Molnár and Csala *Applied Network Science* (2023) 8:71
https://doi.org/10.1007/s41109-023-00599-8

Applied Network Science

RESEARCH Open Access

Threshold sensitivity of the production network topology

Eszter Molnár[1,2,3] and Dénes Csala[1,3,4*]

*Correspondence: d.csala@lancaster.ac.uk

[1] Department of Engineering, Lancaster University, Lancaster, UK
[2] Material Social Futures Doctoral Training Centre, Lancaster University, Lancaster, UK
[3] Faculty of Economics and Business Administration, Babeş-Bolyai University, Cluj-Napoca, Romania
[4] Economics Observatory, School of Economics, University of Bristol, Bristol, UK

Abstract

Industries today are tightly interconnected, necessitating a systematic perspective in understanding the complexity of relations. Employing network science, the literature constructs dense production networks to address this challenge. However, handling this high density involves carefully choosing the level of pruning to retain as much information as possible. Yet, current research lacks comprehensive insight into the extent of distortion the data removal produces in the network structure. Our paper aims to examine how this widespread thresholding method changes the production network's topology. We do this by studying the network topology and centrality metrics under various thresholds on inter-industry networks derived from the US input-output accounts. We find that altering even minor threshold values significantly reshapes the network's structure. Core industries serving as hubs are also affected. Hence, research using the production network framework to explain the propagation of local shocks and disturbances should also take into account that even low-value monetary transactions contribute to the interrelatedness and complexity of production networks.

Keywords: Production network, Input-output analysis, Monetary transactions, Network science, Threshold

Introduction

No industry can exist and develop individually. All sectors are interdependent through the exchange of products and services (Xu and Liang 2019). Given the increasing complexity of our society, we cannot examine a problem without the help of multidisciplinary approaches. Inter-industry relations are profoundly interconnected; therefore, we must use techniques that view industries from a systematic perspective and understand the complexity of relations. In this study, we use network and data science tools to explore traditional economic science issues, specifically the question of industry interrelatedness through the lens of complex production networks.

From the economic perspective, an industry is a branch of an economy that produces closely related raw materials, goods, or services, while from the business point of view, it is a group of companies that are related based on their primary business activities. There are dozens of industry classifications, and these classifications are typically grouped into larger categories called sectors.

Leontief's (Leontief 1970; Dietzenbacher and Lahr 2008) economic input-output models represent, in mathematical form, the monetary transactions between industry sectors. They specify what goods and services (output) are consumed by other industries (input).

The network science approach to input-output models is not a novel concept. There are mainly two basic approaches currently being adopted in this research area. One is the analysis from a supply chain perspective, using company-level data (Wu 2015; Brintrup and Ledwoch 2018; Perera et al. 2017), and the other is the industry perspective, using country or global-level input-output accounts.

From the industry perspective approach, a considerable amount of literature has been published using the World Input-Output Database (Xu and Liang 2019; Soyyiğit and Boz 2018; Grazzini and Spelta 2015; Baldwin and Lopez-Gonzalez 2015; Soyyiğit and Çırpıcı 2017), covering 40 countries in the 2013 release and 43 countries in the 2016 release, all with 35 (2013) and 56 (2016) sectors (Timmer et al. 2015). Although this data source can cover most countries, it only contains information on very aggregated sectors. In recent years, researchers have also investigated various approaches to the input-output transaction data of the US economy as systematised by the Bureau of Economic Analysis (BEA) https://www.bea.gov/data/industries/input-output-accounts-data. Most of the studies focused on the sector and summary level of the input-output accounts containing 21 (sector-level) and 71 (summary-level) aggregated industries (Foerster and Choi 2017; Duan 2012).

Sectoral inter-dependencies are pivotal in connecting microeconomic shocks with business fluctuations and cycles, especially during the supply-chain fluctuations of today. Researchers claim that production networks derived from input-output models provide a valuable account in opening the black box of co-movement and propagation mechanisms that shape aggregate outcomes. Carvalho (2014) points out that production network research can advise scholars on the origins of aggregate fluctuations and policymakers on how to be ready and recuperate from disadvantageous shocks that disturb production chains (Keith et al. 2017; Dykes and Sterman 2017).

Horvath (1998, 2000) and Acemoglu et al. (2012) argue that the structure of the production network plays a crucial role in determining the aggregate behaviour of the system. Network structure is the set of nodes and edges of a network. Nodes representing industries in the production network carry certain quantitative properties, which are represented as weights. The edges representing the monetary transactions between industries also carry weights, but in the case of a production network they are also directed - and the weights vary significantly in terms of directions.

When the production network is outstandingly asymmetric, for example, when few sectors are in a dominant role as suppliers, idiosyncratic shocks lead to aggregate fluctuations. If the production organisation is dominated by a small number of hubs supplying inputs to many different sectors, disturbances at these crucial nodes will affect the global production system, determining losses in production and welfare (Acemoglu et al. 2012). Bigio and La'O (2016) also demonstrate that overall network topology defines the strength of each channel. They compare two production networks: a horizontal and a vertical economy, and show that the network centrality of sectors matters for how they affect aggregate output. Carvalho (2014) also advocates for the significance of network topology by comparing the amplification of micro-level volatility and the network multiplier in a horizontal economy with no input trade, a vertical economy with a source and a sink, and a star/hub-and-spoke economy with a central node/s.

