ANDRAS BORSOS | MARTIN STANCSICS

UNFOLDING THE HIDDEN STRUCTURE OF THE HUNGARIAN MULTI-LAYER FIRM NETWORK

MNB OCCASIONAL PAPERS | 139

2020
NOVEMBER

UNFOLDING THE HIDDEN STRUCTURE OF THE HUNGARIAN MULTI-LAYER FIRM NETWORK

MNB OCCASIONAL PAPERS | 139

2020
NOVEMBER

MNB Occasional Papers 2020/139

Unfolding the hidden structure of the Hungarian multi-layer firm network

(A többrétegű magyar cégházózat struktúrájának feltárása)

Written by Andras Borsos* **, Martin Stancsics*** ****

Budapest, November 2020

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

Szabadság tér 9., H-1054 Budapest

www.mnb.hu

ISSN 1585 5678 (on-line)

*Central European University

**Central Bank of Hungary

***Central Bank of Hungary

****We benefited from discussions with Eszter Balogh, Peter Bauer, Zsuzsanna Hosszu, Janos Kertesz, Peter Lang, Bence Mero and Zsolt Olah. We are grateful for the comments received at the NetSci, CCS, CERS and NetWorkshop conferences. This paper has grown out of the PhD dissertation of the first author at Central European University. Contacts: borsosa@mnb.hu, stancsicsma@mnb.hu

Contents

Abstract

In this paper we offer an explorative, mainly descriptive analysis about the multi-layered network of Hungarian firms. To conduct this study, we obtained access to firm-level supplier information, on which we could superimpose also the ownership background of all Hungarian companies. Our primary focus was to explore the topological origins of shock propagation phenomena among firms. As both the supplier and the ownership layers are considered to be among the most significant shock-transmitting mediums, our data is ideal to gain insight into previously unobserved structural drivers of spreading processes. We found (i) several topological traits on micro-, meso-, and macro-scale as well, which can be responsible for facilitating contagious processes via supplier links; (ii) we could also identify separated blocks of the economy (representing different production chains) within which shocks can more freely spread in the system; (iii) furthermore, we could assess the significance of economic entities regarding the extent they can influence the economy via their ownership relations.

JEL: C63, C67, C81, G32, L23.

Keywords: ownership network, production network, supply chain, spillover, shock propagation, contagion.

Összefoglaló

Jelen tanulmány egy feltáró, többnyire leíró jellegű elemzést kínál a magyar cégek közötti többszintű hálózat bemutatására. Annak érdekében, hogy ez az adatintenzív kutatás létrejöhessen, hozzáférést nyertünk a magyar cégek közötti beszállítói kapcsolatokat tartalmazó adatokhoz, amihez hozzárendeltük a cégek tulajdonosi hátterének hálózatát is. Mivel mind a szállítói, mind a tulajdonosi réteget a legjelentősebb sokközvetítő kapcsolatrendszerek közé sorolhatjuk, a rendelkezésünkre álló adatbázis lehetővé teszi, hogy betekintést nyerjünk a vállalatok közti terjedési hatások korábban nem megfigyelhető strukturális hátterébe. Az eredményeink között (i) kimutattunk számos (mikro-, mezo- és makroskálán mérhető) topológiai jellemzőt, amelyek felelősek lehetnek a szállítói hálózatban történő terjedési mechanizmusokért; (ii) azonosítottuk a gazdaság elkülönülő csoportosulásait (amelyek különböző termelési láncokat reprezentálnak), (iii) továbbá felmértük az egyes szereplők jelentőségét a tulajdonosi kapcsolatokon keresztül kifejtett befolyásolási potenciál szempontjából.

1 Introduction

In the past decade we experienced a vast surge in interest towards modeling and analyzing interdependencies among companies. In this study we offer a general, descriptive exploration of the topological structure of firm networks using Hungarian data. The structure of these networks is a key element in understanding the governing forces behind any kind of spreading phenomena among firms. Although the topology of the underlying graph might play a different role for the various types of shocks, our work is relevant for a wide range of applications, such as productivity spillovers (Liu et al. (2000), Gorg and Strobl (2001)), spreading of financial shocks (Demir et al. (2018), Costello (2020)), or upstream and downstream supply chain disruptions (Barrot and Sauvagnat (2016), Carvalho et al. (2016)).

Examining the system of interfirm connections has been present in the economic literature at least since Leontief's seminal work on the structure of the American economy (Leontief (1951)). The roots of the recent increased enthusiasm in using more fine-grained, firm-level disaggregation are twofold: (i) developments in data availability and (ii) new conceptual innovations.

As a part of the universal pattern of increased accessibility to micro-level data, it became possible in some countries to obtain comprehensive datasets about firm-level connections. Previously, researchers who wanted to consider firm connections in their analyses could use either a sample of the given network or a higher aggregation level (e.g. industry or country). Both approaches turned out to suffer from serious limitations. When measured on a sample, even the most elementary characteristics of networks (e.g. the density or the average degree) require non-trivial corrections, which can be very different depending on the sampling method (Granovetter (1976)), while more sophisticated analyses on samples are hindered severely by potential distortions (Frank (1971)). The other option is to use a completely observed, but aggregated system, however, this approach has other caveats. One seemingly obvious drawback is that during the process of aggregation we lose information not only about the heterogeneity of the actors, but also about their connections among each other. However, it was not evident at all for a long time in economics (at least from the point of view of macroeconomics) whether disregarding the observation of firm-level events and characteristics is relevant or not.

As this debate flared up and gained a lot of attention recently, it leads to the second, more theoretical branch of factors giving popularity to granular firm network analysis in economics. An important milestone in the development of this field was the rejection of the traditional argument of Lucas Jr (1977) about the diversification of shocks in the economy. The former consensus was that firm-level idiosyncratic events do not have any influence on the macroeconomic scale as they cancel each other out based on the law of large numbers. However, Gabaix (2011) and Acemoglu et al. (2012) showed, that due to the heterogeneity of the firms and the topology of their connections stress events of the largest companies cannot be offset by smaller firms even if the shocks are uncorrelated[1]. Another vastly influential branch of research on economic networks is the topic of contagions in the financial sector (e.g. Gai and Kapadia (2010), Acemoglu et al. (2015)), which direction gained momentum after the widespread recognition of the topology of bank networks as one of the main drivers of instability. In parallel to these trends in macroeconomics and macro-finance, the empirical literature on supply chain contagion gained popularity as well. In these papers researchers try to measure the extent of spreading of some exogenous event (e.g. natural catastrophe or policy shock) on the supplier network. See for example Barrot and Sauvagnat (2016), Carvalho et al. (2016), Demir et al. (2018), Boehm et al. (2019), but Carvalho and Tahbaz-Salehi (2018) and Bernard and Moxnes (2018) offer reviews of the broader literature on production networks.

These results about the unexpected directions and rate of shock spreading proved that the interconnectedness of the economic and financial systems is a vastly influential aspect of many processes of the economy. However, this observation cannot be simply interpreted that more connections means higher potential for any kind of spreading phenomena. E.g the seminal work of Elliott et al. (2014) showed, that the contagion potential in a financial system depends on the network structure on a non-monotonic way: Diversification (having more economic partners) increases the size of the connected components in the

[1] They showed that the distribution of company size (and also the direct and indirect demand towards a given company's products) could be described well using power-low distribution, in which there is a relatively high probability of extremely large observations. Depending on the exponent of the distribution, the assumptions of the law of large numbers could be violated.

network initially, but after a while this process actually leads to more diversified, more resilient systems. Similar logic can be observed in the case of dual source strategies in supply chain management[2].

These developments triggered a new wave for economic network analysis as network theory offered a novel way of thinking about the structure of the economy by representing it as a complex system, in which non-linear interactions and feedbacks create emergent phenomena. For now, the network-based approach has become part of the mainstream in several areas. Nevertheless, the first step before one could integrate these –- now sometimes almost fully observable -– networks into economic models should be the thorough exploration of the data. Most importantly we have to examine the high resolution of the topological structure, which is the distinguishing feature compared to the previous aggregate observations. However, one can mainly find theoretical works aiming to describe these systems (e.g. Benhabib et al. (2010), Dutta and Jackson (2003), Goyal (2012), Jackson (2010)), and just a very limited number of empirical papers.

In one of these works, Watanabe et al. (2015) offers a detailed analysis of trade connections of 400,000 Japanese firms. Although it is still about a sample of firms in the country, they were the first to analyze a supplier network of this extent. Dhyne et al. (2015) describes the production network of Belgium from the point of view of its integration into the world trade network. This work was one of the first members of a series of papers about the Belgian production network (Kikkawa et al. (2019), Magerman et al. (2016), Tintelnot et al. (2018)), however all these further research projects focused on economic questions with a higher abstraction level and not on the network itself. Another branch of empirical research considers ownership relations among companies, which is also a recently often examined layer of firm networks. These papers usually use the Orbis database[3] to analyze the global ownership network of companies: Using the same or highly overlapping data sources Glattfelder (2013) offered a methodology to extract the backbone of the global ownership network, Vitali et al. (2011) showed that there is a very high concentration in this network with a group of core companies, Vitali and Battiston (2011) examined the ownership structures' embedding in the geographical space, Vitali and Battiston (2014) explored the community structure of the global ownership network, Heemskerk and Takes (2016) described the multipolar nature of the global political economy, and Garcia-Bernardo et al. (2017) tried to identify offshore financial centers. Some other layers of corporate connections were studied as well, however, often only on smaller samples of firms. Zajac and Westphal (1996), Battiston et al. (2003) and Davis et al. (2003) looked at the network of interlocking board members and decision makers. Innovation dynamics have been also considered on networks of R&D partnerships e.g. by Tomasello et al. (2017), while there are numerous studies focusing on stock price correlation-based network of listed companies (e.g. Tumminello et al. (2010)).

Most of the papers until now dealt with only a single layer of corporate networks. A very recent exception is Jeude et al. (2019) which study features four layers: ownership links and board member overlaps among a very large sample of firms; furthermore R&D collaborations and stock correlation between listed companies. In our research we want to contribute to this direction in the literature by attempting to unfold the non-trivial characteristics of the multi-layered network of firms using Hungarian data. At the Central Bank of Hungary, we could create a uniquely rich dataset by having access to supplier transaction information among firms, on which we can superimpose their ownership links as well in the period between 2014-2017. The supplier information is coming from firms' VAT reports collected by the National Tax and Customs Administration of Hungary. In this data we can observe trade links among Hungarian firms where the tax content of the transactions between two firms exceeds EUR 3000 in the given year. Considering the ownership data we used the OPTEN dataset of more than 400 000 Hungarian firms. To further enrich the scope of our analysis we added other micro-level datasets of the Central Bank of Hungary. These made it possible to use the detailed characteristics of firms (coming from their balance sheets and profit and loss statements) as additional attributes of the nodes.

Both the supplier and the ownership layers are among the most significant shock-transmitting mediums; thus, our data is ideal to explore the topological origins of the above described spreading phenomena. As these data are usually not collected for research purposes, and economists are often unfamiliar with the specific and unique characteristics of network data, we also contribute by highlighting the significant amount of preprocessing which is necessary to use these information in line with the economic expectations and interpretations. The methodological approach of our analysis consists of elements coming not only from economics, but also from the network science literature (Newman (2018), Barabási et al. (2016)), which can provide us with suitable tools to explore the underlying structure of the firm network on micro-, meso- and macro-scales as well.

[2] Firms often establish more than one link for a given input to decrease their sensitivity to disruptions in the supply chain.

[3] Orbis is a company database provided by Bureau van Dijk (which is a Moody's Analytics business information publisher).

Regarding our results, we firstly considered the ownership structure of the Hungarian economy in order to distinguish transactions within and between ownership groups. As this system is much sparser than the supplier layer, and it consists of many small components, we could not analyze it on the level of a giant component. However, we could still measure its most important characteristics and we gained insight into the typical motifs in the ownership structure of firms.

Furthermore, the ownership data enabled us to explore the network of economic actors from another angle as it conveys invaluable information about the direct and indirect influence of the observed entities. Based on our corrected measure of control, we could also analyze the distribution of control in the economy. We found that more than 40% of the control is associated with the top 100 owners in the Hungarian economy. We also assessed the role of different groups formed based on numerous dimensions, such as the nationality, the legal category or the HQ location of the owners.

Regarding the supplier layer, we have found several topological patterns of the production network which can be responsible for facilitating contagious processes: (i) despite the low density of the network we can identify a giant component which encompasses more than 94% of the nodes. (ii) The average longest path length among the firms in this component is around five[4], which indicates small-worldness in the network. (iii) The long-tailed degree distribution ensures the presence of hubs, that can be key actors in spreading shocks. (iv) Contagions can be further promoted by micro-level motifs: there is an unexpectedly high probability of reciprocal dyads and closed triangles, which can amplify shocks via local feedback loops.

We could gain other valuable insights about the system by exploring its meso-level configuration. We identified well-defined and occasionally overlapping community structure, which reflects closely the production chains of different segments in the Hungarian economy. This grouping allowed us to assess firms' capacity to connect communities, which measure can be used as a proxy for shock transmitting ability between the otherwise separated chains of production. We found that firms in the transportation and infrastructure sectors, and firms with high productivity and high export rate have the most important role in connecting different blocks of the economy. In addition to the topological information, we also added several node attributes to the network. Using these variables, we measured strong homophily[5] based on several firm characteristics, most notably in the case of productivity, profitability and geographical location, however, these traits are much weaker in terms of separating the network than the supply chains identified by our community detection method.

As the scope of this paper covers only the preparation and exploration of these special datasets, there are plenty of options for further research directions. Thus, here we only list a few options we want to explore as next steps in our research. We are planning to enhance the banking system contagion model of the Central Bank of Hungary by including real-economy feedback effects coming from the production network. In order to be able to do this, we also have to estimate the spillover effects of financial shocks in the supply chain. Furthermore, by exploiting the presence of both the supplier and the ownership data, it is also possible to examine the vertical and horizontal M&A activities, or more generally, the link formation strategies of the observed firms.

The rest of the paper is organized as follows. Section 2 discusses the analysis on the ownership layer, while Section 3 describes the exploration of the supplier network. Section 4 summarizes the results, limitations and further research plans.

[4] Here we did not consider the directions of the links as shocks can spread in both directions depending on the process.

[5] The tendency of the formation of links between similar nodes.

2 Ownership network

At the Central Bank of Hungary we obtained access to OPTEN's firm-level ownership data[6] of more than 400 000 Hungarian firms for the period between 2015-2019. Besides the ownership links, the data contains some attributes of the firms and their owners as well, but we can also merge this data to other micro-level datasets in order to use additional characteristics of the companies. A more detailed description of the quality and the cleaning of the OPTEN data can be found in Appendix A.

2.1 NETWORK TERMINOLOGY AND DEFINITIONS

To accomplish a formal analysis of the ownership network we have to introduce some basic concepts to represent a network in a mathematically interpretable way:

- In graph theory, the number of links connected to a given node (i) is called the *degree* (k_i). If the links have directions, we distinguish the *indegree* (k_i^{in}) and *outdegree* (k_i^{out}), showing the number of links coming in and going out in the case of a particular node.
- The links can also have weights which correspond to the ownership share in our data. In the case of a weighted network, we can calculate the *strength* (s_i) of a node instead of its degree by summing up the weights of the links associated with the given node.
- If we want to refer to the whole network, the simplest – although computationally often very inefficient – way is to represent it as an *adjacency matrix* (A or in the case of weighted networks W), where $A_{i,j}$ (or $W_{i,j}$) corresponds to the ownership share of actor i in actor j. The size of this matrix is $(m + n) \times (m + n)$ where m and n are the number of firms and the number of individuals in the network respectively.
- The density of a (sub)graph is defined as the ratio of the number of edges to the number of possible edges in the network[7].
- A *(connected) component* is a subgraph of the network where at least one path exists between every pair of nodes. We can distinguish between *strong* and *weak* forms of connectedness. The former requires that only directed paths can be considered, while the latter ignores the direction of the links.
- The *local clustering coefficient* shows the probability that two neighbors of a node are connected to each other as well forming closed triads.
- The *average shortest path length* shows the average number of steps it takes to get from one member of the network to another. It is calculated by finding the shortest path between all pairs of nodes, and taking the average over them.

2.2 TOPOLOGICAL ANALYSIS OF THE OWNERSHIP NETWORK

Although we analyzed the data for every year which we could observe, the topological structure showed very similar results for all the observed periods; therefore, we present here only the description of the year 2017, which is the last year for which we have access to every datasets we are using during the analysis. (The basic description of the network in other years can be found in Appendix A.)

The network consists of more than 1 million nodes (firms and individuals as well) and almost the same number of edges (ownership relation between nodes), which implies that the average indegree (or outdegree) is somewhat less than one (Figure 1). The first important observation is that the network is not connected, i.e. it consists of many (259 138) components (using the

[6] OPTEN is a Hungarian firm-level data provider company.

[7] This definition is valid in the case of simple graphs, where there are no self-links or multi-edges between the nodes).

weak form of connectedness), among which even the largest one contains only around 11% of the nodes, while all the others have maximum a few 100 members (Figure 2). One can also consider strongly connected components, but they would capture only partial information about ownership structures. For instance, if two individuals are owners of a firm, we would observe only one of them within one component (as there is no directed path between the two owners). This way, the largest strongly connected component contains only 19 nodes.

Due to the low edge density of the network the size of the giant component remains rather limited as it encompasses only 11% of the nodes. [8] This largest component with an average degree of 4.4 is not as sparse as the network in general, and it also features some interesting characteristics:

- Its degree distribution is fat-tailed indicating the presence of actors with outstanding influence (Figure 3).
- The average shortest path length is 13.78. This value might seem low, however, it is not as low as it is typical in the case of many observed small-world networks. One additional reason for this high number can be the fact that a large portion of nodes represent individuals, which are restricted to have outgoing links only.
- We also calculated the local clustering coefficient, which is more than 0.18. It is a way higher probability than having a link between two randomly chosen nodes. This result reveals an important structural pattern in the network which is worth examining in more details by calculating motif statistics.

As it can be seen on Figure 4, there are 16 types of motifs consisting of three nodes (Davis and Leinhardt (1967)). In the case of a sparse network, the vast majority of the cases fall into the first, disconnected category, however, the distribution of the remaining motifs is very uneven. Although motif c) depicts only dyadic connections, it is interesting to observe that there are many reciprocal ownership relationships in the network. As individuals cannot be owned, these dyads can be formed only between firms. Based on motifs d) and e) it occurs more often that a firm have more than one owners, than having more than one firms in an actor's ownership, which implies difference in the *indegree* and *outdegree* distributions. While there are only 2896 observations of the simplest chain structure shown by motif f), we could find many ownership connections intertwined in more convoluted ways, e.g. following the patterns of motifs i), l) and m). Furthermore, it is surprising to notice the high number of instances in the case of motif p) which illustrates a fully connected triad with all the possible links among the nodes.

2.3 MEASURING *INFLUENCE* AND *CONTROL* OF OWNERS

Our data make it possible to assess the significance of economic entities from the point of view of the extent they can influence and control the economy via their ownership relations. In order to enable us to properly analyze this aspect of the ownership structure, we had to define measures for the manifestation of the economic actors' power. Although our methods sometimes differ from the analysis made by Glattfelder (2013), we often follow the approach and terminology of that paper in this section.

In order to carry out this analysis we have to apply a few corrections to the raw ownership information. Firstly, it is not obvious at all, how much actual power is entailed to a given ownership share. Secondly, we want to consider not only direct, but also indirect ownership links to gain accurate assessment about the influence of a given actor in the economy. Thirdly, we also need to take into account some measure of the sizes of owned firms. In the following subsections we describe our approach to deal with these points.

