the observed network structure in 2007, vertical and horizontal searching are, on average, about the same order of magnitude. Due to the different scales of structural effects, however, comparing effect sizes is misleading when assessing their relative importance for network formation. Therefore, Table 3 presents odds calculated using standardized coefficients. As a consequence, the magnitudes of the closure and equivalence effect differ. As expected, unconditionally on other effects, both tendencies are more pronounced for tied industries in the effect's sense. The odds ratios of the structural equivalence and transitive ties effect ($\frac{1.118}{1.02} = 1.096 > 1$ and $1.956 > 1$ respectively) indicate this. A similar statement applies to the conditional relevance of effects. Note that since the transitivity effect is a dummy variable and Table 3 shows averages, its coefficient is unaffected by grouping industry pairs or linkages along the lines of technological complementarity. However, the imbalance between the relevance of effects is highest for industries building a triad in the sense of transitive closure (row 2). Even when both effects are most pronounced (ex-post tied in terms of both patterns), this statement retains its validity. In fact, the horizontal search effect is relatively moderate and only outweighs tendencies toward transitivity in the absence of indirect vertical linkages (row 1). In general, the structural equivalence effect seems to be relatively less important than vertical relations for the formation of ties because it exhibits little variation across tied and untied industries. This conclusion at least partly mitigates the value of horizontal input diffusion or rather its power for explaining the heterogenous roles industries play in the connectivity of the production network. Nevertheless, its statistical significance remains robust.

6. Conclusion

This paper puts forward a new approach for assessing the dynamic structure of production networks. The empirical model simulates the unobservable dependencies between the industry-specific formation of input relations. For this purpose, it utilizes the endogeneity of input-output relations, which potentially bias estimates in regression models. Through so-called structural effects, the model allows the neighborhood of industries to affect the odds of adopting intermediates. Existing relations serve as information channels guiding the search for new inputs. This novel approach provides a framework for the systematic examination of production networks. In terms of structural tendencies or patterns, the presented work is to my knowledge and besides Carvalho and Voigtländer (2014) the only attempt to explicitly target the endogenous formation of networks. While results coincide, the presented evidence is more comprehensive when it comes to disentangling multiple distinct structural effects. The identified structural effects indicate that indirect connectivity in the form of horizontal and vertical relations of industries is an integral part of the evolution of individual relations and the network, whereby the latter dominates the former. Moreover, the model class resembles theoretical mechanisms more closely because it directly incorporates (network) factors in individual input decisions. In other words, it links the aspects at the micro level to the macroeconomic view of production networks.

Results have clear implications for current and future research focusing on micro-macro transmissions. First, the neighboring industries significantly affect input-output relations, suggesting that the revision of individual linkages eventually generates cascading network changes. Hence, theories have to consider dynamic interdependencies between economic agents. Viewing production networks as structures solely affected by exogenous factors misses out on the endogenous component. In order to fully capture the effects of idiosyncratic shocks, extensive effects in the form of cascading tie adoptions and their impact on the propagation mechanism have to be considered. Although the next logical step for understanding the role of production networks for the macroeconomy, the availability of network data complicates the empirical examination of such interaction effects.

Second, the heterogeneity in the US production network is partly endogenous. That is, hubs and mediators distinguish themselves through the structure of their direct and indirect upstream neighborhood and, contrary to most theories (e.g., Acemoglu and Azar 2020; Oberfield 2018), not necessarily via the efficiency of production technologies. In particular, hubs can result from production technologies that are particularly close to that of other industries expressed through the network distance between pairs of industries. In this sense, hubs possess general-purpose technologies. However, results do not directly relate to the fundamental mechanism behind micro-macro models, which trace aggregate fluctuations back to the dominant role of specific industries; rather they question how these models explain the endogenous emergence of hubs or the heterogeneity of industries with respect to their ties.

These implications for the current literature raise the need to reassess existing theories with respect to the compatibility with the documented tendencies. Since this paper does not focus on the detailed mechanism behind structural patterns, investigating the implications for the network structure of the most prominent theories regarding endogenous production networks may provide a narrative for the reported findings beyond searching behavior. In addition, a comparative study offers the possibility to connect effect sizes and economic theory. These issues are left for future research.