The US Bureau of Economic Analysis publishes the Industry Economic Accounts generally at three levels of detail: sector (21 industry groups), summary (71 industry groups), and detail (405 industry groups). For example, at the sector-level, Durable goods are included between the 21 industry groups. This sector at the summary-level contains Primary metals, Machinery, Computer and electronic products and 8 other sub-sectors. While broken down even further to detail-level, the Primary metals include 10 industries (Iron and steel mills and ferroalloy manufacturing, Ferrous metal foundries, etc.), the Machinery includes 28 industries (Farm machinery and equipment manufacturing, Semiconductor machinery manufacturing, etc.), and Computer and electronic products include 20 industries (Electronic computer manufacturing, Telephone apparatus manufacturing, etc.). These industry classifications are all grouped hierarchically into three levels.

Although extensive research has been carried out from an industry perspective, just a few studies exist that develop a network of at least 400 detail-level industries. This can be a key problem because the networks built on a highly aggregated level with just a few nodes and connections might not represent the industry interdependencies accurately.

On the one hand, the topology of a detail-level industry network can differ considerably from an aggregate-level network. The first one tends to be way denser with more lower-weighted connections, thereby behaving differently. On the other hand, some essential links could be hidden in an aggregate-level network. For example, embedded in a highly-weighted connection, several detail-level links that might be more important than the other present aggregated ones could have been hidden. The whole map of industrial interdependence could change when analysing these separately. The detail-level input-output account data might allow us to discover a more representative picture not just with more separable sectors, but with way more supporting connections between industries, in number and validity. This increased granularity of understanding allows for better identification of key industries undergoing change and more efficient tracking and understanding of innovation.

Most scholars, like Carvalho, use a threshold value and consider only a percentage of the monetary transactions to be present in the production network. As the monetary transactions will become the edges between industry-industry pairs when constructing the production network, they do this mainly to make the data more manageable. Initially, the input-output models have almost i^2 supporting transactions, with i being the number of industries. The production networks obtained from such detail-level input-output models are exceptionally dense and challenging to analyse without offering much insight.

This problem can be mitigated by carefully choosing the level of intersection to minimise the network structure distortion caused by data removal (Radicchi et al. 2011; De Benedictis et al. 2014). The goal is to prune a tiny fraction of the total weight so that as much information as possible is preserved (García-Algarra et al. 2019). Hence, most scholars impose a threshold and exclude a percentage of the smaller monetary transactions—thus "severing" these links in the production network. In our study, we refer to this cut-off point as threshold ζ.

Collectively, these studies outline a critical need to examine how the threshold ζ influences the topology of the production network and, thus, the propagation mechanisms and aggregate fluctuations. While we might agree that production networks can provide a bridge between micro and macro (Carvalho 2014), very little is currently known about how much the threshold value ζ chosen by scholars distorts this bridge.

In this study, we discover the considerable change in the production network topology to the threshold value ζ by analysing a detail-level input-output model with a high number of sub-sectors to understand the most accurate structure of the network of industries. The Data and methodology section presents the chosen dataset, the production network model, and the topological and node-level analysis.

Data and methodology

Data source

We use the detail-level United States Input-Output Accounts Data (US Input-Output Accounts Data) to conduct this research. Leontief's work (Leontief 1970; Dietzenbacher and Lahr 2008) is the foundation of the Bureau of Economic Analysis' work today in creating the U.S. Input-Output Accounts. The Bureau of Economic Analysis publishes the Industry Economic Accounts generally at three levels of detail: sector (21 industry groups), summary (71 industry groups), and detail (405 industry groups). Estimates at the detail level are produced roughly every five years, with the last two releases being from 2007 and 2012. While there are Historical Benchmark Input-Output Tables from 1947 until 2002, these reflect industry definitions that vary across years, and BEA advises that they should not be used as a time series. BEA uses its Industry Codes for the three levels of detail, but it also defines how these relate to the 2012 North American Industry Classification System (NAICS) code structure (North American Industry Classification System).

We use the last two releases, the 2007 and 2012 Total Requirements table - Industry by Industry, including 405 industries and inter-industry purchases (Total Requirements Data). Additional explanations regarding the Total Requirements derivation can be found on BEA's website (Total Requirements Definition).

Production network generation from data

The BEA Total Requirements table has a network representation in which an element w_{ij} represents the nominal amount of goods i used as input by sector j, with $i, j = 1,..., N$, where N is the number of sectors.