2.3.1 DISTRIBUTION OF OWNERSHIP SHARES

To assess the influencing ability entailed to the observed ownership in our data we have to consider at least two potential distortions. The first one considers the assumption that ownership shares correspond to voting share. Although there are several common practices in corporate governance to deviate from the one-share-one-vote principle, it is credible to assume (for instance based on Silanes et al. (1999)) that in the vast majority of cases we can use ownership as a proxy for influence manifested in the voting rights. The second bias, however, may require more effort to correct for. Owners or shareholders

[8] In network theory the term *giant component* can sometimes be defined in a rigorous way, however, very often it is a rather loosely used concept. Here we simply mean the largest component which includes a significant portion of the nodes.

of a company can be considering not necessarily just as individuals but rather as rivaling voting blocks. In this mindset it is obviously incorrect to assume perfect proportionality between ownership and effective influence. The most common example to illustrate the difference between the two is the following distribution of ownership shares: 49% – 49% and 2%. In this case all three owners have practically the same influence as any two of them can form a block to gain majority.

One can find numerous similar examples, but it is far from being obvious how to create a correction which covers as many of these situations as possible, but it is still tractable computationally. Several so called power indices were proposed regarding this problem (see e.g.Leech (2002)), but there is no consensus in the literature on best practices. Because of its simplicity and efficacy we decided to apply the method proposed in Glattfelder (2013). The underlying idea of this measure is that the actual influence of an owner depends not only on its own ownership share, but also on the distribution of ownership shares of the other owners. The more dispersed the ownership structure is, the higher the influence of the given owner is. To calculate this concentration-corrected measure of influence they are using a version of the Herfindahl-index in the following way:

$$H_{i,j} := \frac{W_{i,j}^2}{\sum_{l \in P_j^{in}} W_{l,j}^2} \tag{1}$$

where $H_{i,j}$ is the corrected ownership share of owner i in firm j, $W_{i,j}$ is the original ownership share and P_j^{in} is the set of indices of neighbors (owners) connected to j by incoming links. This measure can take values in the interval $(0,1]$. If $H_{i,j}$ is close to one, it means that firm i has almost exclusive influence on firm j. Based on this measure we can calculate the *direct influence* of any owner by summing up all of its influence scores:

$$h_i := \sum_{j \in P_i^{out}} H_{i,j} \tag{2}$$

where P_i^{out} is the set of indices of neighbors connected to i by outgoing links.

2.3.2 DIRECT AND INDIRECT OWNERSHIP

An obvious shortcoming of Eqs.(2) is that it only considers direct ownership links. To account for indirect connections Brioschi et al. (1989) proposed a method called the *integrated model*. The main component of their approach can be written in the form of a recursive computation:

$$\widetilde{H_{i,j}} := H_{i,j} + \sum_{n \in P_i^{out}} H_{i,n} \widetilde{H_{n,j}} \tag{3}$$

where $\widetilde{H}$ denotes *integrated influence*. The interpretation of this formula is that the actual influence of owner A on a firm B consists of two elements: the direct influence of owner A on firm B and the integrated influence on firm B by other firms owned by owner A. This expression can be written in matrix form as well:

$$\widetilde{H} = H + H\widetilde{H} \tag{4}$$

which gives the following solution:

$$\widetilde{H} = (I - H)^{-1} H. \tag{5}$$

Although Brioschi et al. (1989) showed that the mathematical requirements to conduct this calculation are always satisfied in an ownership network, there still can be computational constraints if the matrix representing the ownership network is large. In our case the inversion of the $(I-H)$ matrix was prohibitive, therefore, we calculated its Neumann-series approximation:

$$(I-H)^{-1} = I + H + H^2 + H^3 + \dots \tag{6}$$

This method is intuitively interpretable since the k^{th} power of an adjacency matrix gives us the number of the routes with length k between two nodes. If we add up all the powers, we will cover all the indirect links in the network in the end. Due to the large memory requirement of storing large matrices, we could compute the approximation only up to the 6th power. This happened because although the original matrix is very sparse, it is not necessarily true for its inverse, or even for the higher powers of it. However, as the average length of the shortest paths is relatively short in any component of the network, six steps can cover the vast majority of the relevant ownership links. Consequently, the elements of the resulting matrix ($\widetilde{H}$) can be interpreted as the *total influence* of owner A on firm B.

Another limitation of this calculation is that we cannot observe global ultimate beneficiary owners (UBOs) as we only see such ownership relations, for which at least one of the endpoints of the links is a Hungarian firm. As ownership ties between foreign entities would be necessary to trace the exact paths of more convoluted offshore activities, we cannot take these into account in the investigation of the owners' indirect influence over the Hungarian economy.

2.3.3 WEIGHTING OF INFLUENCE

Our measures so far did not consider any information regarding the significance of the owned firms. By adding a non-topological node attribute to correct for this could enhance the precision of the assessment of owners' influence considerably. We decided to use the simple approach to multiply the matrix of *total influences* ($\widetilde{H}$) by the vector of some approximation of firm values. While recognizing the depth of the methodologies in corporate valuation, the amount of firms in this exercise grants justification for opting for the simplest possible option to evaluate firms. Some of the obvious candidates as proxies for firm value could be e.g. capitalization (for listed companies), or total asset value (for smaller firms). The resulting *total controlled value* measure could be formulated like this:

$$\widetilde{c_i} := \sum_{j \in P_i^{out}} \widetilde{H_{i,j}} v_j \tag{7}$$

where v_i denotes the i's firm value.

However, the OPTEN data does not contain any variable which we could use as a proxy of firms' value, therefore we had to join another firm-level dataset coming from the Hungarian Tax Authority containing the balance sheets and profit and loss statements of firms. Unfortunately, the overlap between these datasets is not perfect, i.e. we cannot match the required firm characteristics to almost 22% of the firms in the ownership network.

A more serious caveat of this approach is the multiplication of firms' value when one conducts the aforementioned computation. The problem arises due to the fact that the value of a given firm contains the proportional part of the value of the companies owned by this firm as well. (A more detailed illustration of this problem can be found in Appendix B.)

To solve this issue, we wanted to find a node attribute, which is independent from the ownership structure, but conveys some information about firms' weight in the economy. A better candidate to meet these requirements is the value added of firms, which can be directly applied as a replacement for the previous value proxies. With this solution we only have to make a slight modification on Eqs.(7) by replacing total assets with the real value added of firms:

$$\widetilde{c_i} := \sum_{j \in P_i^{out}} \widetilde{H_{i,j}} rva_j \tag{8}$$

Although this measure is clearly not an ideal proxy for firms' value, it gives more accurate results if one wants to compare the control of different economic actors than the naive approach of using traditional firm size variables.

Based on this measure, we could calculate the empirical cumulative distribution of the total control of owners, which can be seen in Figure 6. It is important to note, that neither the total influence nor the total control of a given firm include itself. The interpretation of this plot is then the following: the top right corner of the diagram represents 100% of the owners controlling 100% of the the economy's value added, and the first data point in the lower left-hand corner denotes the most important owner. The red lines indicate that the top 10 owners control more than 17%, and the top 100 owners control more than 40% of the economy (measured by the real value added of firms).[9]

2.4 INFLUENCE AND CONTROL BASED ON THE OWNERS' ATTRIBUTES

Besides the ownership links, the OPTEN data contains some attributes of the owners as well. Most importantly, we can see whether the owner is a firm or an individual as well as the country level location of its headquarter. We calculated the *direct* and *total influence* and *control* measures aggregated along the dimensions of Hungarian/foreign and firm/individual owners. As it is shown in Figure 7, there is only a small gap between the *direct* and *total* versions, however, the difference is way more pronounced between the *influence* and *control* results. As the average value added of companies owned by foreign owners and by firms is much higher, their significance is heavily underestimated in the case of the unweighted influence measures. Moreover, Figure 8 reveals that foreign firms have the biggest role among these categories by controlling around 37% of the value added in the Hungarian economy. Hungarian firms and individuals have almost the same amount of *total control* (31%), while foreign individuals have much smaller significance by exercising less then 1% control.

We can make similar analysis on a more disaggregated level concerning the significance of foreign countries in the Hungarian economy (Figure 9)[10]. As holdings and special purpose firms designed for tax optimization might distort the results especially in the case of the *control* measures, *direct* and *total influence* might be a better indicator for foreign countries' importance in the Hungarian economy. For example The Netherlands is generally not as important economic partner for Hungary as Germany or Austria, but there are several large companies which control their Hungarian subsidiaries through holding entities with headquarters in The Netherlands. As the organizational structure of these transnational companies can change frequently based on their strategic decisions, we can observe significant changes in the control measures of countries, while their influence remains relatively stable over the examined years (as it can be seen if one compares Figure 9 and Figure 10).

If we focus only on the owners belonging to the *Hungarian firm* category, we can examine more disaggregated levels by adding further attributes from our additional firm dataset. Figure 11 and 12 shows our measures of significance of owner firms aggregated based on their head quarters' location at the level of counties (NUTS 3) and regions (NUTS 2) of Hungary. Although these diagrams are calculated based on partial data without considering the role of individuals and foreign entities, the results are in line with intuition that the more developed areas such as the capital and the counties with major towns play a more important role in the ownership network. (E.g. Fejer county is a traditional hub for large Hungarian industrial companies, such as Videoton, Dunaferr, Kofem.)

We can also use firm size categories as an alternative aggregation dimension (Figure 13). Although micro-enterprises have the largest role based on every measure, it can be misleading to rely on only one type of metrics as *influence* greatly underestimates the significance of large companies.

We carried out this analysis also based on the NACE industry categories of the Hungarian owner firms (Figure 14). We can see the dominance of the finance and insurance industries in the *control* measures, however, the *construction* and the *professional, scientific and technical activities*[11] industries are even more influential based on their *influence*.

[9] In this calculation we did not assume any strategic cooperation between owners to control firms, and we did not take into account the fact, that having 50% + ε ownership share can often be sufficient to fully control a company.

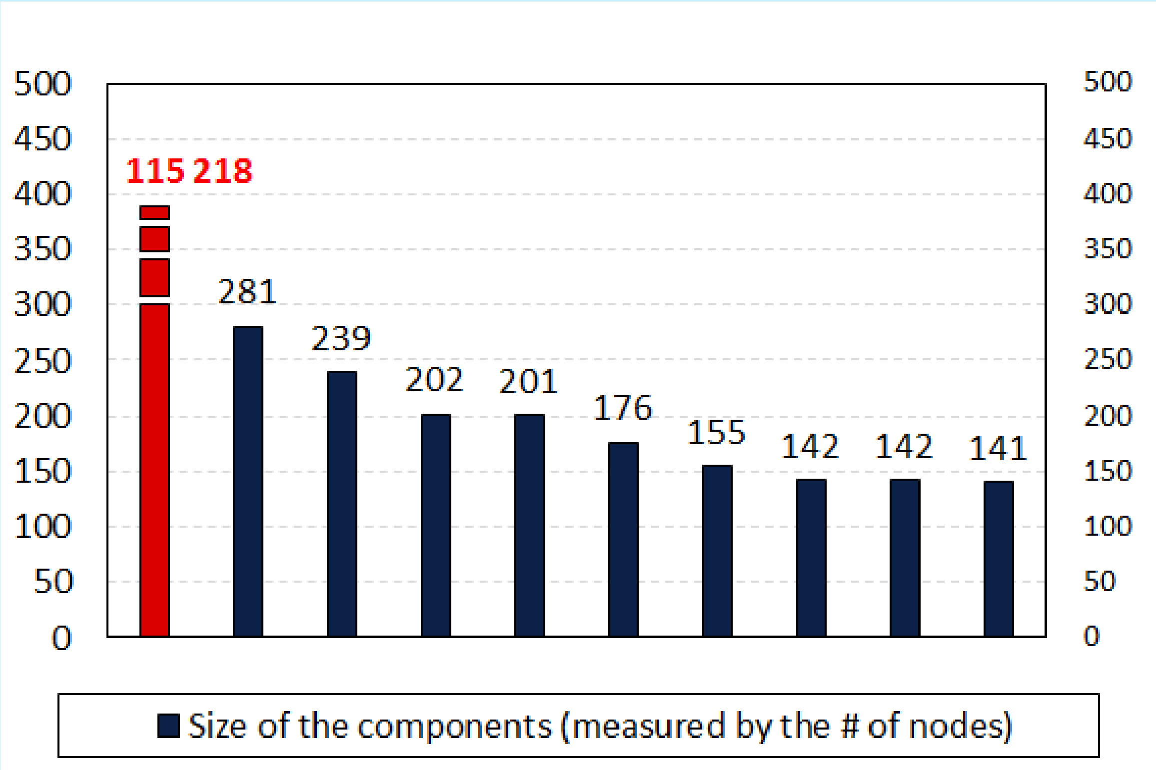
[10] If a firm operating in Hungary is a foreign-owned firm and it owns other firms, then foreign influence/control includes not only these owned firms but also the foreign-owned firm itself.

[11] The "Prof., sci., tech. activities" category refers to professional, scientific and technical activities, which contains legal, auditing, consulting services as well as scientific and technical (e.g. architectural) services.

Figure 1
Basic description of the ownership network

	Total network	Largest component
Number of nodes	1 029 487	115 218
Number of edges	963 744	253 840
Density	9e-7	1.9e-5
Average degree	1.86	4.4
Shortest path lengths (avg.)	-	13.78
Shortest path lengths (st.dev.)	-	5.91
Local clustering (avg.)	-	0.18
Local clustering (st.dev.)	-	0.2

Figure 2
Size distribution of the 10 largest components



Based on 2017 data.

Figure 3
Degree distribution of the 2017 ownership network's giant component

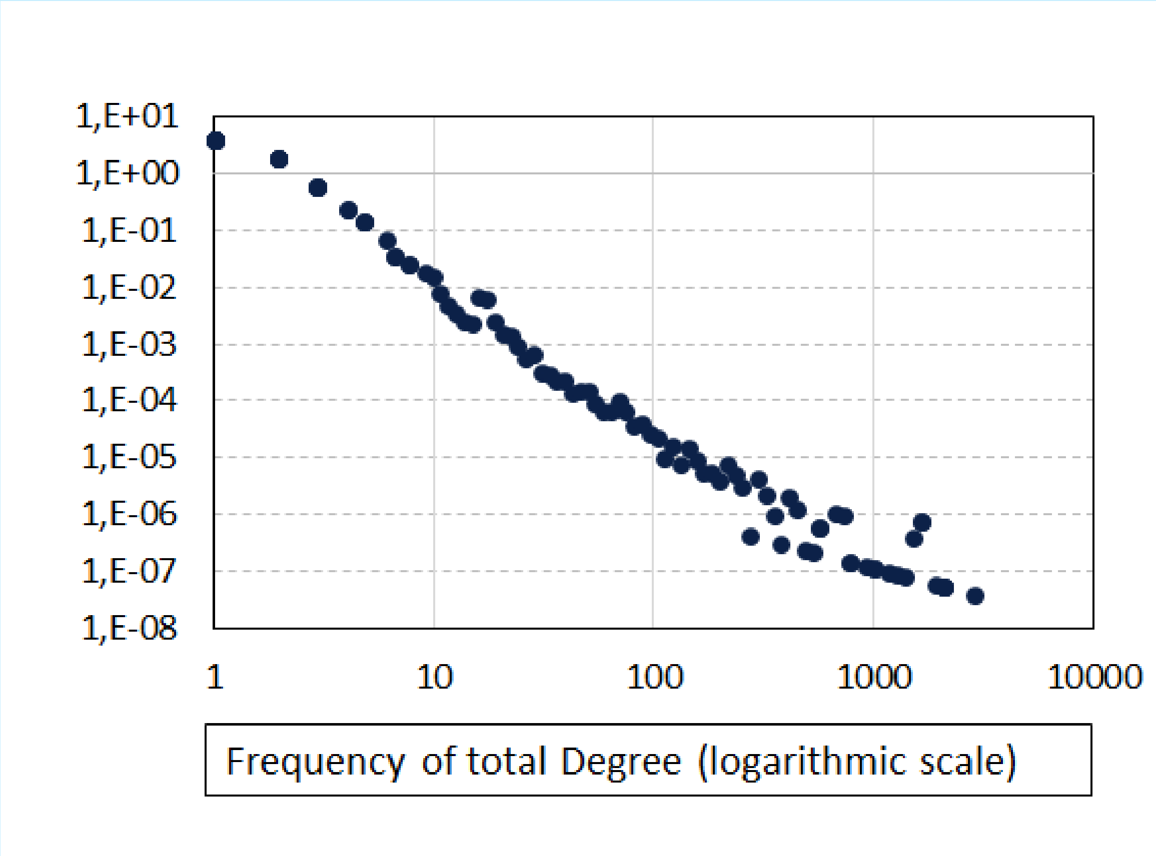
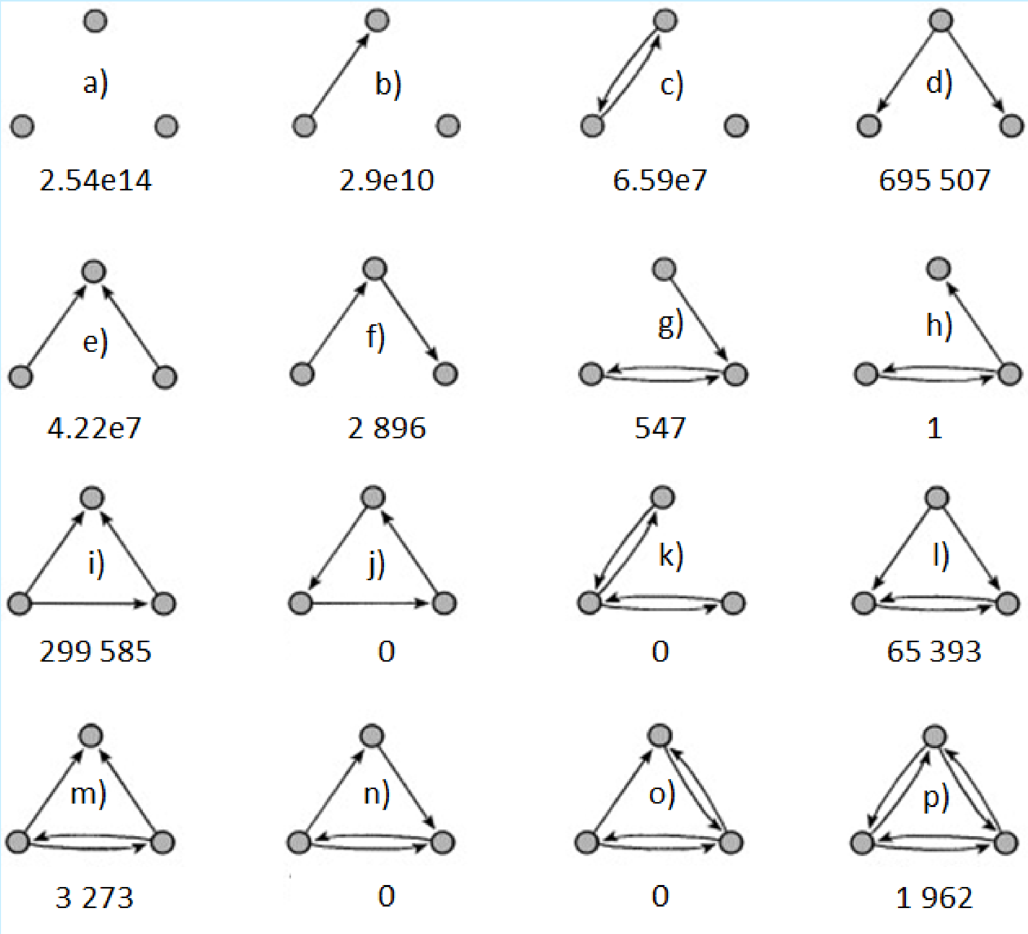
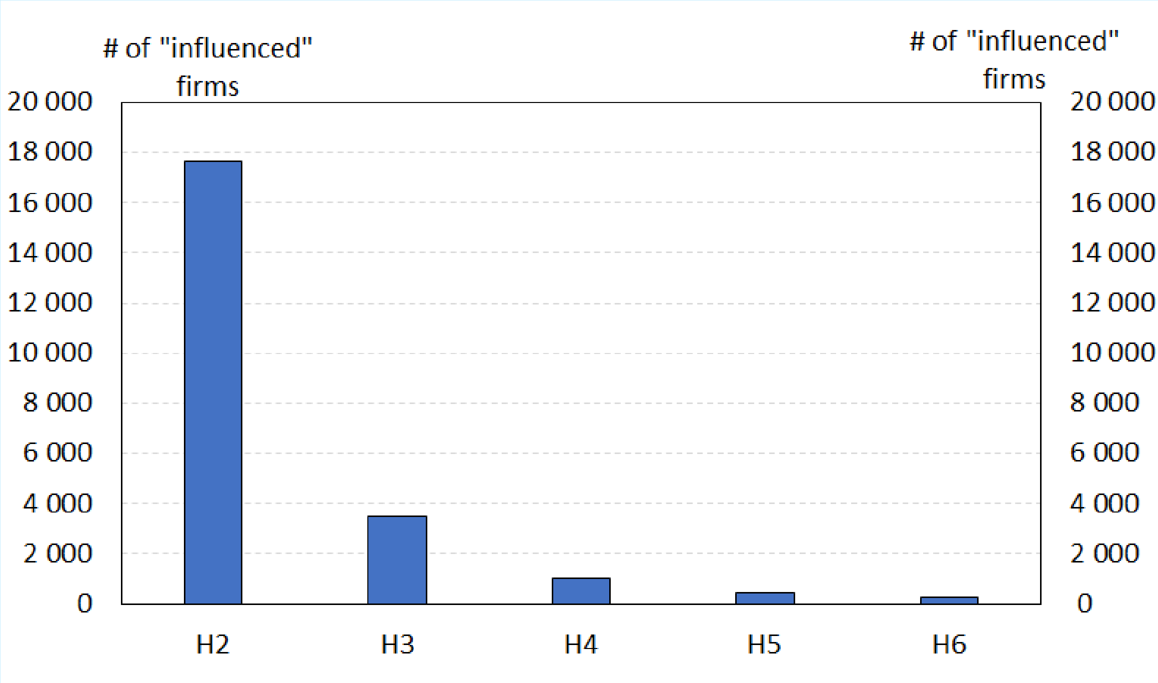


Figure 4
Motif statistics of the 2017 ownership network's giant component



Estimated on a sample.