Bibliography

Acemoglu, Daron, Ufuk Akcigit and William Kerr. 2016. 'Networks and the Macroeconomy: An Empirical Exploration'. *NBER Macroeconomics Annual* 30 (1): 273–335.

Acemoglu, Daron and Pablo D. Azar. 2020. 'Endogenous Production Networks'. *Econometrica* 88 (1): 33–82.

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar and Alireza Tahbaz-Salehi. 2012. 'The Network Origins of Aggregate Fluctuations'. *Econometrica* 80 (5): 1977–2016.

Atalay, Enghin. 2017. 'How Important Are Sectoral Shocks?' *American Economic Journal: Macroeconomics* 9 (4): 254–280.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo and Matthew O. Jackson. 2013. 'The Diffusion of Microfinance'. *Science* 341 (6144).

Barrot, Jean-Noël and Julien Sauvagnat. 2016. 'Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks'. *The Quarterly Journal of Economics* 131 (3): 1543–1592.

Becker, Randy A., Wayne B. Gray and Jordan Marvakov. 2021. 'BER-CES Manufacturing Industry Database (1958-2018, version 2021a)'. National Bureau of Economic Research. Accessed 2 May 2023. https://www.nber.org/research/data/nber-ces-manufacturing-industry-database.

Bellamy, Marcus A., Soumen Ghosh and Manpreet Hora. 2014. 'The influence of supply network structure on firm innovation'. *Journal of Operations Management* 32 (6): 357–373.

Blum, Julius R. 1954. 'Approximation Methods which Converge with Probability one'. *The Annals of Mathematical Statistics* 25 (2): 382–386.

Boehm, Christoph E., Aaron Flaaen and Nitya Pandalai-Nayar. 2019. 'Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake'. *The Review of Economics and Statistics* 101 (1): 60–75.

Bresnahan, Timothy F. and M. Trajtenberg. 1995. 'General purpose technologies 'Engines of growth'?' *Journal of Econometrics* 65 (1): 83–108.

Calvó-Armengol, Antoni and Matthew O. Jackson. 2004. 'The Effects of Social Networks on Employment and Inequality'. *American Economic Review* 94, no. 3 (May): 426–454.

Carvalho, Vasco M. 2014. 'From Micro to Macro via Production Networks'. *Journal of Economic Perspectives* 28 (4): 23–48.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito and Alireza Tahbaz-Salehi. 2020. 'Supply Chain Disruptions: Evidence from the Great East Japan Earthquake'. *The Quarterly Journal of Economics* 136 (2): 1255–1321.

Carvalho, Vasco M. and Nico Voigtländer. 2014. 'Input Diffusion and the Evolution of Production Networks'. *NBER Working Paper Series,* no. 20025.

Conley, Timothy G. and Bill Dupor. 2003. 'A Spatial Analysis of Sectoral Complementarity'. *Journal of Political Economy* 111 (2): 311–352.

Conley, Timothy G. and Christopher R. Udry. 2010. 'Learning about a New Technology: Pineapple in Ghana'. *American Economic Review* 100 (1): 35–69.

Dhyne, Emmanuel, Ayumu Ken Kikkawa, Magne Mogstad and Felix Tintelnot. 2020. 'Trade and Domestic Production Networks'. *The Review of Economic Studies* 88 (2): 643–668.

Elliott, Matthew, Benjamin Golub and Matthew V. Leduc. 2022. 'Supply Network Formation and Fragility'. *American Economic Review* 112 (8): 2701–2747.

Foerster, Andrew T., Pierre-Daniel G. Sarte and Mark W. Watson. 2011. 'Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production'. *Journal of Political Economy* 119 (1): 1–38.

Forni, Mario and Sergio Paba. 2003. 'Spillovers and the growth of local industries'. *The Journal of Industrial Economics* 50 (2): 151–171.

Gabaix, Xavier. 2011. 'The Granular Origins of Aggregate Fluctuations'. *Econometrica* 79 (3): 733–772.