For a two-industry example of a network from the Total Requirements table, see Fig. 1. The network built from this database is a directed weighted graph. The vertices are the industries (i and j), and the directed connections are the monetary transactions, the nominal flow of goods between sectors. The weight of each link is the economic value, representing how much an industry supports the development of the other (w_{ij} and w_{ji}). For example, industry i requires w_{ij} dollar input from industry j to produce one dollar output to final users (typically $w_{ij} \neq w_{ji}$).

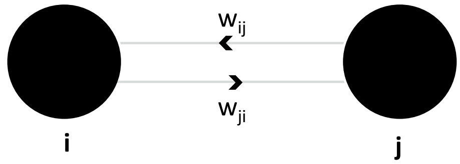

Fig. 1 Example of a two-industry production network. The vertices *i* and *j* are the industries, and the connection weights w_{ij} and w_{ji} are the normalised monetary transaction values between sectors

The input-output models also indicate if the output of an industry is required as input to the same industry (w_{ii}). For example, the Oil and gas extraction industry could produce the oil and gas to power its own equipment, or the Computer design and manufacturing industry can produce computers that are used to plan the next generation of computers (Economic Input-Output Life Cycle Assessment). We don't use these self-sector transactions in our analysis.

Threshold ζ

After building the production network from the Input-Output Accounts Data, we define thirty different edge cut-off thresholds ζ from 0.00001 to 0.15 with 0.005 equal intervals. The threshold ζ is associated with the value of the links, representing inter-industry trade in monetary expression. For example, at the end of our analysis, when the threshold value ζ is 0.15; we only consider those connections that weigh at least 0.15. In other words, only those monetary transactions become edges in the resulting production network where at least 0.15 dollars is needed from one industry to produce one dollar output in another. At this threshold ζ, in the 2012 production network, only 142 industries (nodes) from the starting 405 and 156 monetary transactions (edges) are present. We choose this long interval and these small steps to represent the best topology change according to the cut-off ζ.

Algorithms for random and scale-free models

Then, we *prune* the graph according to the considered cut-off thresholds ζ. We generate a random and scale-free graph with the same parameters as our production network at each cut-off point ζ. For both, we use the NetworkX graph generator algorithms (NetworkX library).

For the random graph, we use the directed *G(n,m)* algorithm, with the parameters n being the node number and m the edge number. We give the same node and edge number for the random graph generator algorithm as in our original production network.

For the scale-free graph, we use the algorithm implemented after Bollobás et al. (2003) with the same node number and by calculating α, β and γ parameters to fit our production network. The probability of adding a new node in this algorithm is distributed between the parameters α, β and γ, hence the sum of these should be 1. α is the probability of adding a new node connected to an existing node chosen randomly according to the in-degree distribution, while γ is the opposite of this: adding a new node according to the out-degree distribution. β is the likelihood of adding an edge between two existing nodes. One existing node is chosen randomly according to the in-degree distribution, and the other is chosen randomly according to the out-degree distribution.

In simple terms, α and γ define whether the directed network is scale-free from the in-degree or the out-degree perspective. We represent these parameters by checking the the original production network's top in-degree and out-degree values. In this context, this means comparing the top buyer to the top supplier in terms of the number of different transactions.

Degree distribution comparison (Kolmogorov–Smirnov test)

However, there are some other methods, too (Barrat et al. 2004; Newman 2005); we choose the Kolmogorov–Smirnov (KS) statistical test to compare our observed degree distribution to what we would expect from a random and a scale-free graph with the same parameters because it is widely used when comparing degree distributions for networks (Deng et al. 2011; Muchnik et al. 2013; Gómez et al. 2008; Gjermëni 2017).

As our production network is a directed network, we distinguish between in-degree and out-degree:

$$k_{in,i} = \sum_{j=1}^{N} a_{ij} \qquad k_{out,i} = \sum_{j=1}^{N} a_{ji} \tag{1}$$

where $k_{in,i}$ is the incoming degree of node i, representing the number of incoming edges onto the node, a_{ij} is 1 if there is a directed transaction between industry i (buyer) and industry j (seller). In other words, if industry i requires the output of industry j; otherwise it is 0. The outgoing degree is $k_{out,i}$, representing the number of links that point from node i to other nodes. In this case, a_{ji} is 1, if the industry i is a supplier to industry j. In these measures, self-sector transactions, indicating if the output of an industry is required as input to the same industry, are not included (a_{ii}).

We separately analyse and compare in-degree and out-degree distribution at each threshold ζ by observing when the network behaves like a random or scale-free graph. We use the SciPy two-sample KS statistical test based on Hodges (1958). The KS statistic quantifies the distance between the empirical distribution functions of the two samples (the observed degree distribution and the random/scale-free graph's degree distribution). The smaller this value is, the more likely the two samples are drawn from the same distribution.

Centrality measures

We calculate two network centrality measures for every node at six cut-off points ζ from 0.0 to 0.1 with 0.02 equal intervals on the directed weighted graph. In this interval, most monetary transactions are disregarded at minimal threshold values (see the distribution of the edge weights in Fig. 2), and at 0.1 dollars, there are still more connections than nodes in the production network. With 6 cut-off points ζ in this area, we can illustrate well that the ranking of the prominent sectors is also highly dependent on the threshold value.