Figure 5
Additional indirect influence by the number of steps in the ownership network.



H# refers to a given power of the original influence matrix. The "# of influenced firms" is equal to the sum of the elements of a given matrix.

Figure 6
Cumulative distribution of owners' total control in 2017

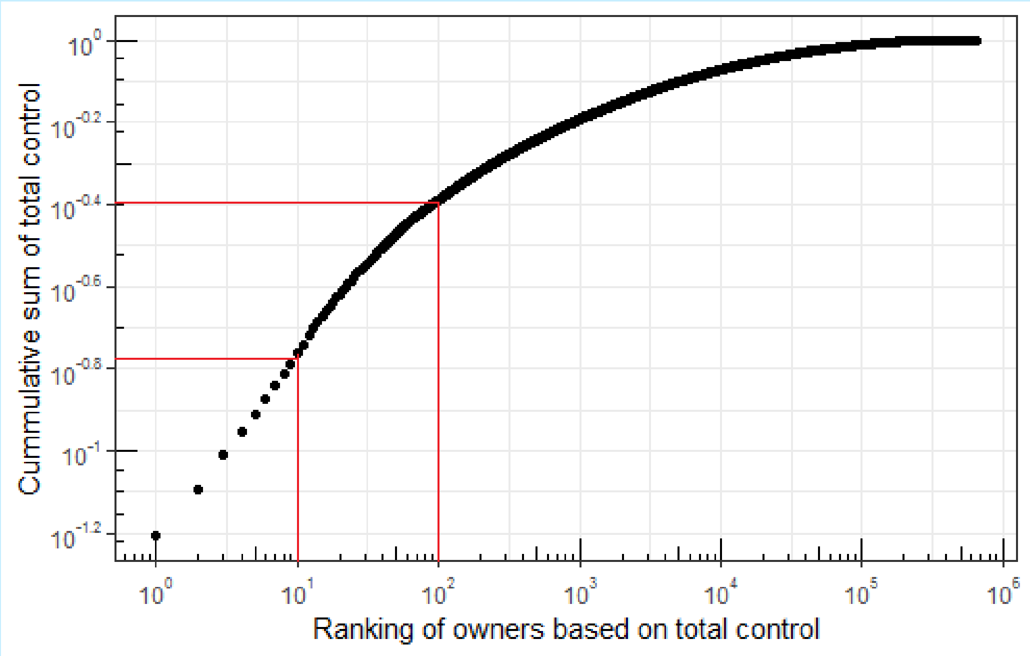


Figure 7
Total/direct control/influence of owners

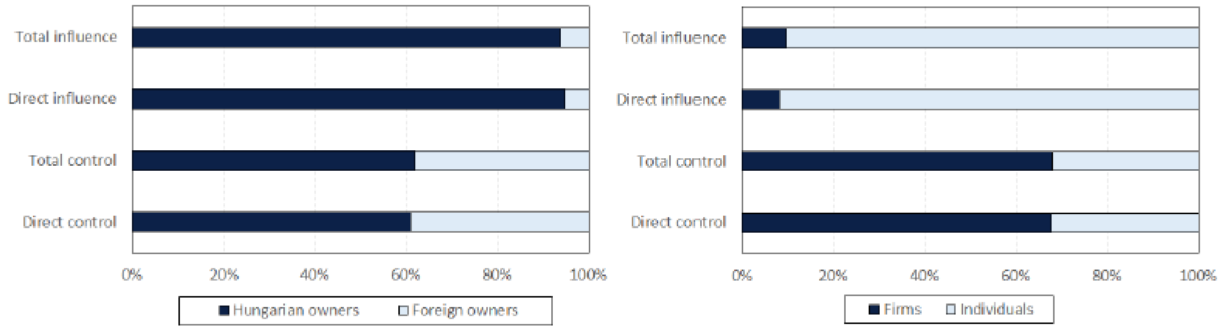
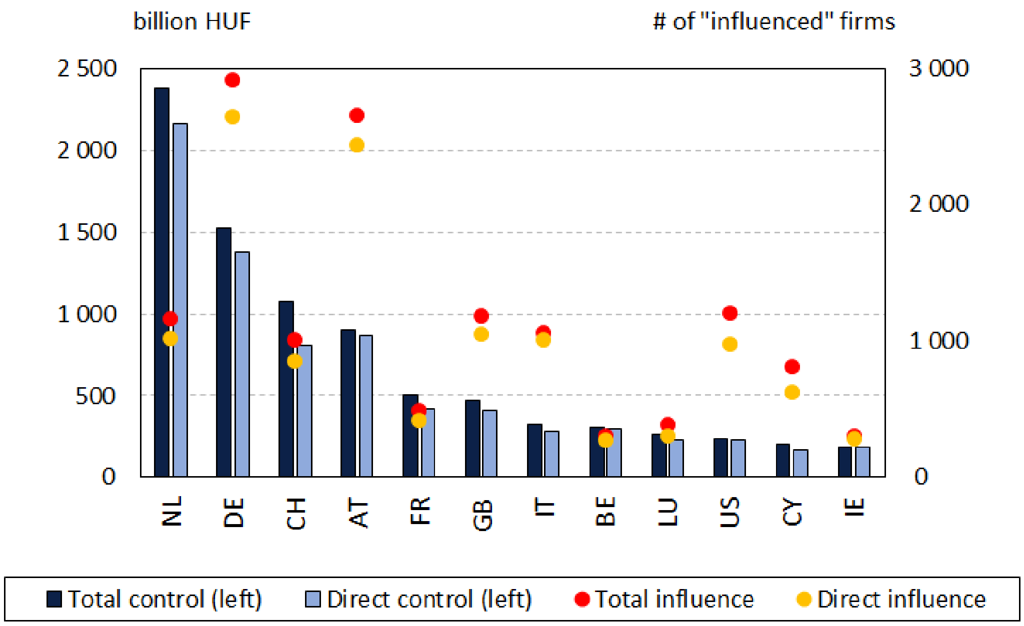


Figure 8
Total control (in 1000 billion HUF) based on the owners' attributes

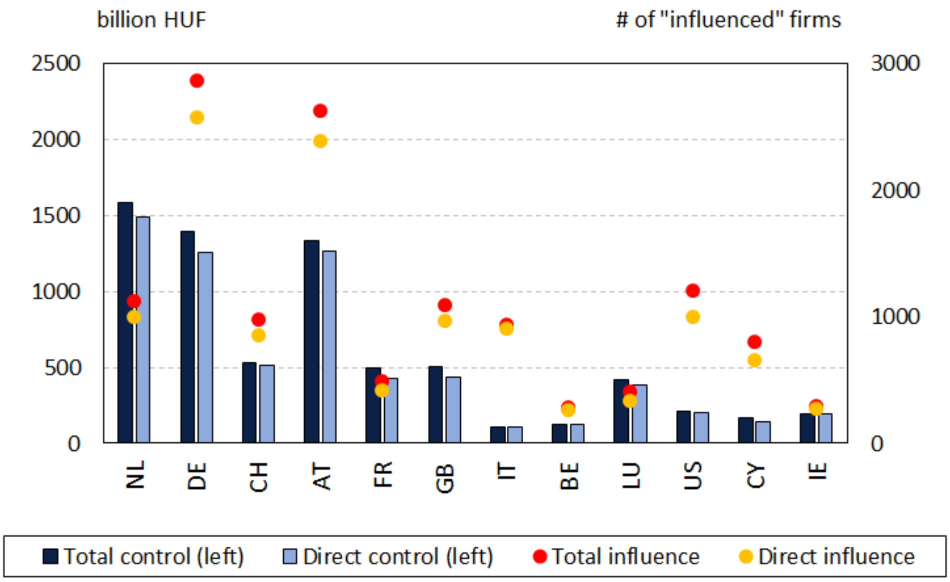
	Hungarian owners	Foreign owners	Sum
Firms	8.88 (31%)	10.58 (37%)	19.47 (68%)
Individuals	8.91 (31%)	0.26 (1%)	9.17 (32%)
Sum	17.79 (62%)	10.85 (38%)	28.65 (100%)

Figure 9
Total/direct control/influence of countries in 2017.



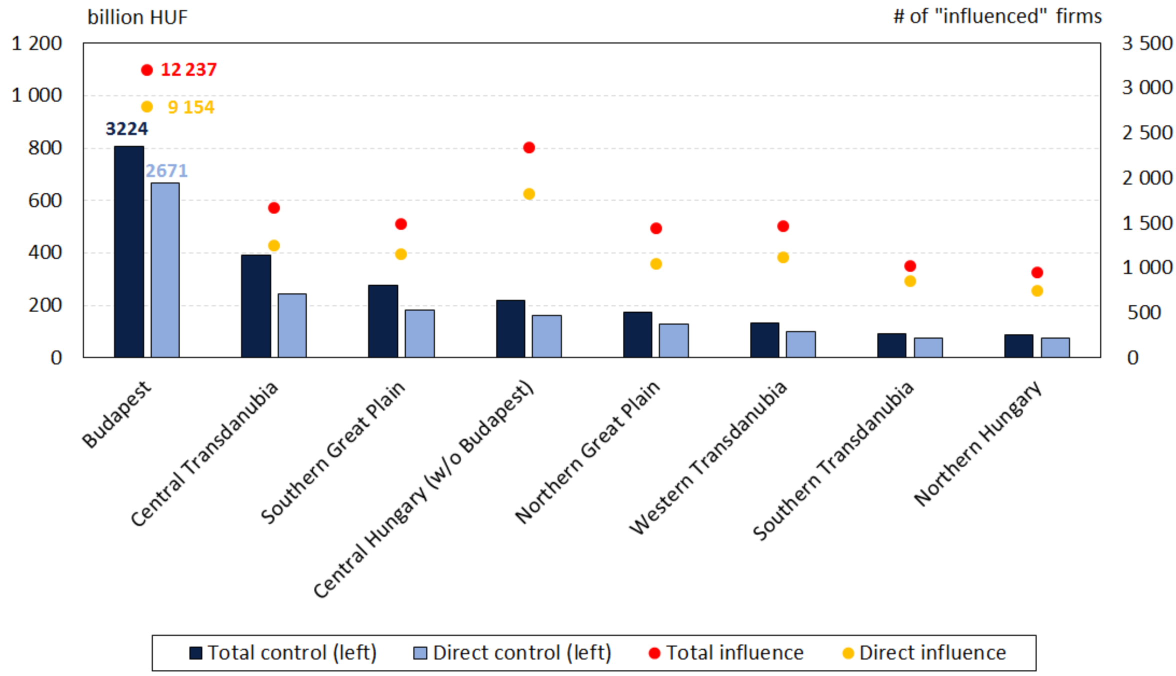
The "# of influenced firms" is equal to the sum of the total influence of owners belonging to a given country.

Figure 10
Total/direct control/influence of countries in 2016.



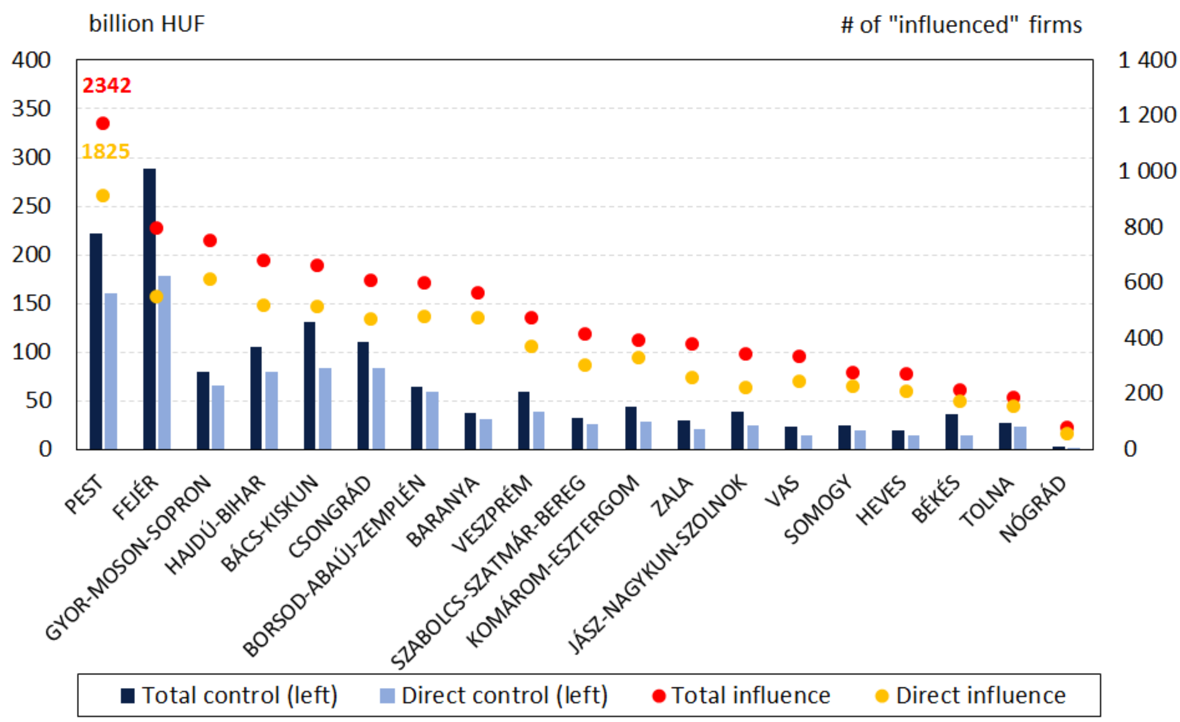
The "# of influenced firms" is equal to the sum of the total influence of owners belonging to a given country.

Figure 11
Total/direct control/influence of regions in 2017



The "# of influenced firms" is equal to the sum of the total influence of owners belonging to a given region.

Figure 12
Total/direct control/influence of counties in 2017

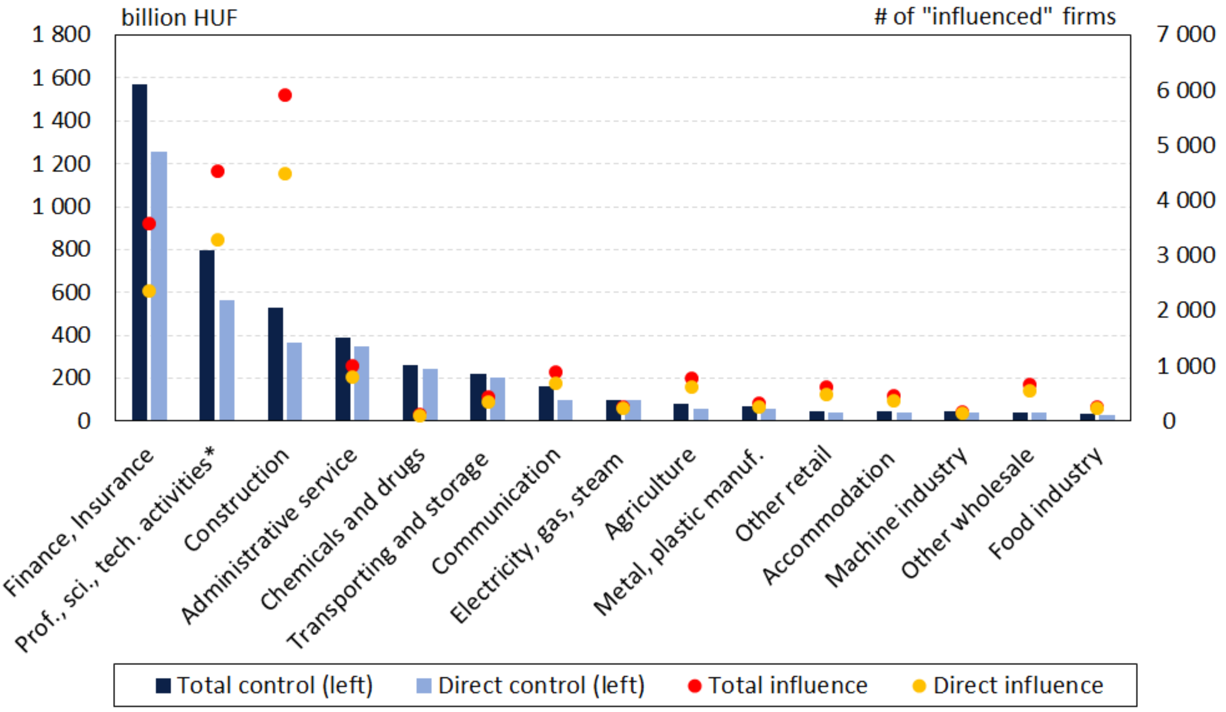


The "# of influenced firms" is equal to the sum of the total influence of owners belonging to a given county.

Figure 13
Total/direct control/influence based on size of the owner firm

Firm size category	Total control (billion HUF)	Direct control (billion HUF)	Total influence	Direct influence
Micro	2 090	1 418	16 690	12 279
Small	398	325	3 214	2 691
Medium	286	226	1 448	1 236
Large	1 052	956	719	545

Figure 14
Total/direct control/influence of the top 15 industries in 2017



The "# of influenced firms" is equal to the sum of the total influence of owners belonging to a given industry.