Giovanni, Julian di and Andrei A. Levchenko. 2010. 'Putting the Parts Together: Trade, Vertical Linkages, and Business Cycle Comovement'. *American Economic Journal: Macroeconomics* 2 (2): 95–124.

Granovetter, Mark S. 1973. 'The Strength of Weak Ties'. *American Journal of Sociology* 78 (6): 1360–1380.

Guadagni, Peter M. and John D. C. Little. 1983. 'A Logit Model of Brand Choice Calibrated on Scanner Data'. *Marketing Science* 2 (3): 203–238.

Gualdi, Stanislao and Antoine Mandel. 2018. ‘Endogenous growth in production networks’. *Journal of Evolutionary Economics* 29 (1): 91–117.

Hanneke, Steve, Wenjie Fu and Eric P. Xing. 2010. ‘Discrete temporal models of social networks’. *Electronic Journal of Statistics* 4 (none): 585–605.

Hauknes, Johan and Mark Knell. 2009. ‘Embodied knowledge and sectoral linkages: An input–output approach to the interaction of high- and low-tech industries’. *Research Policy* 38 (3): 459–469.

Jackson, Matthew O. 2011. ‘An Overview of Social Networks and Economic Applications’. In *Handbook of Social Economics,* edited by Jess Benhabib, Alberto Bisin and Matthew O. Jackson, 511–585.

Jackson, Matthew O. and Brian W. Rogers. 2007. ‘Meeting Strangers and Friends of Friends: How Random Are Social Networks?’ *American Economic Review* 97 (3): 890–915.

Koskinen, Johan H. and Tom A. B. Snijders. 2007. ‘Bayesian inference for dynamic social network data’. *Journal of Statistical Planning and Inference* 137 (12): 3930–3938.

Lafontaine, Francine and Margaret Slade. 2007. ‘Vertical Integration and Firm Boundaries: The Evidence’. *Journal of Economic Literature* 45 (3): 629–685.

Leifeld, Philip and Skyler J. Cranmer. 2019. ‘A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model’. *Network Science* 7 (1): 20–51.

Long, John B. and Charles I. Plosser. 1983. ‘Real Business Cycles’. *Journal of Political Economy* 91 (1): 39–69.

Lucas, Robert E. 1977. ‘Understanding business cycles’. *Carnegie-Rochester Conference Series on Public Policy* 5 (1): 7–29.

McFadden, Daniel. 1974. ‘The measurement of urban travel demand’. *Journal of Public Economics* 3 (4): 303–328.

Newman, M. E. J. 2002. ‘Assortative Mixing in Networks’. *Physical Review Letters* 89 (20): 208701.

———. 2003. ‘Mixing patterns in networks’. *Physical Review E* 67 (2): 026126.

Oberfield, Ezra. 2018. ‘A Theory of Input-Output Architecture’. *Econometrica* 86 (2): 559–589.

Pesaran, M. Hashem and Cynthia Fan Yang. 2020. 'Econometric analysis of production networks with dominant units'. *Journal of Econometrics* 219 (2): 507–541.

Ripley, Ruth M., Tom A. B. Snijders, Zsófia Boda, András Vörös and Paulina Preciado. 2023. *Manual for RSiena.* Accessed 12 May 2023.

Robbins, Herbert and Sutton Monro. 1951. 'A Stochastic Approximation Method'. *The Annals of Mathematical Statistics* 22 (3): 400–407.

Saviotti, Pier Paolo and Andreas Pyka. 2008. 'Micro and macro dynamics: Industry life cycles, inter-sector coordination and aggregate growth'. *Journal of Evolutionary Economics* 18 (2): 167–182.

Snijders, Tom A. B. 2001. 'The Statistical Evaluation of Social Network Dynamics'. *Sociological Methodology* 31 (1): 361–395.

———. 2005. 'Models for longitudinal network data'. In *Models and Methods in Social Network Analysis,* edited by Peter J. Carrington, John Scott and Stanley Wasserman, 215–247. Cambridge University Press.