The degree is defined as the total number of links of a node. The production network is weighted and directed; thereby, we compute the weighted in-degree (in-strength) and the weighted out-degree (out-strength) of each node:

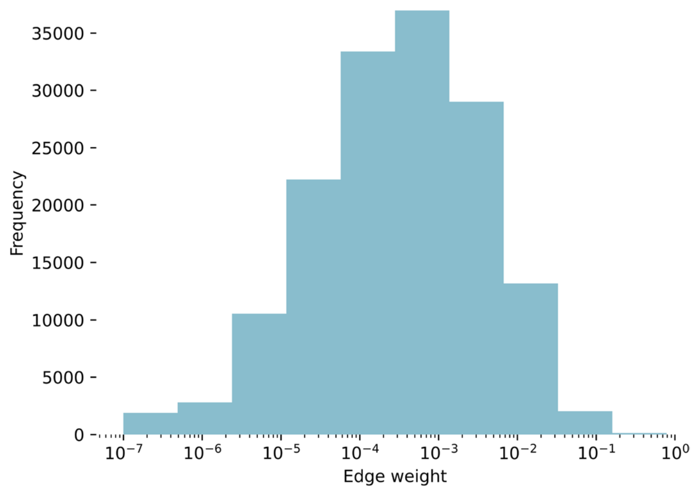

Fig. 2 Global edge weight distribution (Lin-log scale). This figure shows the edge weight distribution in the production network. The edge weights (normalised monetary transaction values) are shown on the horizontal axis on a logarithmic scale, and the vertical axis shows the number of edges on a linear scale

$$k^w_{in,i} = \sum_{j=1}^{N} w_{ij} \qquad k^w_{out,i} = \sum_{j=1}^{N} w_{ji} \tag{2}$$

where $k^w_{in,i}$ is the in-strength of industry i, representing the sum of all incoming flows of goods, the nominal inputs used by the sector i, w_{ij} is the transaction value between industry i (buyer) and industry j (seller). In other words, the required amount from industry j to produce one dollar of output from industry i. The outgoing degree for industry i is $k^w_{out,i}$, representing the sum of the outflow of goods from node i to other nodes. In this case, w_{ji} is the nominal input that industry i supplies to industry j. We don't use self-sector transaction weights representing in what amount the output of an industry is required as input to the same industry (w_{ii}).

We choose the weighted out-degree and the weighted *PageRank* out-degree centrality (Page et al. 1999; Langville and Meyer 2004) as our key metrics. The first represents how big and indispensable an industry is in monetary terms, *globally,* in the entire production network, and the second represents how key an industry is in terms of *location* in the network.

The reason behind the PageRank algorithm is that a node is systemically significant if its neighbours are important and/or the neighbours of the neighbours are important. However, the production network is a directed network, and in directed networks, we distinguish the links based on their directions. Defining the centrally located nodes in undirected networks is relatively straightforward, while it is much more complicated in directed networks. Most centrality metrics in network research, such as the PageRank, are based on in-degree (the number of edges pointing to the node). Hence, in this case, based on the number of suppliers. Although, the production network is quite a particular network in this sense. One might ask if those industries are the central ones that need the most resources or the ones that pump the most supply into the network. The critical sectors that dominate economic activity are all hidden if we only consider the in-degree-based centrality metrics. That being the case, we calculate the weighted PageRank out-degree centrality on the reversed production network. The inverted network contains the same nodes and edges, but the direction of the edges is reversed.

At each threshold ζ, we show the top 20 central industries and compare how they changed according to the cut-off.

Besides this, we also compare the weighted in-degree and out-degree distribution without a cut-off ζ to dig deeper into the asymmetry analysis.

Results

Our results provide a first insight into the production networks' topology change according to the threshold value ζ.

We first analyse the topological features and find that the industry inter-dependence network topology is highly sensitive to the chosen threshold ζ. Starting with the smaller monetary value transactions, as we disregard the higher and higher ones, the production network's topology transforms; from the out-degree perspective, it leans towards a scale-free structure and, from the in-degree perspective, towards a random structure.

Figure 2 shows the edge weight distribution throughout the whole production network. In other words, the normalised monetary transaction values distribution in the national economy as it is present in the US input-output accounts. It is quite apparent from this figure that the vast majority of monetary transactions are minimal in their value. Therefore, the very small threshold ζ values cut out the most transactions.

Figure 3 shows the Kolmogorov–Smirnov statistical test's value at various edge cut-off thresholds ζ. We compare the production networks' observed in-degree and out-degree distribution at each threshold ζ to what we would expect from a random and scale-free network with the same parameters. The closer the KS statistic value is to 0, the more likely the two samples are drawn from the same distribution.