3 Supplier network

We obtained access to firm-level supplier connection data containing trade links among Hungarian firms where the tax content of the transactions between two firms exceeds EUR 3000 in the given year (Figure 15). This data is available between 2014-2017, and it is coming from Hungarian firms' VAT reports collected by the National Tax and Customs Administration of Hungary. Similarly to the ownership data, the most intuitive way to handle this network is to think about it as an *adjacency matrix* (A or in the case of weighted networks W), where each cell corresponds to the purchased value of the firm in the row dimension from the firm in the column dimension ($A_{i,j}$ or $W_{i,j}$). That is, an outward link starting from a given node denotes that it buys from the firm toward which the arrow points.

It is important to emphasize that this data is directly not comparable to the typical industry-level, symmetric input-output tables due to some fundamental conceptual differences. The firm level data contains information only about trade relationships above the regulatory threshold, and only for those products and services which are subject to the VAT. (Albeit it also ignores the reverse VAT situations.) However, the data includes transactions between firms which are not necessarily residents in Hungary if the fulfillment of the transaction happened in Hungary. In the I/O table of the Hungarian Central Statistical Office (HCSO) these trades often belong to the foreign trade category (although in some cases they would not be part of the I/O table statistics at all). Moreover, we cannot see in the granular data any further information about the products and services, thus, it is impossible to know if a purchase happened for investment reasons, which would be handled differently in the I/O table than a purchase due trade purposes. A further problem can arise when the invoice comes from a trader firm, in which cases the source of the product is unknown. If it was imported, then it will be categorized as foreign trade in the HCSO I/O tables. There are some differences between the industry classifications as well, and also in the calculation of the prices. The HCSO I/O table uses basic prices, which are different from the market price as they do not include margins, transportation costs and the net position of the taxes and allowances. Due to all these factors, the industry-level aggregation of our granular data results in a completely different table than the I/O matrix produced by the HCSO.

3.1 PREPARATION OF THE SUPPLIER NETWORK DATA

The links of the network change significantly from one year to another because there are a lot of one-off, incidental transactions. More than 50% of the links disappear between the observed periods, and new links emerge in a similar extent. As these relationships are not particularly relevant from the point of view of spreading processes and they increase the noise in our measurements, we filtered the network to contain only long-term supplier connections. We consider a link *long-term connection* (i) if there were at least two transactions between the parties, and (ii) if there is at least one quarter time difference between the first and the last transaction between the two firms (Figure 16). Even with these mild requirements, only 54% of the links are long-term, however, these cover 93% of the aggregate trade volume in the network.

Another source of distortion we have to deal with is that there is no general rule for VAT reporting in the case of firms belonging to the same ownership-based group. Sometimes they file their VAT reports collectively, but as it is only optional, there are many firms belonging to a group which report individually. To handle this difficulty we can utilize the ownership layer of the firm connections by applying the following procedure[12]:

- In the case of every ownership link when the *total influence* exceeds 50%, we combined the link's endpoints into a group.
- If some firms in the supplier network belonged to the same group, we replaced them by a new node representing the group.
- We added the links of the original firms to the new "group" node.
- We eliminated the resulting self-loops (within-group links) (Figure 17).

[12] Although we did not consider ownership links with influence weights under the 50% threshold, and we cannot see global ultimate beneficiary owners, we could still cover probably the vast majority of the ordinary intertwinings among firms.

Further details about the features, quality and cleaning of the data can be found in Appendix C.

3.2 TOPOLOGICAL ANALYSIS OF THE SUPPLIER NETWORK

Similarly to the section about the ownership network, we present only the results of one year (2017) as it is sufficiently representative for the whole examined period. (The basic description of the network in other years can be found in Appendix C.)

As this network consists of Hungarian firms only, the number of nodes is much lower than in the case of the ownership network which contained individuals and foreign actors as well. However, the density of this network layer is higher (even after filtering for long-term connections and correcting for ownership groups), which contributes to the emergence of a giant component covering around 94% of the nodes. In this case we decided to focus only on this component as it represents credibly the whole network while all the other isolated parts are negligible in size.

The resulting network consists of 89 778 nodes and 235 913 links. The average total degree in this giant component is around 5.2 (which implies that the average in/out-degree is the half of this value, 2.6), which can be interpreted as the average number of long-term supplier and buyer relationships for firms (Figure 18). This result is difficult to compare to any other datasets in the literature, as other papers usually consider all the transactions (with different thresholds and without filtering for long-term links) among (often only a sample of larger) firms.

Regarding the degree distribution of the graph we can see slightly different figures depending on whether we consider the *total, in-, or outdegree* of the nodes. For all the three measures the distributions have fat tails, however, we do not encounter as many extreme values in the case of the number of suppliers of firms (*outdegree*) as in the case of the number of buyers (*indegree*) (Figure 19). Despite of this disparity, we could ascertain that there are firms in the network which can be considered *hubs*. These agents can play a special role in any spreading process for which the production network is a relevant medium, thus, the identification of these firms and the assessment of their importance can be vastly important.

We can also see that the average shortest path length[13] is below 5 with a standard error as low as 1.1, which implies that shocks can be transmitted easily between firms via hubs in the network. (For further investigation on this see section 3.3.) Furthermore, Table 4 also shows measures of local feedback loops: there is an unexpectedly high probability of reciprocal dyads (11%) [14] and closed triplets (7.8%), which can further amplify shocks and promote contagions.

As it can be seen on Figure 20, we assessed the frequency of the different triadic motifs also in the supplier network. In line with the observed difference in the distribution of in- and outdegrees, we can see based on motifs d) and e) that there are much more instances of firms having multiple buyers than having multiple suppliers. While in the sample there are only a few observations of the simplest triadic loop formation shown by motif j), there are several appearances of loops hidden in the more convoluted motifs, such as n), o) and p). These more complicated motifs are often observed e.g. among wholesale trader firms and manufacturers operating on the upstream part of supply chains. (E.g. two wholesale trader of chemical materials are trading with each other to both directions, and both of them are connected to a chemical material producer.) Based on this, the network is far from being acyclic which observation can be in juxtaposition with the results of McNerney (2009), in which paper the authors examined input-output economic systems, and found that economies tend to be acyclic at the scale of triadic patterns on industry level.

3.3 SUPPLIER CONNECTIONS BASED ON FIRM CHARACTERISTICS

By connecting this data to the locations, balance sheets and profit and loss statements of firms we could carry out analysis not only based on topological information but also using several node and link attributes.[15] We examined several firm characteristics, but we present here only those cases where we found meaningful patterns.

[13] The average shortest path length is the average number of steps along the shortest paths for all possible pairs of network nodes

[14] In the case os reciprocity we cannot filter for repurchases, which can somewhat inflate this measure.

[15] As some of these characteristics of the firms are not trivial to consolidate based on their ownership background, in this section we considered the firms as they were present in the data originally.

We examined the supplier relationship between firms based on their (labor) productivity. We assigned each firm into groups formed based on productivity deciles, and then we collapsed the network to this aggregation level. This way, the resulting 10 by 10 matrix shows the flows in the supplier network among firms grouped based on they labor productivity. However, although the number of firms will be the same in every group, the size of the firms, and consequently the number (and weight) of their supplier connections can be different. To control for this effect, we could compare the observed flows to a null model which gives us the expected flow between any two groups if the links were formed randomly in the network. To perform this calculation we divided the observed flows by the product of the *outdegree* of the group in the row dimension and the *indegree* of the group in the column dimension (or in the case of weighted networks we can use the *strengths*)[16]:

$$\widetilde{W_{p,q}} = \frac{W_{p,q}}{S_p^{out} S_q^{in}} \tag{9}$$

where $W_{p,q}$ denotes the original and $\widetilde{W_{p,q}}$ denotes the normalized flow between productivity groups p and q. S_p^{out} is the sum of the strengths of outgoing links for group p and S_q^{in} is the sum of the strengths of incoming links for group q.

Figure 21 shows some homophily based on productivity. The cells near the lower-left corner, but at some level also near the upper-right corner are darker, indicating stronger linkages between firms with similar productivity levels. These observations can have several connections to the existing literature, however, in this paper we do not try to identify the factors leading to this pattern or assess the potential consequences of this network structure.

Trade connections between groups with very different productivity are also very polarized: Productive firms sell much more to the less productive firms than the other way around. We quantified this polarization between the groups using the following formula which is based on Iino and Iyetomi (2012):

$$P_{p,q} = \frac{W_{p,q} - W_{q,p}}{W_{p,q} + W_{q,p}} \tag{10}$$

where $P_{p,q}$ denotes the polarization ratio.

This formulation of the polarization shows the typical direction of trade between groups in the row and in the column dimensions. The polarization matrix is antisymmetric, i.e. $P_{p,q} = -P_{q,p}$. If the relationship between groups p and q is only one-directional, $P_{p,q}$ will be ± 1 (the sign depends on the direction), while if the flow is the same to both directions, $P_{p,q} = 0$. On Figure 22 we can see that the more productive a group is the larger its dominance is in the trade relationships with less productive groups.

On Figure 23, we can see a weaker but similar pattern if we use return of assets (ROA) instead of labor productivity (as it can be expected due to the high correlation between productivity and profitability measures).

We can calculate similar measures to assess geographical clustering as well (Figure 24). The diagonal elements of the matrix are clearly darker than the off-diagonals, which shows that trade connection within regions are stronger than between regions, indicating the presence of location-based preferences in link formation.

As in this case it would be also informative to see the absolute magnitude of the trade connections (i.e. without comparing them to a null model) to assess the dominance of different regions, we can simply consider the number of links between the different regions as well. According to Figure 25 the dominance of Budapest is apparent: Firms in the capital have way more connection than in any other region, and they have a lot of trade links to all the other regions as well. Furthermore, firms in Budapest typically supplied more to firms in other regions than the other way around.

3.4 EXPLORING THE COMMUNITY STRUCTURE OF THE SUPPLIER NETWORK

An often observed characteristics of real-world social and economic networks is that they have a mesoscopic structure which can be best described by the concept of *communities*. A network is regarded to have a community structure if its nodes can

[16] This normalization is in the spirit of the *configuration model* which generates uncorrelated random networks with a given degree sequence.

be grouped into internally densely connected sets (which potentially overlap), i.e. members of a community are relatively densely connected within their group, but there are only sparser connections between the groups Yang et al. (2010) (Figure 26). Identifying community structures can be very revealing about a complex system, as the observation of the grouping of the actors based on this dimension is only possible through the examination of the whole network on granular level.

In the case of our production network, community detection can result in the identification of blocks within the economy, in which the coherence is provided by the intricate supplier relations among the constituent firms. As local shocks usually propagate more unimpededly within the surrounding community than between the separated communities, this segmentation of the network can help us tremendously in the more detailed understanding of spreading processes on the supplier network.

There are many algorithms (coming from different disciplines e.g. computer science, biology, mathematics, physics and sociology) which have been developed for identifying communities. (An excellent review of community detection algorithms can be found in Javed et al. (2018).) In our analysis we opted to use a widely-used technique called the ”Louvain-method” which is based on the *modularity* of the network Newman (2006). As these methodologies are not common in economics, we describe the applied algorithms in more details in Appendix D.

3.4.1 THE COMMUNITY STRUCTURE OF THE HUNGARIAN SUPPLIER NETWORK

Our procedure detected 249 communities, however, the ten largest already contain almost 80% of the firms, so we concentrate only on these in the more detailed analysis. The groups formed based on the community detection results are much more separated than in the case of any of the former grouping variables. Figure 27 shows that the diagonal of the matrix is clearly outstanding compared to the other cells indicating that connections within communities are much stronger than connections between communities.

Similarly to the results of Fujiwara and Aoyama (2010), the most intuitive variable we can use to interpret the communities is their sectoral composition. The largest communities all can be interpreted as a production chain of certain product categories within the economy. E.g. the first group on Figure 28 consists of firms belonging mainly to the food industry, food wholesale and food retail sectors. The second group contains firms from the machine and electronics industry; metal and plastic manufacturers; electricity, gas and steam suppliers. All the other groups can be similarly well interpreted as blocks containing chains of production of a well defined product category.

If we examine the polarization on Figure 29, we can see clear patterns only for two out of the ten largest groups. In the case of Group B (which corresponds to the machine and electronics production chain) we can see that they buy a lot of intermediate inputs from other blocks, but they do not supply to them in similar extent. In the case of Group J (which corresponds mainly to logistics, insurance and motor vehicle retail) the polarization is exactly the opposite: this block supplies to all the other segments of the economy way more than it buys from them as inputs. A natural explanation contributing to these result can be the unobserved export and import activities of firms in these segments of the economy.

3.4.2 BRIDGES BETWEEN COMMUNITIES OF THE SUPPLIER NETWORK

Although we saw that the communities are highly separated, from the point of view of shock propagation it is still crucial to examine how these large blocks of the economy are connected to each other. Firms having supplier partners belonging to other communities create bridges between distant parts of the network, and therefore, propagate the spreading of contagious processes in the whole system.

The simplest measure to capture firms’ shock transmitting ability is to consider the number of links of a node which are pointing to other communities. As it is shown on Figure 30, although the degree of a node is correlated with the number of links pointing to other communities, there is still a large variation which is not explained just by the degree. To further investigate the firm characteristics associated with our dependant variable, we used a simple regression analysis. We found that firms in the transportation and infrastructure sectors, and firms with high value added and high export sales rate have particularly many outside connections (Figure 32).

We considered another approach as well to assess firms’ intercommunity shock spreading ability. Our community detection methodology was so far incapable to identify overlaps between the communities. However, using a different representation of our network based on Evans and Lambiotte (2010) and Ahn et al. (2010) we are able to take into account this feature as well. This would make it possible to see if a firm is part of multiple communities which would be an indication for its increased ability to transmit shocks.

As a first step to obtain this measure, we have to transform our graph into a so called *line graph*. Nodes of the *line graph* are the links of the original graph, and two nodes (links) are connected if they had a shared endpoint(in the original graph where they were links) (Figure 31).

Using our standard community detection method on the line graph we can assign the links of the original graph into communities. As links of a node in the original graph can belong to more than one communities now, we can count for every node how many communities its links belong to. We can use this number of group memberships of a firm to measure its potential to transmit shocks.

By putting this measure into the same specification as before, we got somewhat different results compared to the first regression (Figure 32). Based on the overlapping community approach, the firms which are typically associated with multiple groups often belong to sectors which provide products and services not directly related to the production activity of their buyers (e.g. insurance, catering, administration, etc.). This result suggests that these firms are not necessarily very influential in spreading shocks, which finding might suggest a more optimistic interpretation about the resilience of the economy: Although the different blocks of the firm network are accessible to each other, but the firms which are truly influential from the point of view of shock propagation are usually only connected to a few communities.

Since there are many firms without any connections outside of its community, we also considered the possibility that the excess zeros in the distribution are generated separately from the data generating process of the count values. As our dependent variables are heavily dispersed, we used a zero-inflated negative binomial regression to explore this alternative approach. Figure 33 shows the results for the count model. When we consider the model with the number of links pointing to other communities as the dependent variable we can see that the firm size and the export activity are the most influential factors, however, if we use the number of group memberships, also firms operating in wholesale industries seem to have a higher level of embeddedness in other communities.

Regarding the process governing the presence of the excess zeros, in the case of the number of outside links the degree and the export activity of firms are negatively associated with the probability of excess zeros, while operating in the food, agriculture, household goods retail and household goods wholesale industries have positive coefficients. In the case of the number of community memberships, the firm size and the variables indicating the geographic region of firms seem to have larger influence on predicting excess zeros. (Figure 34)

Figure 15
Supplier connections among Hungarian firms in 2017

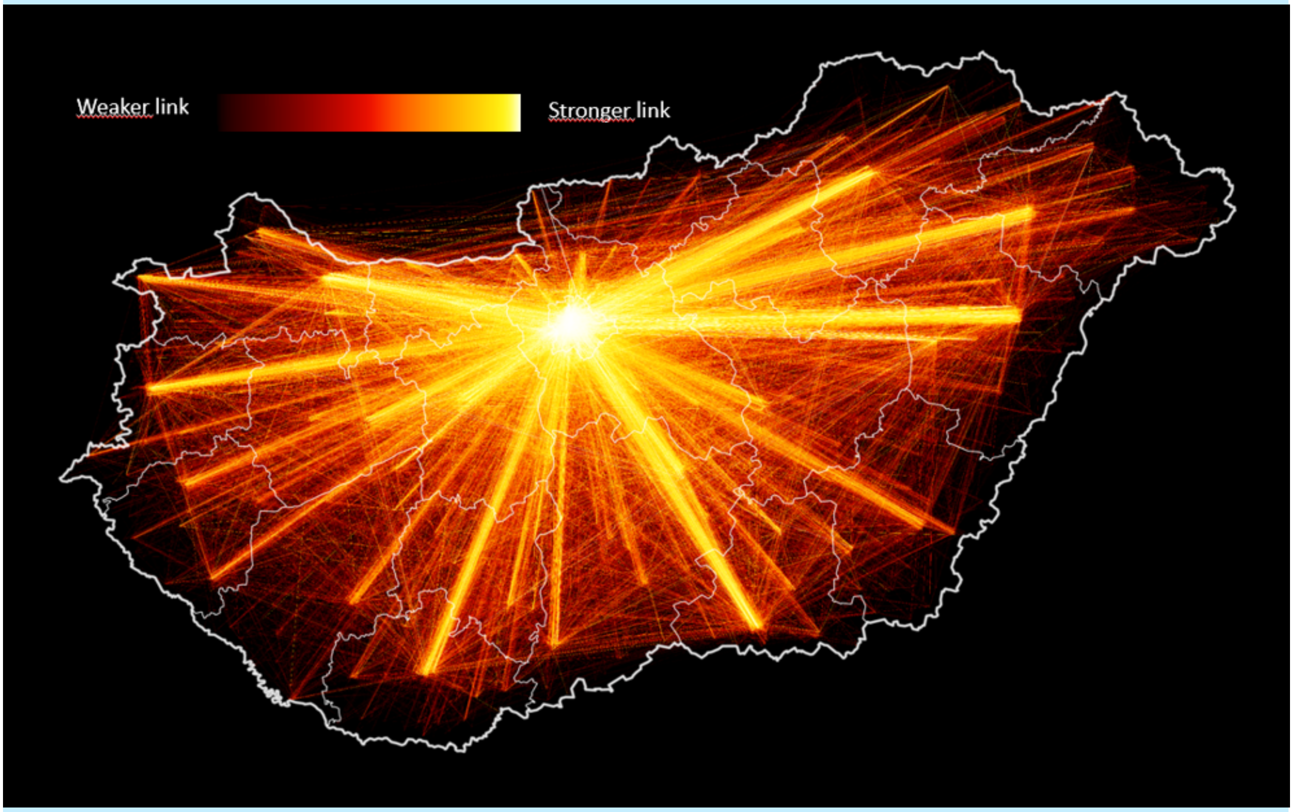


Figure 16
Temporal stability of the supplier connections

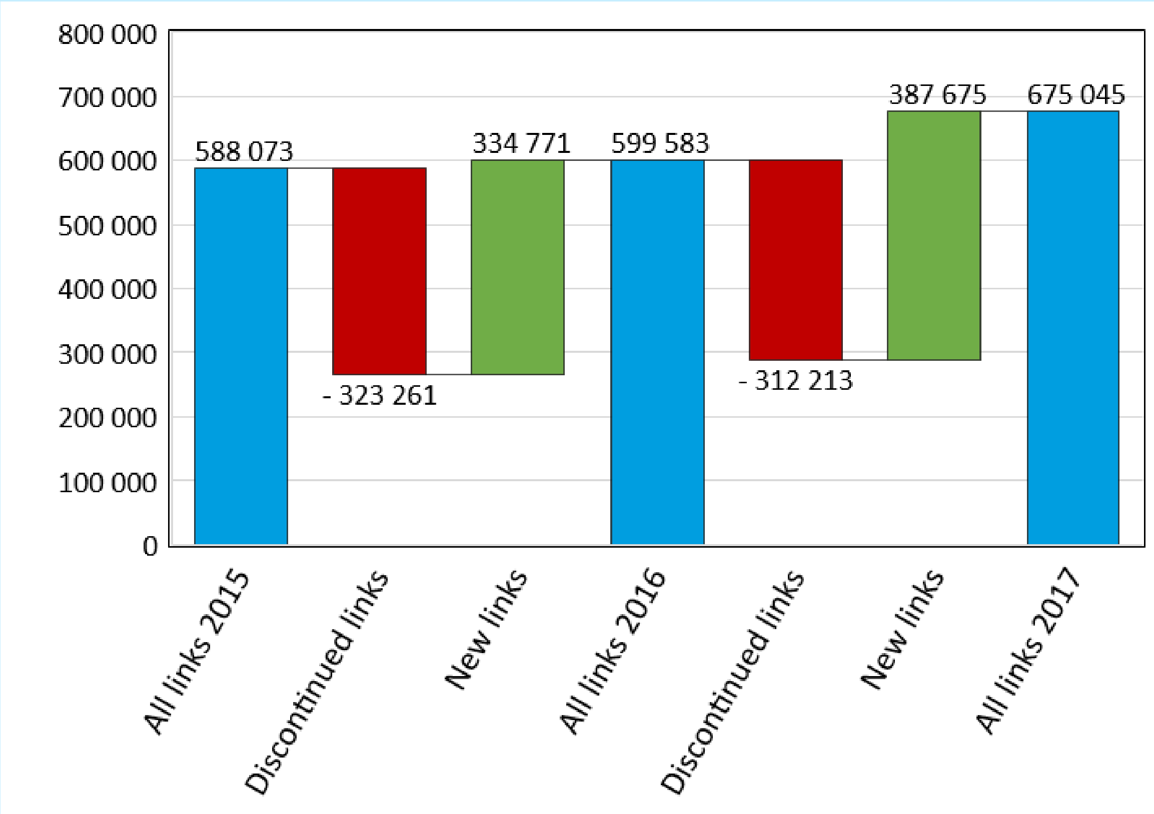


Figure 17
During the correction of the supplier network we combined together firms belonging to the same ownership group, and we eliminated the links within the groups.