Snijders, Tom A. B., Gerhard G. van de Bunt and Christian E. G. Steglich. 2010a. 'Introduction to stochastic actor-based models for network dynamics'. *Social Networks* 32 (1): 44–60.

Snijders, Tom A. B., Johan H. Koskinen and Michael Schweinberger. 2010b. 'Maximum likelihood estimation for social network dynamics'. *The Annals of Applied Statistics* 4 (2): 567–588.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins and Mark S. Handcock. 2006. 'New Specifications for Exponential Random Graph Models'. *Sociological Methodology* 36 (1): 99–153.

US Bureau of Economic Analysis. 2023. 'Input-Output Accounts Data'. Accessed 5 April 2023. https://www.bea.gov/industry/input-output-accounts-data.

Weitzman, Martin L. 1998. 'Recombinant Growth'. *The Quarterly Journal of Economics* 113 (2): 331–360.

A. Appendix

A.1. Data

Because IO tables are subject to redefinitions of industry classifications, data harmonization is necessary to apply the proposed econometric framework. In particular, the assignment of industries to the superordinate sectors may vary between benchmark tables. However, the BEA provides mappings showing the relation of their industry classification system and NAICS. Together with changes in NAICS codes, the comparison of both code lists gives information about the root of redefinitions.

Basically, the definitions of the BEA summarize NAICS codes. Within a year, these allocations are clear-cut; across benchmark tables, they differ substantially. Yet, their harmonization follows multiple steps. First, codes of consecutive years are matched via the associated NAICS codes, requiring that the match-relevant NAICS codes are constant within this period. During this step, the focus is on correcting for renamings and advances toward a finer or coarser classification (code splits). The outcome of this procedure is a set of mappings whereby each links the BEA industry codes of two consecutive years. Revisions of NAICS codes require special attention because they usually induce changes in the classification of the BEA. In most cases, the aggregation of codes can resolve such ambiguities. Nonetheless, a few industries cannot be assigned due to major redefinitions and remain unconsidered.

Second, the combination of the constructed mappings shows the evolution of the classification system over distant years. Thereby, it tracks the adaptions of industry definitions. Through the repeated pooling of codes starting at the first pair of observations, a harmonized code list results. It defines the aggregation rules for the make and use tables. Note that the harmonized industry classification system is necessarily of lower granularity than the original ones.

Ultimately, the list of IO relations between industries results from the make and use tables. More precisely, industries are linked via commodities or by matching products and intermediates, respectively. However, this procedure leads to numerous weak ties attributed to flows of secondary or by-products. In order to focus on the essential relations and prevent misleading results due to uninformative linkages, output relations associated with products other than the main product of each industry, apparent from their codes, are not taken into account.

A.2. Descriptive Analysis

Table A1.: Assortativity coefficients

		1997	2002	2007	2012
Panel A *Network structure*					
Degree	in-in	-0.101	-0.052	-0.042	-0.050
	out-out	0.001	-0.051	-0.050	-0.049
	out-in	-0.011	-0.010	-0.007	-0.010
	in-out	-0.156	-0.183	-0.185	-0.186
Panel B *Industry characteristics*					
Production worker hours		-0.015	-0.017	0.002	-0.008
Production worker wages		-0.014	-0.017	-0.009	-0.020
Total cost of materials		-0.017	-0.019	-0.019	-0.011
Total value added		-0.017	-0.019	-0.015	-0.013
Total capital stock		-0.019	-0.023	-0.032	-0.031
TFP		0.006	-0.007	0.015	0.032

Note: The table shows assortativity coefficients of the data-augmented and harmonized US production network. They measure the homophily of the networks with respect to industry characteristics as Pearson correlation coefficients between pairs of directly tied industries. Calculations follow Newman (2002) and Newman (2003).
Panel A includes coefficients referring to the in- and outdegrees of industries at the ends of ties or good flows. For example, the in-outdegree assortativity coefficient considers the outdegree of the industry producing the intermediate and the indegree of the consuming (downstream) industry.
Panel B describes assortativity in terms of covariates. Contrary to Panel A, statistics do not use distinct variables for industries.