With the threshold ζ increasing, the edge numbers decline at a very past face in the beginning. From the starting value of more than 150,000 edges, only around a third, 50,000 edges, remain at the minimal threshold of 0.001 dollars. The same goes for the threshold of 0.005 dollars with 20,000 edges and the threshold of 0.01 dollars with 10,000. The node number during these thresholds stays the same. After that, the decline is a bit slower in the edge number, and the nodes start to decrease slowly, too. We end the analysis at a reasonable threshold of 0.15 dollars with 142 nodes and 156 edges.

Neither the in-degree nor the out-degree perspective is similar to the random or scale-free distribution with no threshold or very low threshold value ζ. After a while, the in-degree distribution inclines towards a random structure and the out-degree towards a scale-free.

According to out-degrees, the scale-free shift appears around the 0.05 threshold value ($\zeta \approx 0.05$) when the KS statistic falls below 0.1 (in the 2007 network 0.08, and in the 2012 network 0.07) and the p-value exceeds the critical 0.05 (in the 2007 network 0.15, and in the 2012 network 0.23). It stays in the range until the final ζ threshold value of 0.15 analysed in the 2007 network with a KS statistic of 0.1 and a p-value of 0.48 and in the 2012 network with a KS statistic of 0.09 and a p-value of 0.59. The production network is the most scale-free at the 0.065 ζ threshold value in 2007 (KS $=$ 0.029, p-value $=$ 0.99) and at the 0.085 ζ threshold value in 2012 (KS $=$ 0.021, p-value $=$ 0.99).

Fig. 3 Kolmogorov–Smirnov statistic value of the distributions compared. This figure shows 4 line charts that show the KS statistical test's value at 30 edge cut-off thresholds ζ when comparing the production networks' observed degree distribution to what we would expect from a random and scale-free network with the same parameters, according to year and degree type. The closer the KS statistic value is to 0, the more likely the two samples are drawn from the same distribution. The node and edge numbers are also included according to 4 cut-off points ζ

The in-degree network perspective starts to have a topology more similar to a random network after the 0.05 threshold value (in the 2007 network KS $=0.07$, p-value $=0.21$, and in the 2012 network KS $=0.08$, p-value $=0.18$). Just as with the out-degree perspective, the in-degree stays in the random range too, having at the end of the comparison, at 0.15 threshold in the 2007 network, a KS statistic of 0.12 and a p-value of 0.19 and in the 2012 network a KS statistic of 0.17 and a p-value of 0.09. It is most similar to a random network at the 0.06 ζ threshold value in 2007 (KS $=0.054$, p-value $=0.67$) and at the 0.08 ζ threshold value in 2012 (KS $=0.053$, p-value $=0.78$).

Our results show that the threshold changes the production network structure quite much. We might actually say that even with these minimal step thresholds, the production network becomes a different entity at every cut-off, especially during the first phase of imposing the threshold. However, that is not entirely true, as the same industries and the highest transactions in value remain in the network. While the topology slowly shifts. The most notable change from both dimensions is between the 0.0 and 0.05 thresholds. At the 0.05 threshold while still having the vast majority of the industries and the most important monetary transactions in value (in the 2007 network: 373 nodes, 1154 edges, and in the 2012 network: 363 nodes, 1232 edges) the topology from the in-degree perspective takes the shape of a random network and from the out-degree side the shape of a scale-free structure. At this point, 0.8% of all original transactions are present in the network.

Carvalho (2014), in his production network research, accounts for about 80% of the total value of input trade, which would leave us with 5 000 edges out of 150 000, 3.5% of all transactions in the network, around the 0.02 cut-off point ζ ($\zeta \approx 0.02$).

In-degree and out-degree distributions of the production network behave very differently, as is expected. It is more similar to a random network from the in-degree perspective, whereas, from the out-degree perspective, it leans towards a scale-free topology.

The scale-free nature of trade networks is a broadly analysed subject in production network research, too (Gualdi and Mandel 2016; Liu et al. 2021).

To discover more about this asymmetry, we show the weighted in-degree and out-degree distribution of the entire production network without cut-off ζ in Fig. 4.

Some scholars have already approached the question of asymmetry in directed networks (Grazzini and Spelta 2015; Horvath 1998, 2000; Acemoglu et al. 2012; Luo and Whitney 2015; Wang and Wang 2017). In the production network case, the underlying explanation is quite transparent.

The topological difference comes from the presence of critical sectors that dominate economic activity, the so-called "commanding heights". Vladimir Lenin used this phrase in the early 1920s, referring to the control of key segments of a national economy. The difference between raw and processed materials is evident in the degree distribution comparison. The core industries that drive the economy formed hubs, pushing the topology to the scale-free range from the out-degree perspective. From the in-degree perspective, most sectors need the same amount of resources; therefore, the topology leans towards a random network.

According to size and centrality measures, our results show that some core industries tend to remain on top at each threshold ζ. Figure 5 reveals the node-level threshold ζ sensitivity of the production network. We calculate the top 20 sectors according to weighted out-degree and weighted PageRank out-degree centrality at six cut-off points ζ. We found that the top three industries stay the same: 1. Iron, still mills, ferroalloy manufacturing, 2. Oil and gas extraction, 3. Petroleum refineries.