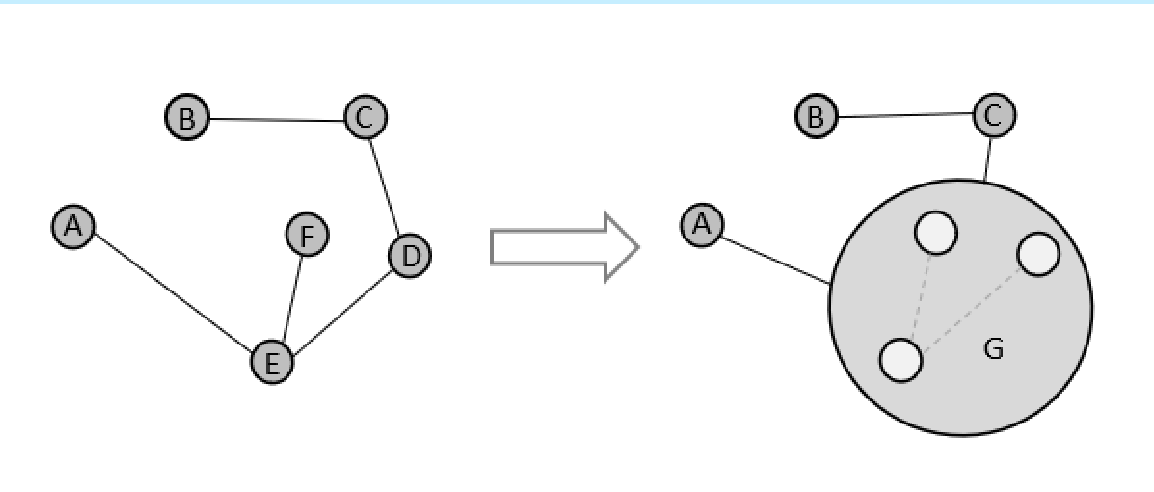


Figure 18
Basic description of the 2017 supplier network's giant component

Number of nodes	89 778
Number of edges	235 913
Density	2.93e-5
Average in-/outdegree	2.63
Shortest path lengths (avg.)	4.92
Shortest path lengths (st.dev.)	1.1
Local clustering (avg.)	0.078
Reciprocity	0.11

Figure 19
Degree distribution of the supplier network in 2017

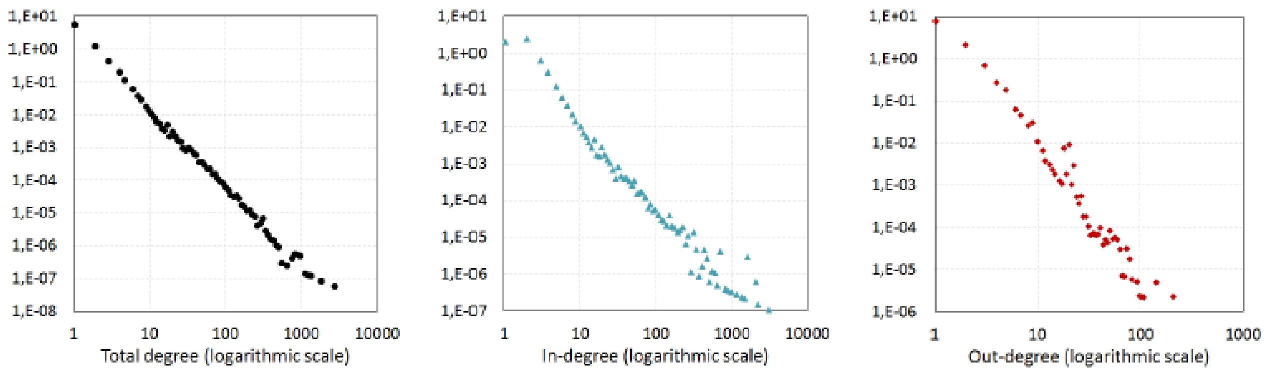


Figure 15: Degree distribution of the supplier network in 2017

Figure 20
Motif statistics of the 2017 supplier network's giant component

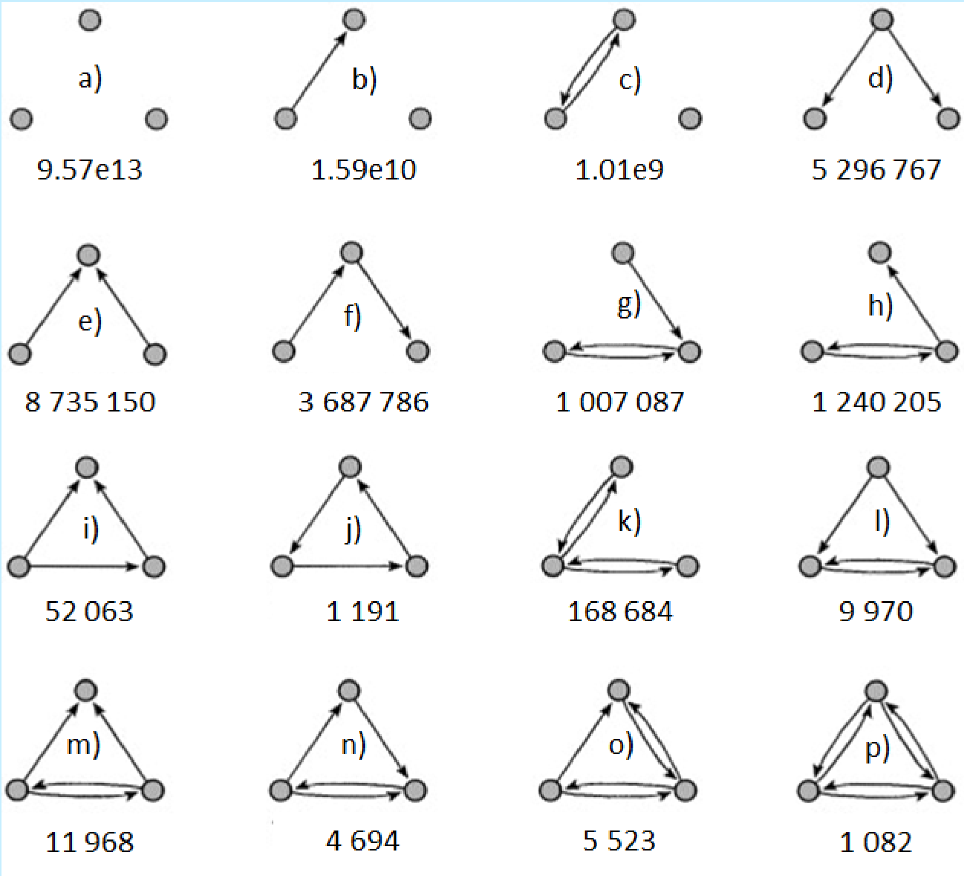
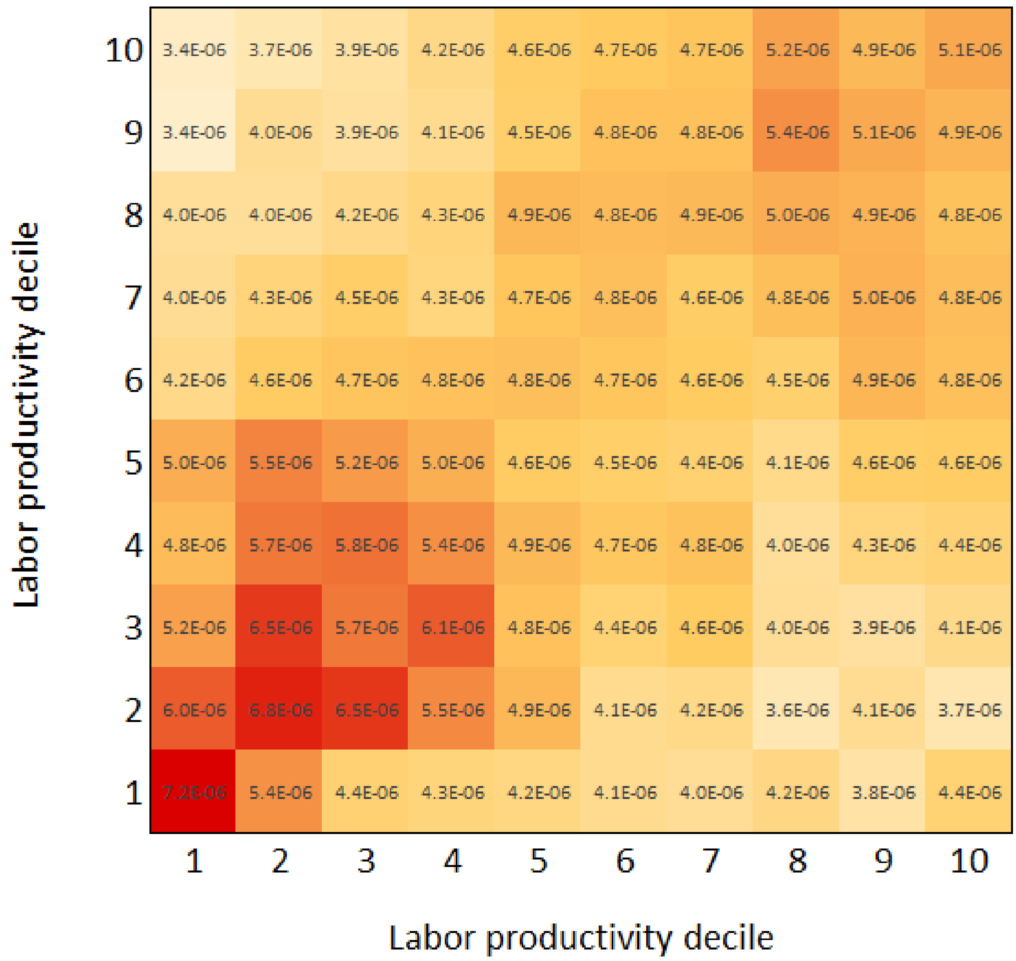
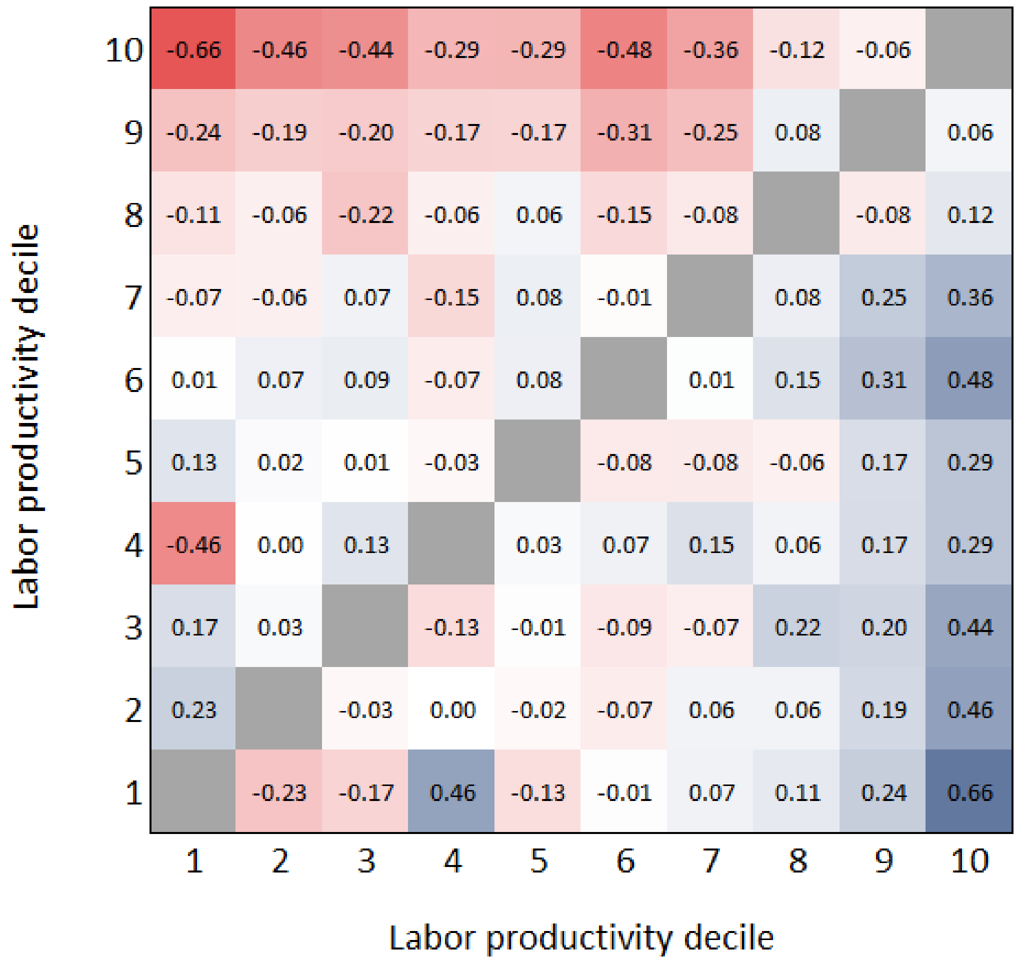


Figure 21
Flows in the supplier network between firms belonging to different productivity-based deciles in 2017



Labor productivity increases from group 1 to 10. Darker coloring indicates stronger trade connection.

Figure 22
Polarization among firms in different labor productivity deciles in 2017



Labor productivity increases from group 1 to 10. Red color means that typically the group in the row dimension supplied to the group in the column dimension, while blue indicates the inverse situation. Darker coloring indicates stronger polarization.

Figure 23
Flows and polarization in the supplier network between firms belonging to different ROA-based deciles in 2017

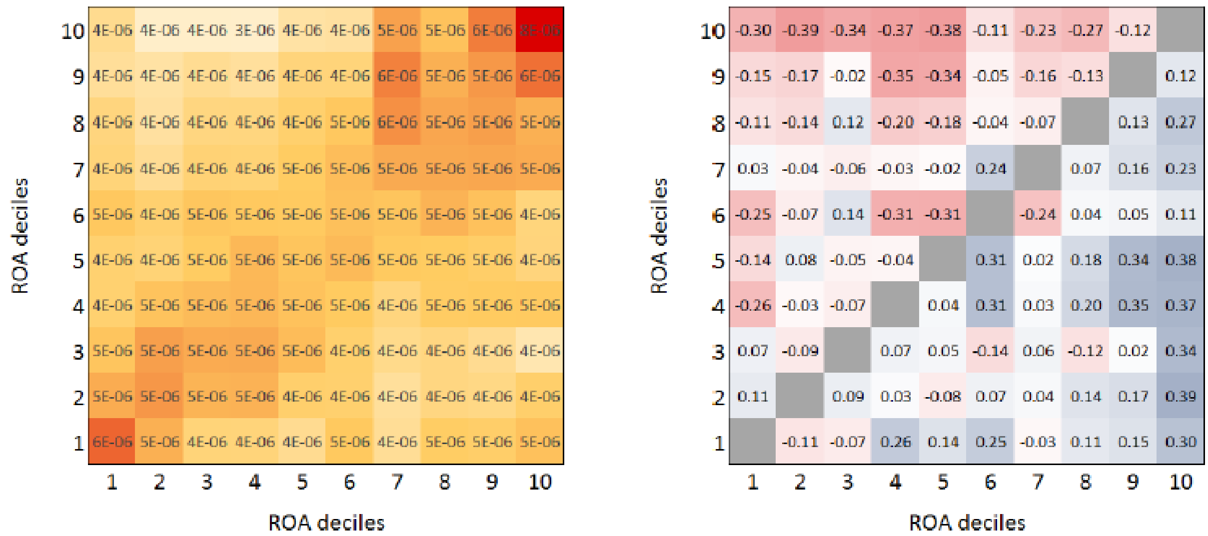


Figure 24
Flows and polarization in the supplier network between firms in different regions in 2017