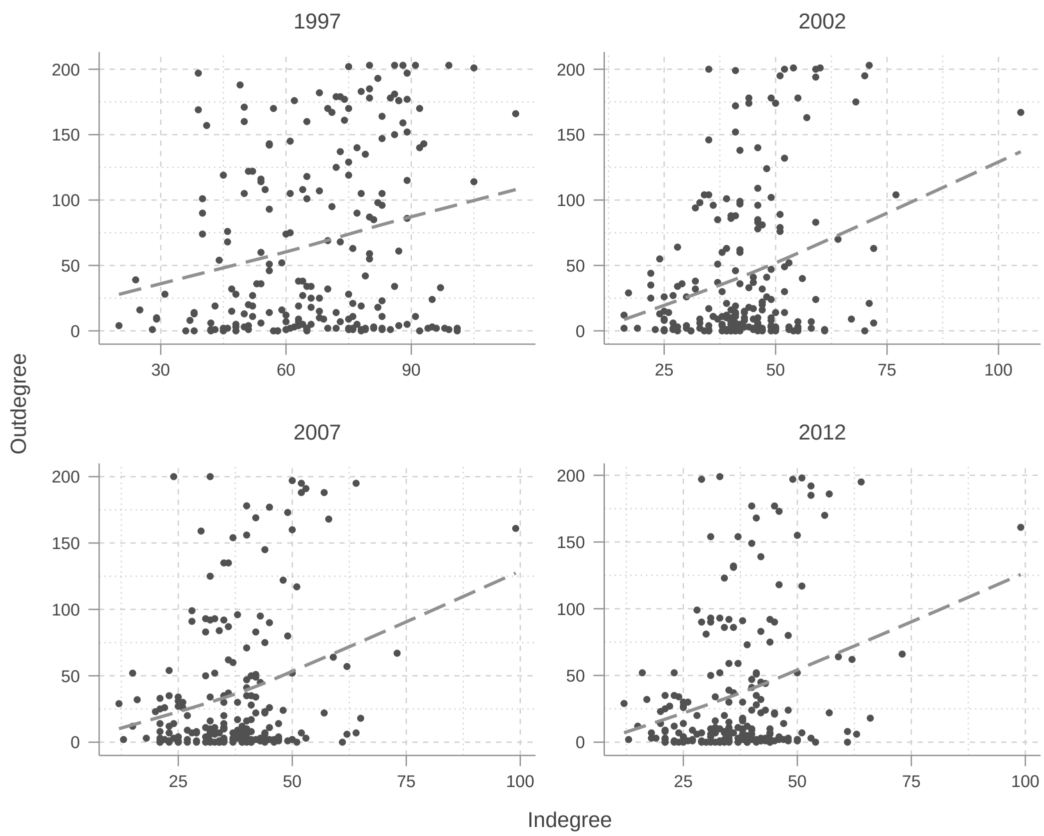


Figure A1.: In- and outdegree

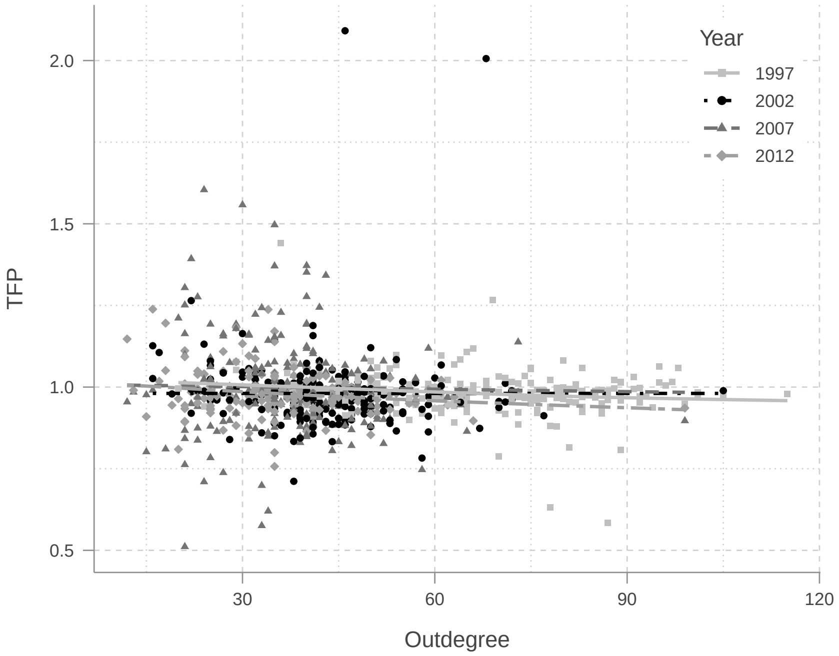


Figure A2.: Hubs and efficiency

A.3. Econometric Framework

Table A2.: Structural effects

Effect name	Formal definition $s_{ik}(\boldsymbol{D})$
Indegree density	$\sum_j d_{ji}$
Reciprocity	$\sum_j d_{ji} d_{ij}$
Transitive ties	$\sum_j d_{ji} \max_h (d_{hi} d_{jh})$
Transitive closure	$d_{ji} \sum_h d_{hi} d_{jh}$
GWESP backward (α)	$\sum_j d_{ji} e^{\alpha} \left[1 - (1 - e^{-\alpha})^{\sum_h d_{hi} d_{jh}}\right]$
Structural equivalence	$\sum_j d_{ji} \sum_{h \neq i,j} (b_0 - \lvert d_{hi} - d_{hj} \rvert)$
	$b_0 = \frac{1}{(T-1)N(N-1)(N-2)}$
	$\times \sum_{t=1}^{T-1} \sum_{h,l} \sum_{m \neq h,l} \lvert d_{mh,t} - d_{ml,t} \rvert$
3-Cycles	$\sum_{j,h} d_{ji} d_{hj} d_{ih}$
Outdegree popularity	$\sum_j d_{ji} \sum_h d_{jh}$
Outdegree activity	$(\sum_j d_{ji})(\sum_j d_{ij})$
Out-indegree assortativity	$\sum_j d_{ji} \left(\sum_h d_{ih}\right) \left(\sum_h d_{hj}\right)$
Covariate alter	$\sum_j d_{ji} x_j$
Covariate ego	$x_i \sum_j d_{ji}$
Covariate similarity	$\sum_j d_{ji} (\mathrm{sim}^x_{ji} - \overline{\mathrm{sim}^x})$
	$\mathrm{sim}^x_{ji} = \frac{\max_{hl}(\lvert x_h - x_l \rvert) - \lvert x_i - x_j \rvert}{\max_{hl}(\lvert x_h - x_l \rvert)}$
	$\overline{\mathrm{sim}^x} = \frac{1}{(T-1)N(N-1)/2} \sum_{t=1}^{T-1} \sum_{h<l} \mathrm{sim}^x_{hl,t}$

Note: The table shows the formal definitions of structural effects, whereby $\boldsymbol{D} = (d_{ji})_{1 \leqslant i,j \leqslant N}$ denotes the adjacency matrix, t the time of observations, and x an arbitrary (time-varying) demeaned node covariate. If not explicitly stated, definitions only incorporate a single network state, not a time-series of a network. However, the structural equivalence and covariate similarity effects require normalization to reduce correlations between effects during simulation. Normalizers are averages over observed network structures of the corresponding structural effect.