Because the threshold ζ change cuts off the minor value transactions, some industries leading the top lists drop to the bottom of the rankings. These are usually the industries that support almost all the other ones but with smaller transactions, such as Truck

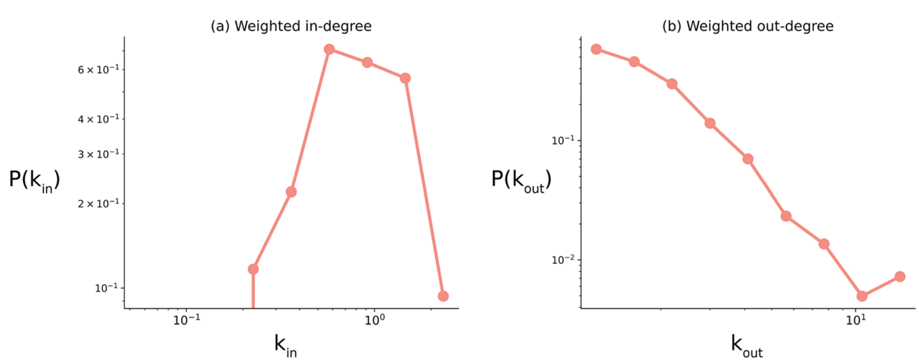

Fig. 4 Node degree distribution of the production network. The node **a** in-degree and **b** out-degree distribution of the entire production network without cut-off in log-log scale and by considering the amount of each monetary transaction (the weights of the edges)

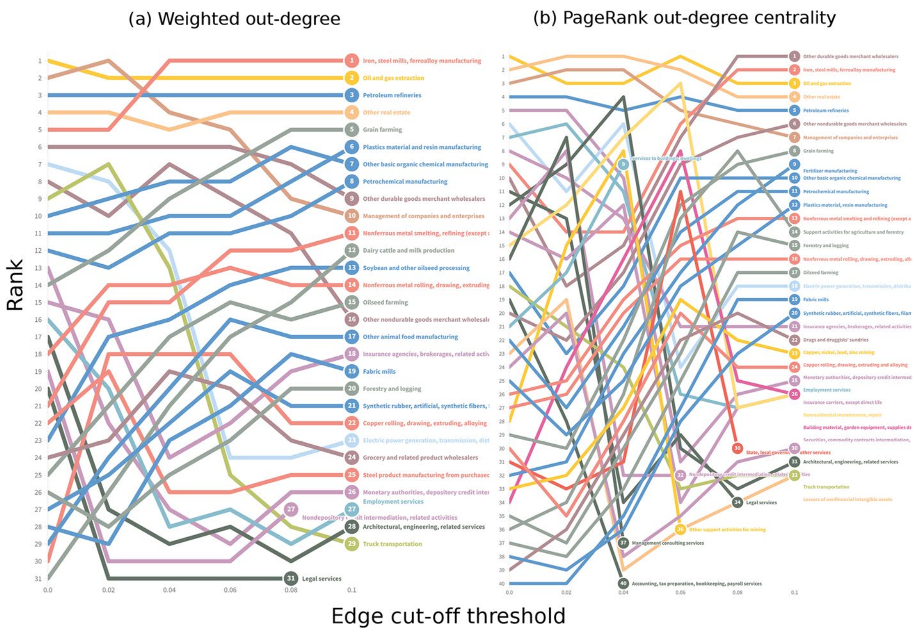

Fig. 5 Top industries according to threshold ζ This figure shows the node-level threshold ζ sensitivity of the production network. We reveal the top 20 sectors according to **a** weighted out-degree and **b** PageRank out-degree centrality at six cut-off points ζ

transportation and Electric power generation, transmission, and distribution. Therefore, we can conclude that the threshold ζ cuts out some significant hubs that amplify aggregate fluctuations.

Discussion

This study aimed to analyse the widely used threshold's influence on the production network topology.

A growing body of literature recognises the importance of production network research, but no attention has been paid yet to the thresholding methods in this framework. Our hypothesis was that changing the percentage of monetary transactions that researchers include in the network (edge cut-off threshold ζ) changes the structure of the network and the core industries to a large extent. Hence, if the topology is highly sensitive to the threshold, that can have further implications for studies based on this.

We prove our hypothesis by examining the network topology and centrality metrics under different thresholds ζ on a network derived from the US input-output accounts data for 2007 and 2012.

We discovered that the industry inter-dependence network is highly exposed to the edge cut-off threshold ζ. The topology of a production network with no threshold or minimal threshold ζ cannot be classified as a random or scale-free network structure. However, as we disregard monetary transactions increasing in number and size, the topology from the in-degree perspective leans towards a random network structure and, from the out-degree side, takes the shape of a scale-free network.

The node-level threshold ζ susceptibility analysis shows us that not just the overall topology changes but some core industries that served as hubs drop to the bottom of the top lists as we decrease the number of transactions included. These industries mainly support numerous other industries with small monetary transactions.