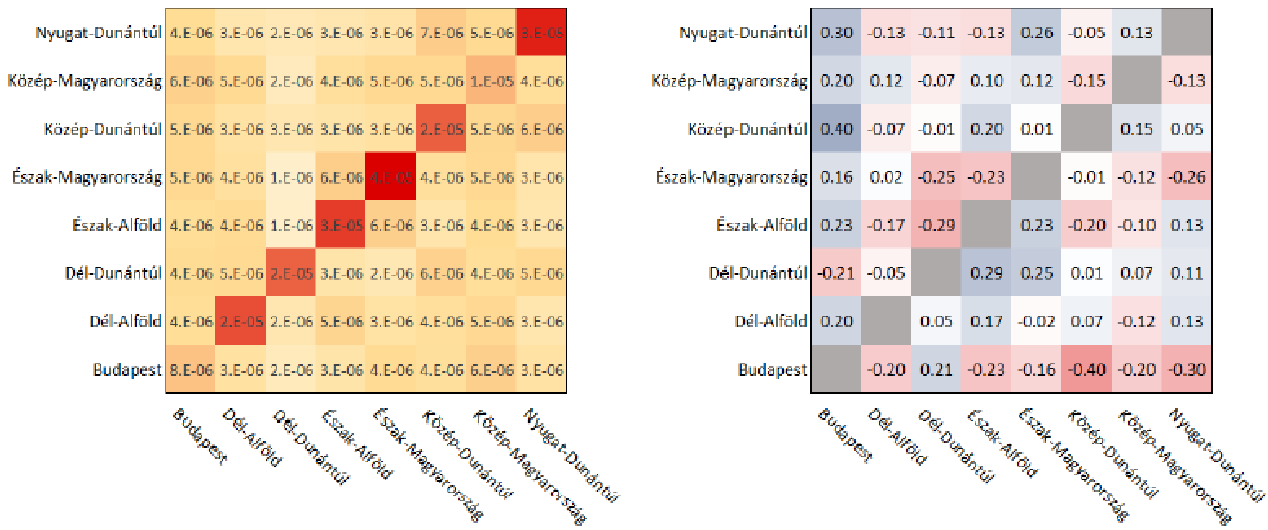
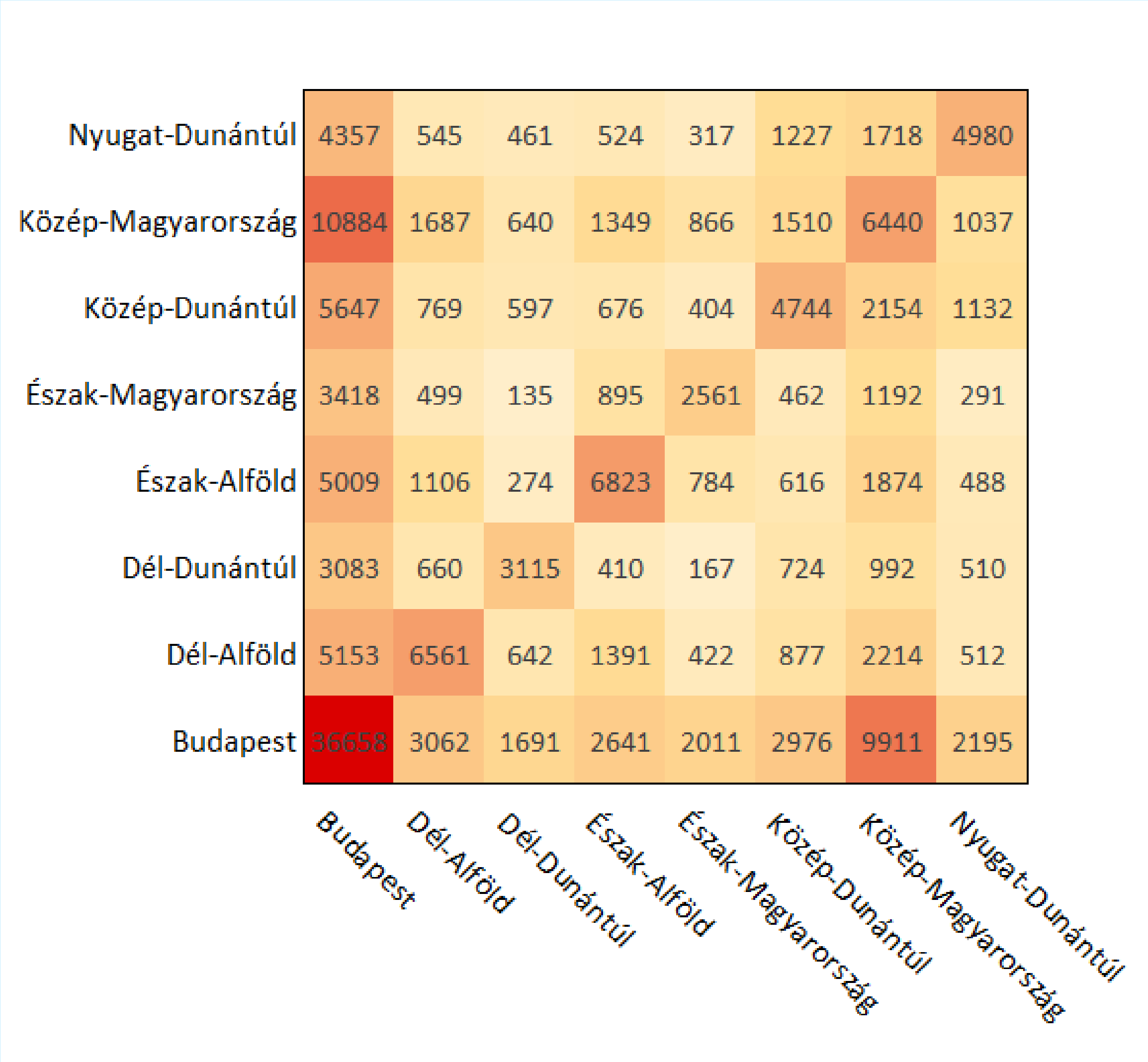


Figure 25
Trade connections in the supplier network between firms in different regions in 2017



Darker coloring indicates stronger trade connection.

Figure 26
Schematic picture of the community structure of a small network

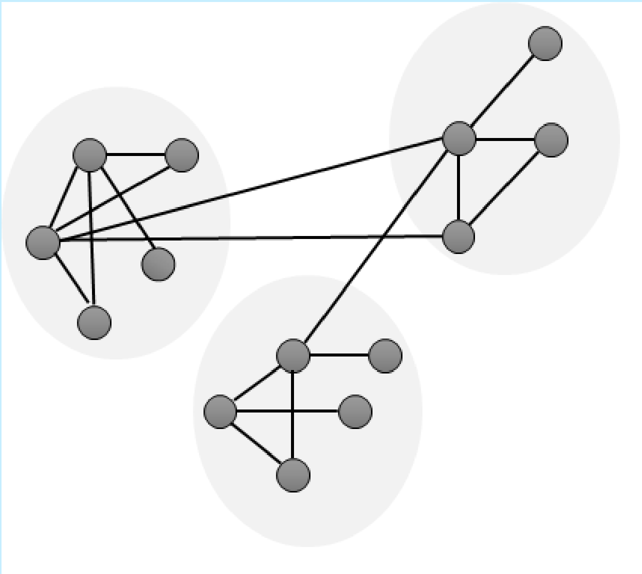
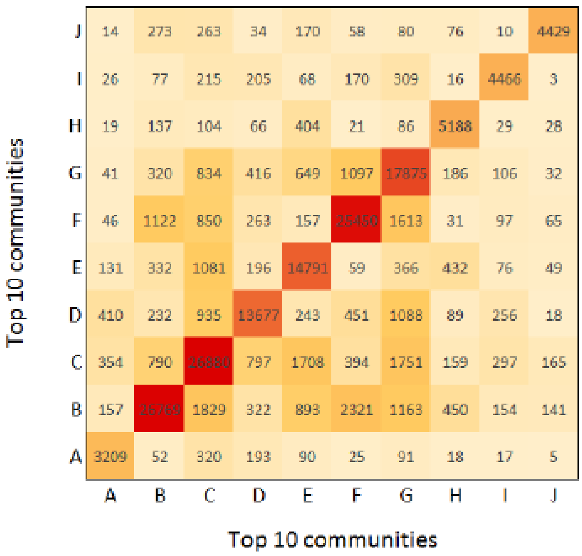
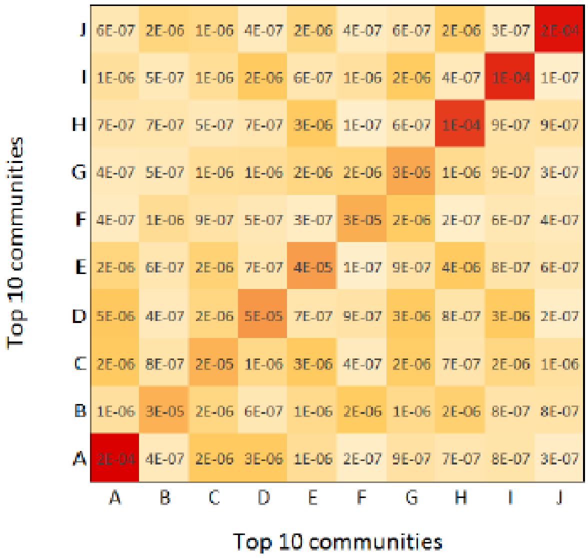


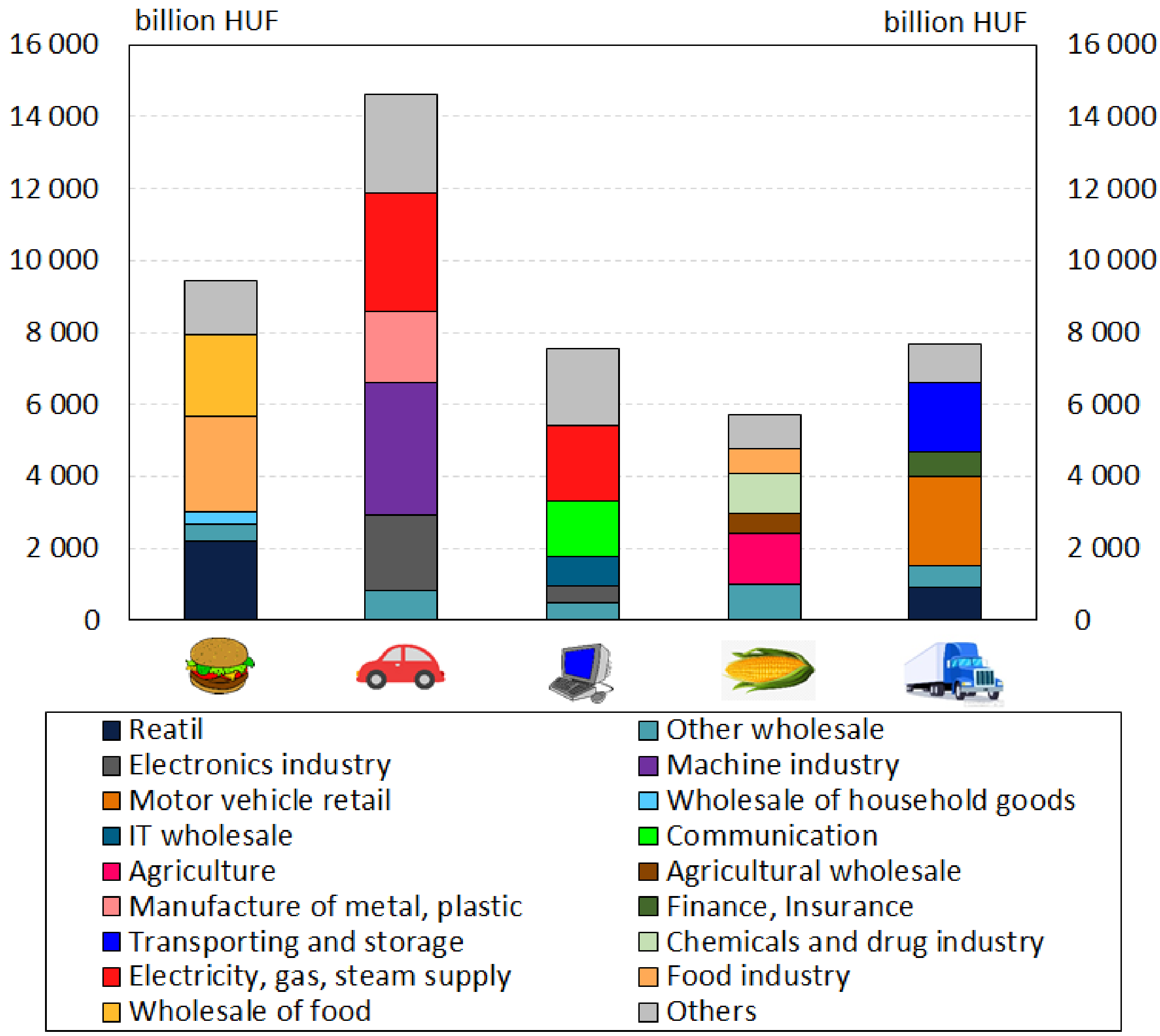
Figure 27
Strength of connection between communities in the supplier network in 2017





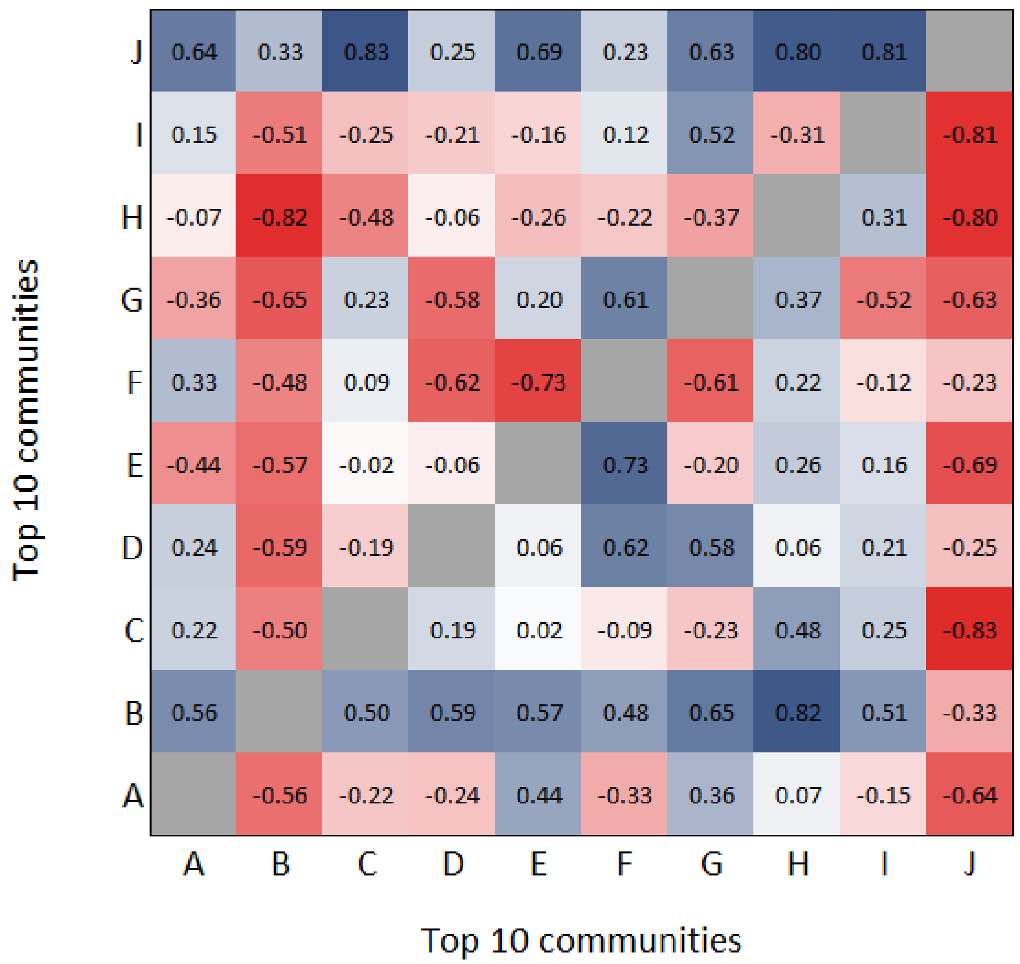
The figure on the left is based on the number of links between communities, while the figure on the right shows the connections compared to the randomly expected number of links between the groups.

Figure 28
Industrial composition based on the size (total assets) of the firms belonging to the top 5 communities in 2017



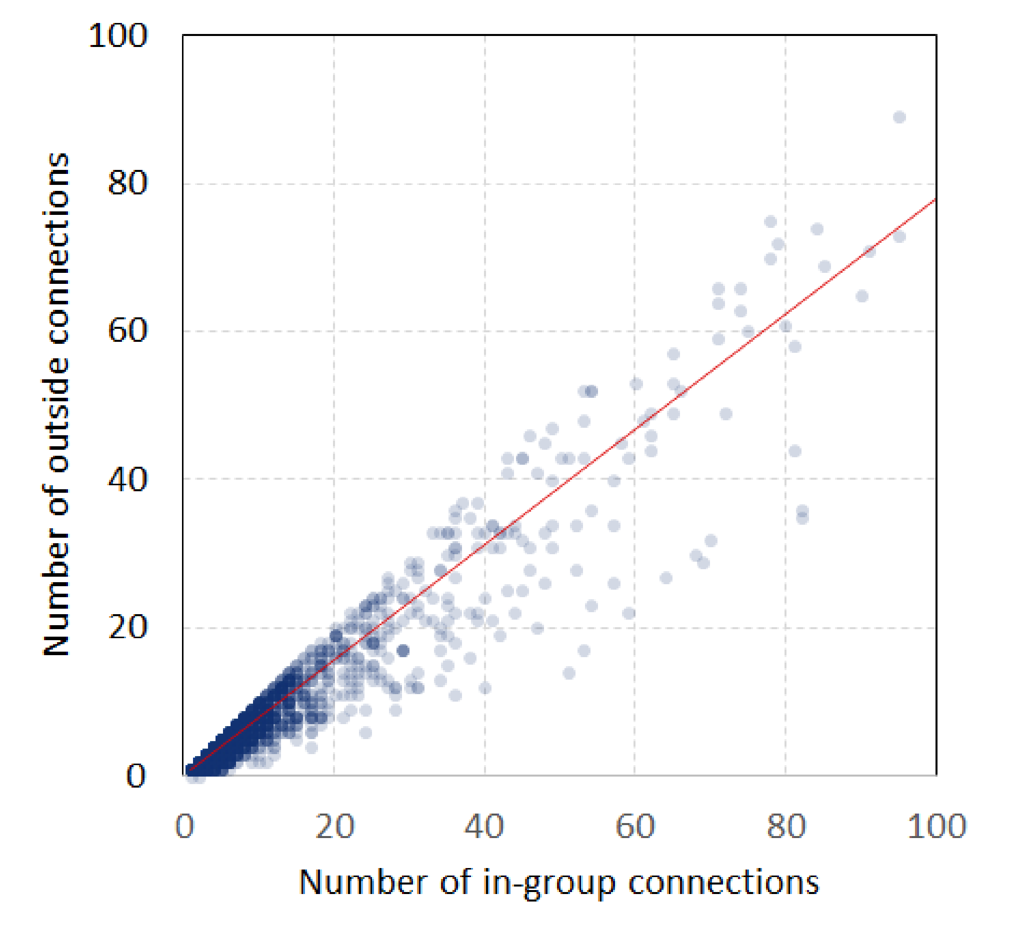
The pictograms indicate the main profile of a given community. E.g. the first group on Figure 24 consists of firms belonging mainly to the food industry, food wholesale and food retail sectors. The second group contains firms from the machine and electronics industry; metal and plastic manufacturers; electricity, gas and steam suppliers. All the other groups can be similarly well interpreted as blocks containing chains of production of a well defined product category.

Figure 29
Polarization among firms in different communities in 2017



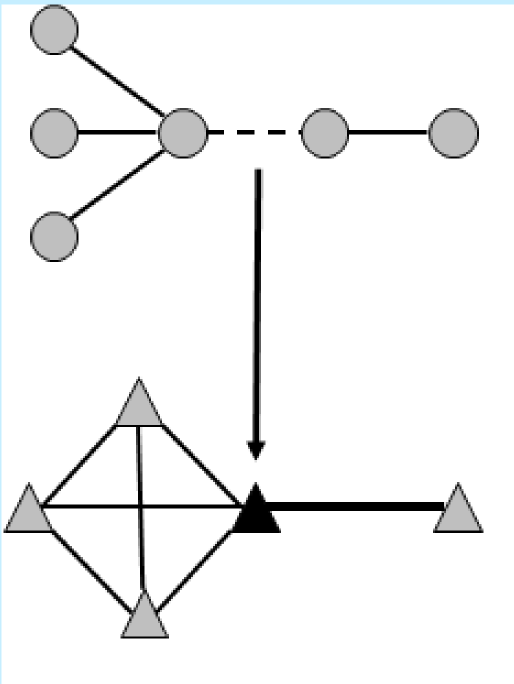
Red color means that typically the group in the row dimension supplied to the group in the column dimension, while blue indicates the inverse situation. Darker coloring indicates stronger polarization.