A.4. The Empirics of Endogenous Network Formation

Table A3.: SOA model results: GWESP

$\log \frac{\Pr(d_{ji}=1)}{\Pr(d_{ji}=0)}$	(1)	(2)	(3)
Panel A *Evaluation function*			
Indegree density	-2.526*** (0.108)	-2.284*** (0.095)	-2.004*** (0.103)
Reciprocity	0.303*** (0.047)	0.296*** (0.048)	0.305*** (0.045)
GWESP backward ($\alpha = 51$)	0.656*** (0.053)	0.655*** (0.049)	0.351*** (0.062)
Structural equivalence	0.015*** (0.002)	0.020*** (0.002)	0.032*** (0.002)
$\mathbb{1}$(1997-2002) $\times$ GWESP backward ($\alpha = 51$)			0.623*** (0.037)
$\mathbb{1}$(1997-2002) $\times$ Structural equivalence			-0.010*** (0.002)
Outdegree popularity	0.006*** (0.001)	0.004*** (0.001)	0.001 (0.001)
TFP alter		-0.980*** (0.150)	-1.083*** (0.165)
Covariates	No	Yes	Yes
Panel B *Rate function*			
1997-2002	52.372 (0.870)	50.816 (0.879)	54.381 (1.211)
2002-2007	7.302 (0.213)	7.297 (0.216)	6.889 (0.196)
2007-2012	0.731 (0.061)	0.737 (0.060)	0.727 (0.061)

Note: The table reports the estimated coefficients $\hat{\boldsymbol{\beta}}$ in Panel A for different SOA model specifications. Panel B shows estimated rate parameters $\hat{\boldsymbol{\theta}}$ and indicates the observation periods used for estimation. Coefficients are marginal effects on the log-odds of forming a tie. Industry covariates comprise TFP, total real value added, total employment, and real capital stock as ego, alter, and similarity effects. Other included structural effects are the 3-cycles, outdegree activity, and out-indegree assortativity effect.
Standard errors are given in parentheses. Based on two-sided t-tests, stars indicate the significance of estimates, i.e., ***$p < 0.01$, **$p < 0.05$ and *$p < 0.1$. Overall maximum convergence ratios are equal to 0.123, 0.145 and 0.2, respectively. Values below 0.25 indicate good convergence of the stochastic approximation method of Snijders (2001) with respect to the MoM equations (see Ripley et al. 2023). Models were estimated with the `RSiena` package.

Table A4.: SOA model results: Robustness checks

$\log \frac{\Pr(d_{ji}=1)}{\Pr(d_{ji}=0)}$	(1)	(2)	(3)	(4)
Panel A *Evaluation function*				
Indegree density	-2.256***	-2.731***	-2.482***	-2.670***
	(0.292)	(0.430)	(0.240)	(0.502)
Reciprocity	0.198*	0.028	0.219**	-0.004
	(0.117)	(0.236)	(0.094)	(0.249)
Transitive ties	0.504**	1.257***		
	(0.209)	(0.379)		
Transitive closure	-0.004	0.079***		
	(0.033)	(0.016)		
GWESP backward ($\alpha = 51$)			0.490***	1.252***
			(0.133)	(0.300)
Structural equivalence	0.038***	0.013*	0.036***	0.027***
	(0.011)	(0.007)	(0.004)	(0.008)
Outdegree popularity	0.001	-0.008**	0.000	-0.001
	(0.003)	(0.003)	(0.001)	(0.003)
TFP alter	1.786***	-0.970**	1.784***	-0.957**
	(0.474)	(0.470)	(0.442)	(0.479)
Panel B *Rate function*				
1997-2002	–	–	–	–
	–	–	–	–
2002-2007	7.396	–	7.392	–
	(0.199)	–	(0.197)	–
2007-2012	0.774	1.017	0.775	1.016
	(0.061)	(0.067)	(0.061)	(0.068)

Note: The table reports the estimated coefficients $\hat{\boldsymbol{\beta}}$ in Panel A for different SOA model specifications. Panel B shows estimated rate parameters $\hat{\boldsymbol{\theta}}$ and indicates the observation periods used for estimation. Coefficients are marginal effects on the log-odds of forming a tie. All specifications control for TFP as ego, alter, and similarity covariate effect. Other included structural effects are the 3-cycles, outdegree activity, and out-indegree assortativity effect.
Standard errors are given in parentheses. Based on two-sided t-tests, stars indicate the significance of estimates, i.e., ***$p < 0.01$, **$p < 0.05$ and *$p < 0.1$. Overall maximum convergence ratios are equal to 0.137, 0.165, 0.118 and 0.14, respectively. Values below 0.25 indicate good convergence of the stochastic approximation method of Snijders (2001) with respect to the MoM equations (see Ripley et al. 2023). Models were estimated with the `RSiena` package.