In his study, Carvalho (2014) mentions that hubs shorten distances. While cutting out these transactions, we lose these hubs and increase distances. These monetary transactions contribute directly to the production networks' connectedness and complexity, which explains the propagation (and sometimes circularity) of local shocks and disturbances. For example, a reasonable question to ask would be, if we do not consider them all, in what proportion will our propagation modelling be distorted?

Even small-value supporting transactions could make a difference. For example, while we might cut out most of the Truck transportation transactions, a disturbance in the industry could impact all the others dependent on it, even with just a small amount. If there is an interruption in Truck transportation, manufacturing industries shake too because, among other effects, probably, the spare parts don't arrive in time. Therefore, we lose some essential propagation mechanisms if we do not consider these transactions.

These small monetary transactions not only directly contribute to propagation, as Truck transportation directly supports other industries, but also function as intermediaries. If we cut them, we ignore all the propagation mechanisms along the chain. The Truck transportation disturbance might impact the manufacturing industries, but this shock might propagate even further in smaller proportion to the third-forth industries connected to the manufacturing ones.

Another aspect that is quite important, but we did not capture through this analysis is that these interdependencies are in no way linear. For example, the Truck transportation industry might contribute only 5% to Car manufacturing. Still, if the Truck transportation industry stops, the other sector falls to zero without access to that crucial part. This limitation of our study can be overcome by examining exactly this resilience aspect in future work.

This study has shown that the threshold value ζ chosen makes a significant difference in the production network's topology; therefore, we must carefully consider and keep in mind the limitations of the research made on production networks using a particular threshold ζ.

As a further step of the research, it would be reasonable to explore the distribution fitting with other methods too (Barrat et al. 2004; Newman 2005; Barabási and Pósfai 2016) and to analyse the change of some concrete propagation mechanisms of particular industries in the context of threshold modification.

Also, another further development of the study could be to examine the clustering structure of the industries in the production network (Fagiolo et al. 2008, 2013; Bartesaghi et al. 2022, 2023), compare it to the sectoral categories, and study how this is distorted through the threshold change.

Acknowledgements

We would like to thank Levente Szász for his valuable advice.

Author contributions

EM and DC designed the research. EM collected, preprocessed, and visualized the data. EM and DC analyzed the data and interpreted the results. EM wrote the manuscript. DC reviewed and revised the manuscript.

Funding
EM acknowledges the support of the Leverhulme Doctoral Scholarships Programme in Material Social Futures.

Availibility of data and materials
The datasets generated and analysed during the current study are available here: https://github.com/molnareszter/Threshold-sensitivity-of-the-production-network-topology

Declarations

Competing interests
The authors declare no competing interests.

Received: 13 June 2023 Accepted: 25 September 2023
Published online: 05 October 2023