Figure 30
The number of links within the firms' communities and the number of connections pointing to other communities



The visualization shows only 10% of the nodes.

Figure 31
Creating „Line graph" from a traditional graph



Nodes of the line graph are the links of the original graph, and two nodes (links) are connected if they had a shared endpoint (in the original graph where they were links.

Figure 32
Regression results

	Dependent variable:	
	# of outside links	# of memberships
	(1)	(2)
degree	0.028***	0.032***
	(0.0005)	(0.001)
log(balance sheet total)	0.225***	0.229***
	(0.008)	(0.017)
log(value added)	0.091***	0.049***
	(0.008)	(0.017)
exporter	0.046***	0.026
	(0.013)	(0.027)
government owned	0.099	−0.381
	(0.166)	(0.276)
foreign owned	−0.161***	0.006
	(0.019)	(0.033)
Wood, paper, printing industry	0.353***	0.111
	(0.080)	(0.089)
Manufacture of metal, plastic	−0.253***	−0.217***
	(0.064)	(0.071)
Finance, Insurance	−0.036	0.747***
	(0.152)	(0.168)
Manufacture of wearing apparel	−0.049***	0.598***
	(0.142)	(0.157)
Accommodation	−0.054	0.409***
	(0.090)	(0.100)
Transporting and storage	0.179***	−0.110
	(0.061)	(0.068)
Chemicals and drug industry	0.992***	−0.813***
	(0.160)	(0.177)
Electricity, gas, steam supply	1.303***	0.131
	(0.173)	(0.191)
Water supply; sewerage	1.145***	−0.411***
	(0.121)	(0.133)
Constant	0.152	−0.865**
	(0.333)	(0.369)
Observations	57,407	57,407
Region FE	✓	✓
Industry FE	✓	✓
SME classification FE	✓	✓
R^2	0.710	0.892
Adjusted R^2	0.710	0.892
Residual Std. Error (df = 57356)	2.463	2.724
F Statistic (df = 50; 57356)	2,807.892***	9,457.715***
Note:		*p<0.1; **p<0.05; ***p<0.01

Among the NACE categories only industries with positive, significant coefficients are listed.

Figure 33
Count model coefficients (negbin with log link)

	Dependent variable:	
	# of outside links	# of memberships
	(1)	(2)
degree	0.028***	0.032***
	(0.0005)	(0.001)
log(balance sheet total)	0.225***	0.229***
	(0.008)	(0.017)
log(value added)	0.091***	0.049***
	(0.008)	(0.017)
exporter	0.046***	0.026
	(0.013)	(0.027)
government owned	0.099	−0.381
	(0.166)	(0.276)
foreign owned	−0.161***	0.006
	(0.019)	(0.033)
Other wholesale	−0.127***	0.259***
	(0.031)	(0.068)
Food industry wholesale	−0.358***	0.376***
	(0.040)	(0.077)
Agricultural wholesale	−0.417***	0.523***
	(0.061)	(0.105)
Constant	−3.122***	−2.874***
	(0.184)	(0.338)
Observations	57,407	57,407
Region FE	✓	✓
Industry FE	✓	✓
SME classification FE	✓	✓
Log Likelihood	-58,968.150	-28,877.760
Note:		*p<0.1; **p<0.05; ***p<0.01

Among the NACE categories only industries with positive, significant coefficients are listed.

Figure 34
Zero-inflation model coefficients (binomial with logit link)

	Dependent variable:	
	# of outside links	# of memberships
	(1)	(2)
degree	−2.611***	−0.545***
	(0.022)	(0.012)
log(balance sheet total)	0.001	−0.213***
	(0.033)	(0.037)
log(value added)	0.121**	0.063
	(0.039)	(0.039)
exporter	−0.354***	−0.053
	(0.072)	(0.065)
government owned	1.519	0.407
	(1.140)	(0.940)
foreign owned	0.062	0.216*
	(0.133)	(0.105)
Southern Great Plain	0.018	0.092
	(0.098)	(0.099)
Southern Transdanubia	0.076	0.325*
	(0.127)	(0.129)
Northern Great Plain	0.067	0.403***
	(0.100)	(0.104)
Northern Hungary	0.104	0.271*
	(0.122)	(0.128)
Central Transdanubia	0.053	0.415***
	(0.105)	(0.109)
Central Hungary	−0.190*	0.131
	(0.087)	(0.088)
Western Transdanubia	0.075	0.512***
	(0.106)	(0.112)
Food industry retail	1.338***	−0.207
	(0.314)	(0.423)
Food industry wholesale	0.709**	−0.239
	(0.240)	(0.191)
Construction, real estate	−0.505**	−0.009
	(0.129)	(0.144)
Agriculture	0.682***	0.308·
	(0.195)	(0.186)
Household goods retail	0.724***	−0.015
	(0.200)	(0.219)
Household goods wholesale	0.583**	0.67
	(0.225)	(0.208)
Manufacture of metal, plastic	0.441**	−0.188
	(0.166)	(0.162)
Constant	1.551	4.832***
	(1.156)	(0.995)
Observations	57,407	57,407
Industry FE	✓	✓
SME classification FE	✓	✓
Log Likelihood	-58,968.150	-28,877.760
Note:		*p<0.1; **p<0.05; ***p<0.01

Among the NACE categories only industries with positive, significant coefficients are listed.

4 Conclusion

There is an increasing interest in several areas of economics towards the inclusion of economic network data and granular connections into the theoretical models and into empirical analyses as well. We contribute to these endeavours by unfolding the basic structure and the unique traits of micro-level network data depicting the most important connections among economic actors. For this investigation we could obtain access to sensitive datasets about the ownership links and the supplier connections of Hungarian firms. Using these sources we built the multi-layer representation of the Hungarian firm network which enabled us to gain insight into its previously unobserved structure. Although this data is almost unmatched in the literature, it is very important to acknowledge that it still has limitations regarding its completeness, furthermore, that it requires careful preparations which is highly dependant on the application.

In the case of the ownership network we described three issues which need to be addressed to avoid serious biases: the distribution of ownership shares, the computation of indirect links and the weighting based on the size of the firms. The resulting ownership system proved to be very sparse and disconnected, however, it still revealed some topological characteristics and the typical motifs at the micro-level. We could also assess the significance of economic entities from the point of view of the extent they can influence and control the economy via their ownership relations. This investigation was also possible at more aggregated levels revealing the role of different groups formed based on several characteristics, such as the nationality, the legal category or the HQ location of the owners.

In the case of the supplier layer, we collapsed the network to the level of ownership groups and defined long-term connections between them. We found several topological characteristics, which can be responsible for facilitating contagious processes. This network has high enough density to the emergence of a giant component, within which we can reach any firm with only a few steps. This is due to the presence of hub nodes which have very high degree and bridge nodes which connect the otherwise isolated blocks of the economy. These blocks were identified by community detections methods and it turned out that they represent the different production chains of certain product categories within the economy. In addition to the topological information, we also used a handful of attributes, based on which we found strong homophily in some firm characteristics.

As the scope of this paper covers only the preparation and exploration of these special datasets, there are plenty of options for further research directions, such as the estimation of various kinds of spillovers among firms, the modeling of micro-level input-output systems, or the exploration of link formation.

5 Appendix

5.1 APPENDIX A – FURTHER DESCRIPTION OF THE OWNERSHIP DATA

Depending on the year, we can observe around 800 000 ownership links among almost 1 million actors covering the ownership structure of around 400 000 firms (Figure 35).

The majority of the owners are individuals, while there are only yearly 51 000 – 66 000 links where the owner is a firm. According to Figure 36, the ratio of firms is much higher in the case of foreign owners than in the case of Hungarian owners (40% as opposed to 6%).

As it is depicted on Figure 37, the links of the network do not change significantly from one year to another. Around 75% of the links are present in all the five observed years.

Furthermore, in the case of 35-40% of the links we can see even the extent of the influence (expressed in percentages). However, sometimes (in 16 000-19 000 cases depending on the year) we found firms with ownership links where the sum of the overall influences exceeded 100%. In these cases we corrected the influence for each owner proportionally to make the sum equal to 100%. For links where the influence information was missing, we used a simple imputation method by dividing the missing amount of influence among the remaining owners. (For example if we observed a link with 50% share, and we could also see that there are two more owners associated with the same firm without influence information, we simply divided the missing 50% between the two remaining owners equally.)

Using firms' anonymized tax numbers as keys, we could connect the ownership data to another dataset coming from the Hungarian Tax Authority, which contains several firms characteristics. Although the overall quality of the data is quite good, it is far from being complete: we cannot see tax numbers in 25-30% of the Hungarian owner firms, and the dataset is even more incomplete in the case of listed companies, for which we rarely observe the ownership structure.

5.2 APPENDIX B – WEIGHTING OF FIRMS' INFLUENCE

As it is depicted in Figure 38.a), the *total controlled value* of firm *A* would be the sum of the following elements:

- 100% of firm *B*'s value via a direct link between *A* and *B*;
- 50% of firm *C*'s value via a direct link between *A* and *C*; and
- 50% of firm *C*'s value via the indirect link between *A* and *C* through *B*.

However, as *B*'s whole value is originated from its ownership over 50% of *C*, the direct link between *A* and *B* and the indirect link between *A* and *C* through *B* overlap. As a result, we inflate *A*'s *total controlled value* stemming from this system of ownership ties.

Figure 38.b) shows a possible correction by reducing firm *B*'s value by the part which comes from its ownership over firm *C*. This method might seem simple, however, if we generalize it for the whole network, it becomes very demanding to implement. We would have to examine every paths in the network and correct all the members of these chains (except the endpoints). As it is computationally infeasible, we turned to a simplification of the problem. We wanted to find a node attribute, which is independent from the ownership structure, but conveys some information about firms' value. The best candidate meeting these requirements is the value added of firms, which can be directly applied as a replacement for the previous value proxies (as it is shown in Figure 38.c)).

5.3 APPENDIX C – FURTHER DESCRIPTION OF THE SUPPLIER DATA

We observed the supplier system between 2014-2017, however, the quality in the first year was very poor probably due to the inexperience of both the authorities and the firms in the new reporting requirements. Because of this reason we considered only the period between 2015-2017 in our analysis. The basic description of the network properties of these years can be seen in Figure 39.

Although the general quality in these years is high, we still had to make a few correction. We filtered for situations where the VAT rate calculated from the supplier network data deviates from the official rate (which is 27% in the examined period). We also corrected, if the tax amount and the purchase value was mixed up. Furthermore, we checked the consistency between the sales revenue of firms (coming from corporate tax declaration data) and the sum of the purchases of a given firm's products and services observed in the supplier network.

Besides the EUR 3000 reporting threshold there is another limitation of the analysis of the supplier network, which is coming from the lack of international trade links in the data. To assess the overall significance of missing links, we connected the supplier network to another dataset (also coming from the Hungarian Tax Authority), which is collected as part of the tax declaration of firms. Due to the sensitivity of these pieces of information, we had to use firms' anonymized tax numbers as keys to merge the different data sources (similarly to the case of the ownership network). This data contains several characteristics related to firms' balance sheets and profit and loss statements, among which we can observe their yearly material costs as well. Although the comparability of these datasets is not perfect, it was possible to compare the sum of the reported supply transactions for a given firm to its aggregate material costs. This exercise revealed that around 50% of the material costs cannot be matched to the supplier network (Figure 40). In the case of the larger firms, the main reason for this disparity is probably the unobserved import, while for smaller firms the value threshold is more likely to be the dominant constrain.

According to the regulation, the frequency of the reporting can be yearly, quarterly or monthly depending on the size of the firms and on the weights of the firms' supplier links. As it is shown on Figure 41, the vast majority of the firms report monthly, however, we used yearly aggregation even in these cases for two reasons: (i) we need longer periods to define long-term relationships; and (ii) the other datasets with which we want to connect this network are also on yearly frequency.

A further interesting feature of this data reporting is that firms are required to submit information not only about their partners whom they are buying from, but also about their buyers. This made it possible to build the supplier network from both directions, and examine their differences. If we construct the network from connections where the subject firm reports its suppliers, we have to face the problem, that there are suppliers not subject to the VAT, which results in missing information. However, when one uses the connections where the subject firm reports its buyers, another problem can emerge, namely that firms are less motivated to report as these transactions entail VAT obligations for them. As both approaches has different shortcomings, and even the reporting periods are not guaranteed to match for the two sides of a given transaction, the overlap is – as expected – far from being perfect. Although it is impossible to precisely assess the extent of these biases, we decided to use the former approach because irregularities connected to VAT declarations are certainly not randomly distributed, therefore, they can distort the results more seriously. This issue offers another research direction connected to the detection of fraudulent activities, however, this topic is outside of the scope of our analysis for now.

5.4 APPENDIX D – DESCRIPTION OF THE COMMUNITY DETECTION METHODOLOGY

The task of detecting communities is closely related to the well-known clustering problems. The reason why there are so many sophisticated network related techniques in the literature is the concern of computational feasibility (coming from the often enormous size and the complexity of the analyzed systems). In almost every approaches of this challenge there are two pressing questions: (i) how to measure how well the network is separated given a particular division, and (ii) what algorithm to use to find the best grouping of the nodes[17].

Our choice of the function describing the fitness of a given partition is a widely-used measure called *modularity* Newman (2006). The intuition behind this approach is that a good division of a network is not merely one in which there are dense

[17] It should be noted that in some cases a partition is not the best approach as there are often overlapping communities.

connections between nodes within modules and sparse connections between nodes in different modules. A better formulation would be to say that we are looking for a partition in which there are *fewer than expected* links between communities and *more than expected* links within communities. Corresponding to this idea, the modularity is the number of links present within communities minus the expected number of links placed at random:[18]

$$Q = \frac{1}{2L}\sum_{i\neq j}\left(A_{i,j} - \frac{k_i k_j}{2L}\right)\delta(C_i, C_j) \tag{11}$$

where Q denotes modularity, L is the number of links in the graph, and the Kronecker delta indicates whether node i and j belong to the same community.

The higher the probability of a link is, the smaller its contribution to the modularity score is. If the sum of the increments in the end is positive, that indicates the possible presence of a community structure. Therefore, our goal is to find the divisions of a network with the largest modularity score. Even in our case with less than 100 thousand nodes, it would be obviously impossible to go over all the possible partitions and calculate the modularity for all of them. Blondel et al. (2008) proposed an agglomerative, multi-level modularity optimization algorithm called the ”Louvain-method” which is based on a hierarchial approach:

- Initially, each node represents a community with a single member.
- Every vertex is moved one-by-one to a community where the modularity is increased in the largest extent by the reallocation of the given node.
- In the second phase, each community is considered a new node on their own, and the process goes back to Step 1.
- The algorithm stops either when all the nodes are assigned in one all-encompassing community, or when we cannot increase the modularity anymore.

Although this type of modularity maximizations has some shortcomings, e.g. it has a resolution limit and problems in detecting overlapping communities or hierarchical structures Javed et al. (2018), it is a widely accepted method which can be relatively simply and reliably implemented even by using R’s Igraph package. However, as the ”Louvain-method” is based on a stochastic algorithm, it can give different results for different realizations. To overcome this problem Tandon et al. (2019) proposed a method in which we aggregate the different realizations using the *fast consensus* procedure:

- Building the consensus graph based on existing links only to avoid too high computational cost. (In the consensus graph two nodes are connected if they belong to the same community.)
- As two nodes do not necessarily belong to the same group in every realization, we can use a threshold, below which we ignore the link. (For instance, if we observe them together in less than 20% of the realizations.)
- Adding „triadic closure” links (because nodes sharing neighbors belong to the same community with higher probability).
- Run our standard community detection algorithm on the consensus graph.
- Iterate the procedure until we reach our tolerance level.

[18] Random placement in this case means the randomization of the links with the preservation of the degree distribution.

Figure 35
Basic description of the ownership data

Year	Number of firms	Number of links
2015	367 857	774 944
2016	385 341	800 112
2017	405 823	824 293
2018	428 092	850 132
2019	454 540	884 059

Figure 36
Number of links with foreign/HUN and individual/firm owners

Year	Foreign individual	Foreign firm	HUN individual	HUN firm
2015	18 908	13 815	704 626	37 595
2016	20 055	14 259	725 618	40 180
2017	21 224	14 686	745 363	43 020
2018	23 157	15 132	765 885	45 958
2019	26 691	16 413	791 120	49 835