References

Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A (2012) The network origins of aggregate fluctuations. Econometrica 80(5):1977–2016
Baldwin R, Lopez-Gonzalez J (2015) Supply-chain trade: a portrait of global patterns and several testable hypotheses. World Econ 38(11):1682–1721. https://doi.org/10.1111/twec.12189
Barabási A-L, Pósfai M (2016) Network science. Cambridge University Press, Cambridge
Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Natl Acad Sci 101(11):3747–3752. https://doi.org/10.1073/pnas.0400087101
Bartesaghi P, Clemente GP, Grassi R (2022) Community structure in the World Trade Network based on communicability distances. J Econ Interact Coord 17(2):405–441. https://doi.org/10.1007/s11403-020-00309-y
Bartesaghi P, Clemente GP, Grassi R (2023) Clustering coefficients as measures of the complex interactions in a directed weighted multilayer network. Physica A 610:128413. https://doi.org/10.1016/j.physa.2022.128413
Bigio S, La'O J (2016) Financial frictions in production networks, p 84
Bollobás B, Borgs C, Chayes J, Riordan O (2003) Directed scale-free graphs. In: Proceedings of the fourteenth annual ACM-SIAM symposium on discrete algorithms, pp 132–139
Brintrup A, Ledwoch A (2018) Supply network science: emergence of a new perspective on a classical field. Chaos Interdiscip J Nonlinear Sci 28(3):033120
Carnegie Mellon University Green Design Institute | Economic Input-Output Life Cycle Assessment (EIO-LCA). http://www.eiolca.net. Accessed 06 Apr 2023
Carvalho VM (2014) From micro to macro via production networks. J Econ Perspect 28(4):23–48
De Benedictis L, Nenci S, Santoni G, Tajoli L, Vicarelli C (2014) Network analysis of world trade using the BACI-CEPII dataset. Glob Econ J 14(03n04):287–343. https://doi.org/10.1515/gej-2014-0032
Deng W, Li W, Cai X, Wang QA (2011) The exponential degree distribution in complex networks: non-equilibrium network theory, numerical simulation and empirical data. Physica A 390(8):1481–1485
Dietzenbacher E, Lahr M (2008) Wassily Leontief and Input-Output Economics. Cambridge Books. Cambridge University Press, Cambridge
Duan W-Q (2012) Modelling the evolution of national economies based on input-output networks. Comput Econ 39(2):145–155
Dykes K, Sterman J (2017) Dynamics of innovation and diffusion in large-scale complex technical systems: the case of wind energy, p 30
Fagiolo G, Reyes J, Schiavo S (2008) On the topological properties of the world trade web: a weighted network analysis. Physica A 387(15):3868–3873. https://doi.org/10.1016/j.physa.2008.01.050
Fagiolo G, Squartini T, Garlaschelli D (2013) Null models of economic networks: the case of the world trade web. J Econ Interact Coord 8(1):75–107. https://doi.org/10.1007/s11403-012-0104-7
Foerster A, Choi J (2017) The changing input-output network structure of the U.S. economy. The Federal Reserve Bank of Kansas City Economic Review
García-Algarra J, Mouronte-López ML, Galeano J (2019) A stochastic generative model of the World Trade Network. Sci Rep 9(1):18539. https://doi.org/10.1038/s41598-019-54979-1
Gjermëni O (2017) Temporal statistical analysis of degree distributions in an undirected landline phone call network graph series. Data 2(4):33
Gómez V, Kaltenbrunner A, López V (2008) Statistical analysis of the social network and discussion threads in slashdot. In: Proceedings of the 17th international conference on world wide web, pp 645–654
Grazzini J, Spelta A (2015) An empirical analysis of the global input-output network and its evolution. In: DISCE—Working Papers del Dipartimento di Economia e Finanza def031, Universit Cattolica del Sacro Cuore, Dipartimenti e Istituti di Scienze Economiche (DISCE)
Gualdi S, Mandel A (2016) On the emergence of scale-free production networks. J Econ Dyn Control 73:61–77. https://doi.org/10.1016/j.jedc.2016.09.012
Hodges JL (1958) The significance probability of the smirnov two-sample test
Horvath M (1998) Cyclicality and sectoral linkages: aggregate fluctuations from independent sectoral shocks. Rev Econ Dyn 1(4):781–808
Horvath M (2000) Sectoral shocks and aggregate fluctuations. J Monet Econ 45(1):69–106
Input-Output Accounts Data | U.S. Bureau of Economic Analysis (BEA). https://www.bea.gov/data/industries/input-output-accounts-data. Accessed 09 March 2023

Keith DR, Sterman JD, Struben J (2017) Supply constraints and waitlists in new product diffusion. Syst Dyn Rev 33:254–279
Langville A, Meyer C (2004) A survey of eigenvector methods of web information retrieval. SIAM Rev 47:135–161
Leontief W (1970) Environmental repercussions and the economic structure: an input-output approach. Rev Econ Stat 52(3):262–271
Liu L, Shen M, Tan C (2021) Scale free is not rare in international trade networks. Sci Rep 11(1):13359. https://doi.org/10.1038/s41598-021-92764-1
Luo J, Whitney DE (2015) Asymmetry in in-degree and out-degree distributions of large-scale industrial networks
Muchnik L, Pei S, Parra LC, Reis SDS, Andrade JS Jr, Havlin S, Makse HA (2013) Origins of power-law degree distribution in the heterogeneity of human activity in social networks. Sci Rep 3(1):1783
NetworkX Python package. https://networkx.org. Accessed 22 Nov 2022
Newman M (2005) Power laws, Pareto distributions and Zipf's law. Contemp Phys 46(5):323–351. https://doi.org/10.1080/00107510500052444
Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking: bringing order to the web
Perera S, Bell MGH, Bliemer MCJ (2017) Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective. Appl Netw Sci 2(1):1–25
Radicchi F, Ramasco JJ, Fortunato S (2011) Information filtering in complex weighted networks. Phys Rev E 83(4):046101. https://doi.org/10.1103/PhysRevE.83.046101
Soyyiğit S, Boz C (2018) Global input-output analysis: a network approach. Yildiz Soc Sci Rev 3:89–102
Soyyiğit S, Çırpıcı YA (2017) An input–output network structure analysis of selected countries. Yildiz Soc Sci Rev 3:65–88
Timmer MP, Dietzenbacher E, Los B, Stehrer R, de Vries GJ (2015) An illustrated user guide to the world input-output database: the case of global automotive production. Rev Int Econ 23(3):575–605
Total Requirements Data | U.S. BEA Input-Output Accounts. https://apps.bea.gov/industry/xls/io-annual/IxI_TR_2007_2012_PRO_DET.xlsx. Accessed 28 Aug 2023
Total Requirements Definition | U.S. Bureau of Economic Analysis (BEA). https://apps.bea.gov/industry/pdf/TotalRequirementsDerivation.pdf. Accessed 17 Jan 2023
Wang L, Wang X (2017) The influences of edge asymmetry on network robustness
Wu L (2015) Centrality of the supply chain network
Xu M, Liang S (2019) Input-output networks offer new insights of economic structure. Physica A 527:121178

Publisher's Note