Figure 37
Duration of links (in years) in the Hungarian firm ownership network

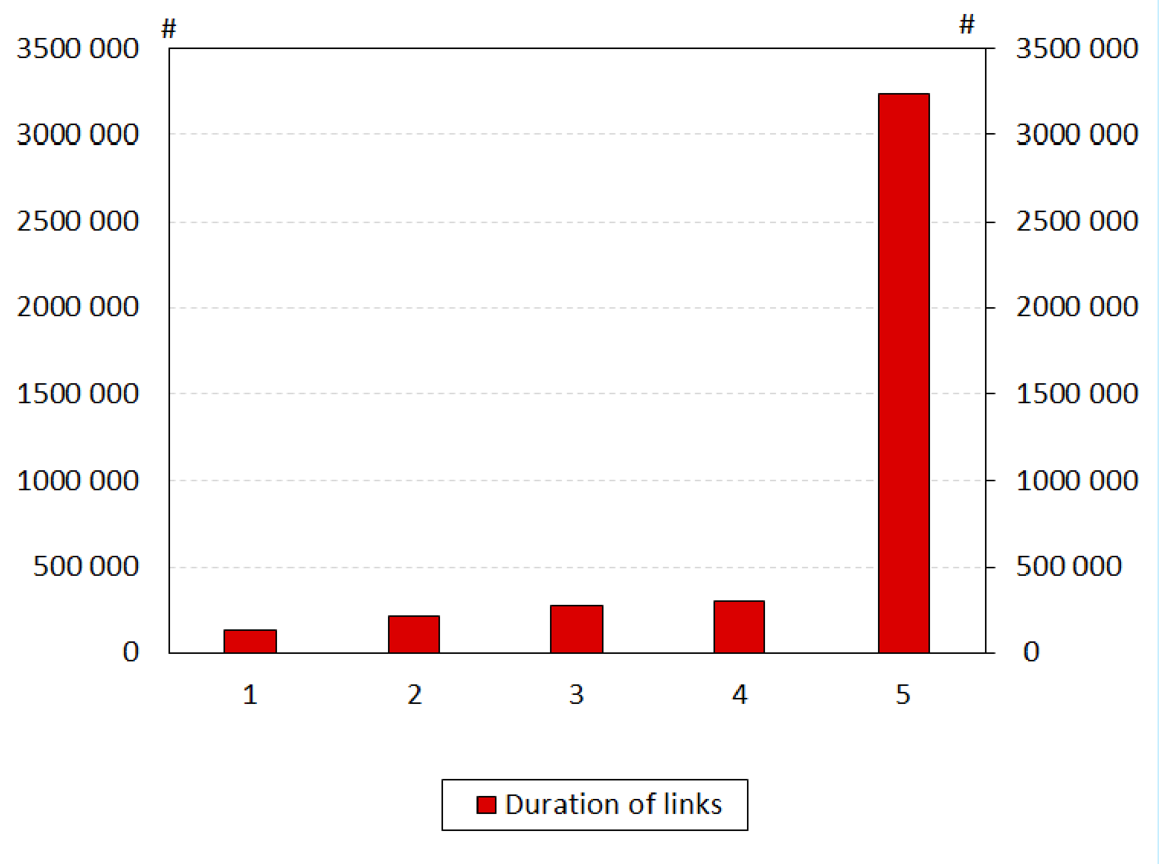
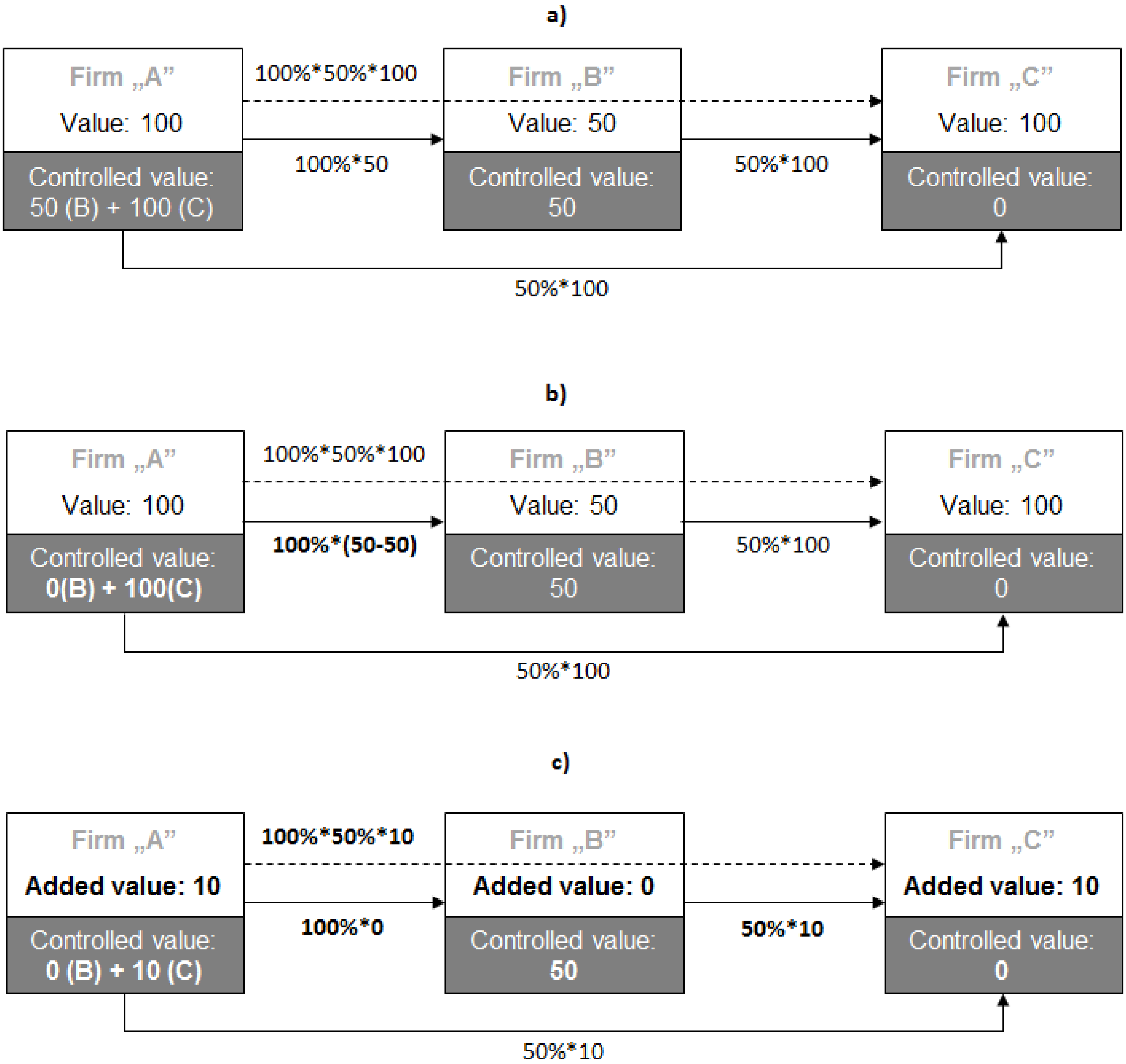


Figure 38
The calculation of indirect control in the ownership network



Direct ownership links are indicated by solid lines, while indirect links represented by dashed lines. Bold font shows changes between a), b) and c).

Figure 39
Description of the supplier network between 2015-2017

	2015	2016	2017
Number of nodes	63 772	79 049	89 778
Number of edges	153 006	205 105	235 913
Density	3.76e-5	3.28e-5	2.93e-5
Average in-/outdegree	2.39	2.59	2.63
Shortest path lengths (avg.)	5.05	4.7	4.92
Shortest path lengths (st.dev.)	1.2	1.1	1.1
Local clustering (avg.)	0.071	0.077	0.078
Reciprocity	0.13	0.11	0.11

Figure 40
Ratio of the sum of supplier transactions and material costs of firms by company sizes.

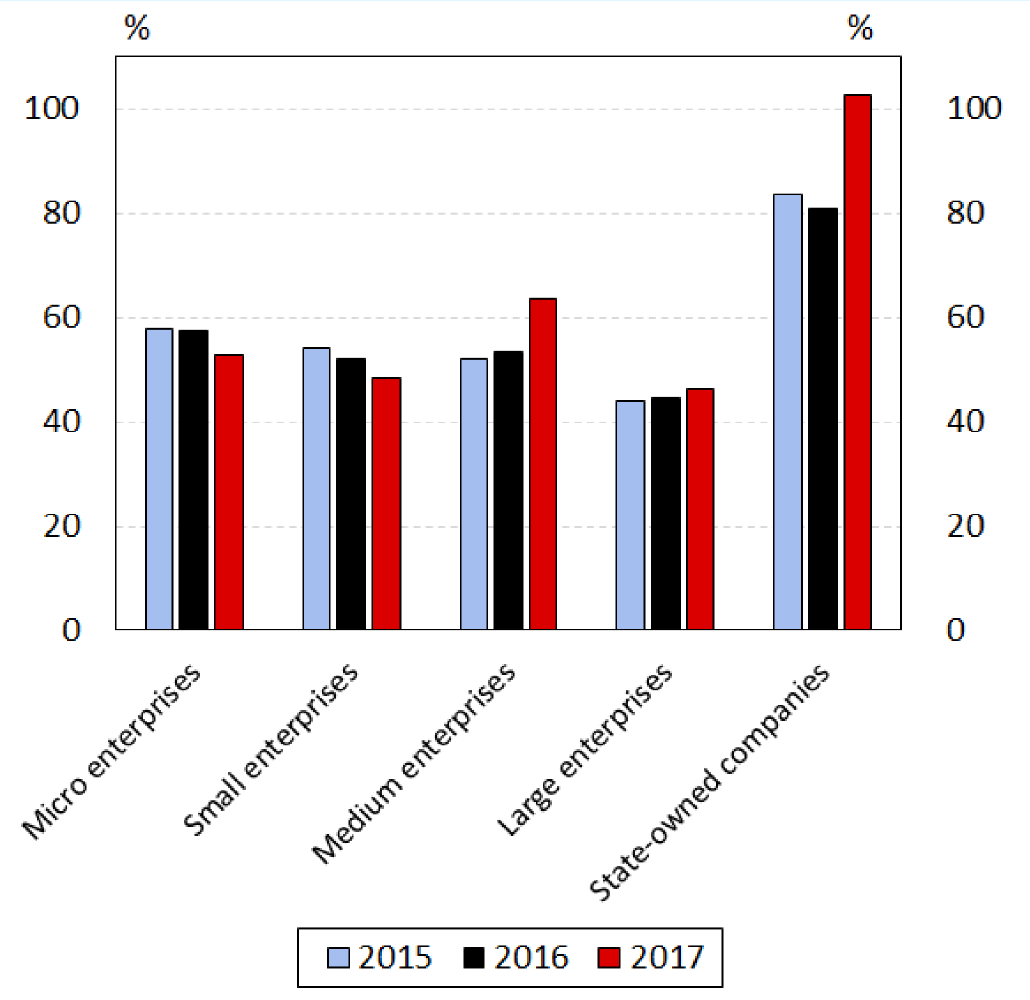
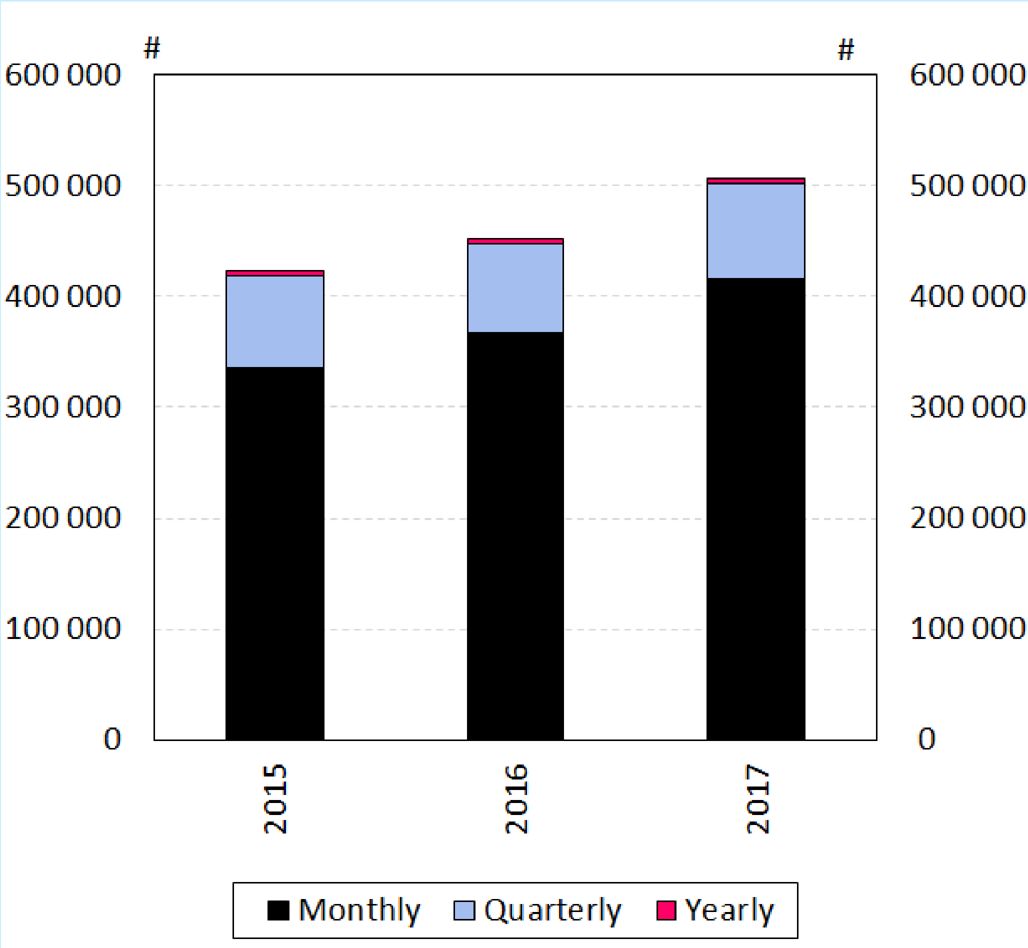


Figure 41
The frequency of firms' reporting about their supplier relationships



References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar and A. Tahbaz-Salehi (2012), 'The network origins of aggregate fluctuations', *Econometrica*, vol. 80 no. 5, pp. 1977–2016

Acemoglu, D., A. Ozdaglar and A. Tahbaz-Salehi (2015), 'Systemic risk and stability in financial networks', *American Economic Review*, vol. 105 no. 2, pp. 564–608

Ahn, Y.-Y., J. P. Bagrow and S. Lehmann (2010), 'Link communities reveal multiscale complexity in networks', *nature*, vol. 466 no. 7307, pp. 761

Barabási, A.-L. (2016), *Network science*, Cambridge university press.

Barrot, J.-N. and J. Sauvagnat (2016), 'Input specificity and the propagation of idiosyncratic shocks in production networks', *The Quarterly Journal of Economics*, vol. 131 no. 3, pp. 1543–1592

Battiston, S., G. Weisbuch and E. Bonabeau (2003), 'Decision spread in the corporate board network', *Advances in Complex Systems*, vol. 6 no. 04, pp. 631–644

Benhabib, J., A. Bisin and M. O. Jackson (2010), *Handbook of social economics*, Elsevier.

Bernard, A. B. and A. Moxnes (2018), 'Networks and trade', *Annual Review of Economics*, vol. 10, pp. 65–85

Blondel, V. D., J.-L. Guillaume, R. Lambiotte and E. Lefebvre (2008), 'Fast unfolding of communities in large networks', *Journal of statistical mechanics: theory and experiment*, vol. 2008 no. 10, pp. P10008

Boehm, C. E., A. Flaaen and N. Pandalai-Nayar (2019), 'Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake', *Review of Economics and Statistics*, vol. 101 no. 1, pp. 60–75

Brioschi, F., L. Buzzacchi and M. G. Colombo (1989), 'Risk capital financing and the separation of ownership and control in business groups', *Journal of Banking & Finance*, vol. 13 no. 4-5, pp. 747–772

Carvalho, V. M. and A. Tahbaz-Salehi (2018), 'Production networks: A primer', ,

Carvalho, V. M., M. Nirei, Y. Saito and A. Tahbaz-Salehi (2016), 'Supply chain disruptions: Evidence from the great east Japan earthquake', *Columbia Business School Research Paper*, no. 17-5,

Costello, A. M. (2020), 'Credit market disruptions and liquidity spillover effects in the supply chain', *Journal of Political Economy*, vol. 128 no. 9, pp. 000–000

Davis, G. F., M. Yoo and W. E. Baker (2003), 'The small world of the American corporate elite, 1982-2001', *Strategic organization*, vol. 1 no. 3, pp. 301–326

Davis, J. A. and S. Leinhardt (1967), 'The structure of positive interpersonal relations in small groups.' ,

Demir, B., B. Javorcik, T. K. Michalski and E. Ors (2018), 'Financial constraints and propagation of shocks in production networks', *Work. Pap., Univ. Oxford, UK*,

Dhyne, E., G. Magerman and S. Rubínová (2015), *The Belgian production network 2002-2012*, tech. rep., NBB Working Paper.

Dutta, B. and M. O. Jackson (2003), 'On the formation of networks and groups', in, *Networks and groups*, pp. 1–15, Springer.

Elliott, M., B. Golub and M. O. Jackson (2014), 'Financial networks and contagion', *American Economic Review*, vol. 104 no. 10, pp. 3115–53

EVANS, T. S. AND R. LAMBIOTTE (2010), 'Line graphs of weighted networks for overlapping communities', *The European Physical Journal B*, vol. 77 no. 2, pp. 265–272

FRANK, O. (1971), *Statistical inference in graphs*, Foa Repro.

FUJIWARA, Y. AND H. AOYAMA (2010), 'Large-scale structure of a nation-wide production network', *The European Physical Journal B*, vol. 77 no. 4, pp. 565–580

GABAIX, X. (2011), 'The granular origins of aggregate fluctuations', *Econometrica*, vol. 79 no. 3, pp. 733–772

GAI, P. AND S. KAPADIA (2010), 'Contagion in financial networks', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 466 no. 2120, pp. 2401–2423

GARCIA-BERNARDO, J., J. FICHTNER, F. W. TAKES AND E. M. HEEMSKERK (2017), 'Uncovering offshore financial centers: Conduits and sinks in the global corporate ownership network', *Scientific Reports*, vol. 7 no. 1, pp. 6246

GLATTFELDER, J. B. (2013), 'Backbone of complex networks of corporations: The flow of control', in, *Decoding Complexity*, pp. 67–93, Springer.

GORG, H. AND E. STROBL (2001), 'Multinational companies and productivity spillovers: A meta-analysis', *The economic journal*, vol. 111 no. 475, pp. F723–F739

GOYAL, S. (2012), *Connections: an introduction to the economics of networks*, Princeton University Press.

GRANOVETTER, M. (1976), 'Network sampling: Some first steps', *American journal of sociology*, vol. 81 no. 6, pp. 1287–1303

HEEMSKERK, E. M. AND F. W. TAKES (2016), 'The corporate elite community structure of global capitalism', *New Political Economy*, vol. 21 no. 1, pp. 90–118

IINO, T. AND H. IYETOMI (2012), 'Directional bias between communities of a production network in Japan', in, *Intelligent decision technologies*, pp. 273–280, Springer.

JACKSON, M. O. (2010), *Social and economic networks*, Princeton university press.

JAVED, M. A., M. S. YOUNIS, S. LATIF, J. QADIR AND A. BAIG (2018), 'Community detection in networks: A multidisciplinary review', *Journal of Network and Computer Applications*, vol. 108, pp. 87–111

DE JEUDE, J. V. L., T. ASTE AND G. CALDARELLI (2019), 'The multilayer structure of corporate networks', *New Journal of Physics*, vol. 21 no. 2, pp. 025002

KIKKAWA, K., G. MAGERMAN AND E. DHYNE (2019), 'Imperfect competition in firm-to-firm trade', *Available at SSRN 3389836*,

LEECH, D. (2002), 'Computation of power indices', ,

LEONTIEF, W. W. (1951), *The structure of American economy, 1919-1939: an empirical application of equilibrium analysis.* Tech. rep.,

LIU, X., P. SILER, C. WANG AND Y. WEI (2000), 'Productivity spillovers from foreign direct investment: Evidence from UK industry level panel data', *Journal of International Business Studies*, vol. 31 no. 3, pp. 407–425

LUCAS JR, R. E. (1977), *Understanding business cycles*,

MAGERMAN, G., K. DE BRUYNE, E. DHYNE AND J. VAN HOVE (2016), *Heterogeneous firms and the micro origins of aggregate fluctuations*, tech. rep., NBB Working Paper.

MCNERNEY, J. (2009), 'Network properties of economic input-output networks', ,

NEWMAN, M. (2018), *Networks*, Oxford university press.

NEWMAN, M. E. (2006), 'Modularity and community structure in networks', *Proceedings of the national academy of sciences*, vol. 103 no. 23, pp. 8577–8582 de Silanes, F. L., R. La Porta and A. Shleifer (1999), 'Corporate ownership around the world', *Journal of Finance*, vol. 54 no. 2, pp. 471–517

Tandon, A., A. Albeshri, V. Thayananthan, W. Alhalabi and S. Fortunato (2019), 'Fast consensus clustering in complex networks', *Physical Review E*, vol. 99 no. 4, pp. 042301

Tintelnot, F., A. K. Kikkawa, M. Mogstad and E. Dhyne (2018), *Trade and domestic production networks*, tech. rep., National Bureau of Economic Research.

Tomasello, M. V., M. Napoletano, A. Garas and F. Schweitzer (2017), 'The rise and fall of R&D networks', *Industrial and corporate change*, vol. 26 no. 4, pp. 617–646

Tumminello, M., F. Lillo and R. N. Mantegna (2010), 'Correlation, hierarchies, and networks in financial markets', *Journal of Economic Behavior & Organization*, vol. 75 no. 1, pp. 40–58

Vitali, S. and S. Battiston (2011), 'Geography versus topology in the european ownership network', *New Journal of Physics*, vol. 13 no. 6, pp. 063021

Vitali, S. and S. Battiston (2014), 'The community structure of the global corporate network', *PloS one*, vol. 9 no. 8, pp. e104655

Vitali, S., J. B. Glattfelder and S. Battiston (2011), 'The network of global corporate control', *PloS one*, vol. 6 no. 10, pp. e25995

Watanabe, T., I. Uesugi and A. Ono (2015), *The economics of interfirm networks*, Springer.

Yang, B., D. Liu and J. Liu (2010), 'Discovering communities from social networks: methodologies and applications', in, *Handbook of social network technologies and applications*, pp. 331–346, Springer.

Zajac, E. J. and J. D. Westphal (1996), *DIRECTOR REPUTATION, CEO/BOARD POWER, AND THE DYNAMICS OF BOARD INTERLOCKS.*

MNB Occasional Papers 2020/139

Unfolding the hidden structure of the Hungarian multi-layer firm network

Budapest, November 2020

mnb.hu

1054 BUDAPEST, SZABADSÁG TÉR 9.