UNIVERSITY OF TARTU

Faculty of Social Sciences

School of Economics and Business Administration

Irina Martyanova

ANALYSIS OF REGIONAL ECONOMIC NETWORK STRUCTURES BASED ON ESTONIAN VAT TRANSACTION DATA

Master's Thesis

Supervisors: Andres Võrk, Taavi Unt, Uku Varblane

Tartu 2019

Name and signature of supervisor…………………………………………………

Allowed for defence on...

(date)

I have written this master's thesis independently. All viewpoints of other authors, literary sources and data from elsewhere used for writing this paper have been referenced.

…………………………………..

(signature of author)

Analysis of regional economic structures based on Estonian VAT transaction data

Abstract

A considerable amount of literature has been published on networks. Still, these studies concentrate mostly on social interactions and the research gap on economic networks is evident. The focus of this paper is twofold and combines economic network theory and regional economic performance analysis. The research is conducted using three main data sources – Estonian VAT tax declaration data, annual accounts data and business registry data for two consecutive years 2016-2017. The results include community detection analysis, analysis of between-region interactions of Estonian geographic regions (according to NUTS classification), regression of the probability of trading relationship and regional analysis of productivity of Estonian firms. Overall, the results indicate that there are regional differences in productivity for Estonian firms that can be associated with lower embeddedness in networks – Tallinn and Tartu show the highest productivity values and higher degree of embeddedness in value chains, while Ida-Viru county has fewer interactions outside its own geographical boundaries. Performance of different industries also varies across regions and different ownership types. Moreover, the analysis of regional interactions proves that distance matters for the formation of firms' partnerships in Estonia but is less significant when a trading relationship has already been established. The results indicate that the effects of current regional network development and specialization should be taken into account in policy making.

Keywords: economic networks, regional analysis, productivity, value chains, VAT declaration data

ACKNOWLEDGEMENT

First, I would like to express my gratitude and appreciation to my supervisors Andres Võrk, Taavi Unt and Uku Varblane from the Centre for Applied Social Sciences at the University of Tartu for their patient guidance, constant encouragement, valuable suggestions and useful critiques during this research. My special thanks are extended to my opponent, Jaan Masso, for the detailed review and constructive recommendations. You supported me greatly during the writing of this thesis and I am sincerely grateful to you.

I am particularly thankful to the staff of Statistics Estonia for granting me access to the data for this research. I also thank Foresight Centre for providing research funding.

Finally, I wish to thank my parents and my friends for their love, enormous support and encouragement throughout my studies.

TABLE OF CONTENTS

1.INTRODUCTION

The last two decades have seen a growing trend to production fragmentation due to openness to trade, information and technology developments. With the tremendous progress of the international trade, many countries specialize in particular production stages and secure their position in global value chain (GVC) based on the country's access to resources, technological progress, etc. (Baker and Miroudot 2013). Openness to international collaboration has triggered the so-called "trade in tasks" (Grossman, Rossi-Hansberg 2008). One of the classic examples is China that specializes mostly in assembly, while other countries act as innovation drivers (Suganuma 2016). Fragmentation of production has allowed the firms to benefit in terms of efficiency and costs. Strong multinational companies appeared on the market and their role is undoubtedly important in the formation of GVC. Cost-minimizing strategies of large corporations are based on dividing production steps between different countries that consequently shapes international networks of companies and their subsidiaries. Such multinationals are usually occupying central positions on the market and therefore their role might even be more significant than the role of the country itself (Baldwin 2016).

Nevertheless, during recent years the development of GVC has stopped. One of the potential causes listed by De Backer, Flaig (2017) is the fact that the maximum complexity of international networks has been reached and therefore firms are finding new solutions for production strategies. When the production chain is highly fragmented, the supply risks rise significantly because of the potential spatial shocks (change in economic policies, transportation issues, natural disasters, etc.). One solution for the companies to minimize international risks within the chain is to concentrate more production stages inside one country. Therefore, the study of within country networks and regional distribution can shed light on the GVC position changes in recent years.

Recent evidence suggests that similar development patterns are noticeable at the firm-level. There exist several motives for firms' cooperation. Firstly, the need for resources has contributed to the formation of input exchange between enterprises. Secondly, instead of accomplishing all the production steps themselves, the firms opt to transfer some of the tasks (usually marketing, transportation, etc.) to third parties (Kraemer et al., 2011). Such collaborations of enterprises form networks that can be either local or global.

According to Goyal (2007), the presence of certain patterns in regional networks shapes the structure of economic activity of the regions. Asheim, Isaksen (2002) underline that access to the same raw materials, similar social and collective values, institutional structure as well as labor market structure contribute to the formation of regional specialization and industrial development. Therefore, collaborative networks may have a decisive significance for firms' and regional performance. Another phenomenon that is introduced in the literature is clustered firms. Previous research such as that conducted by Longhi (2017, p.2) refers to clusters as "geographic concentration of interconnected companies that compete but also co-operate". Clustered firms are found to offer higher wages (Audretsch, Feldman 1996), contribute to knowledge and innovation spillovers (Alcácer, Zhao 2012) and create positive externalities (Porter 1996).

Recently, there has been a renewed interest in inter-firm connections that led to the formation of the cluster support programs as a part of European Comission decision (2008b). Ever since, several countries implemented policies aiming to encourage joint innovation projects, etc. There were about 130 implemented cluster policies in 2013 (Nardone, Muscio, Lopolito 2013). The implementation of cluster policy and industry specific instruments addresses the issues of employment, increasing firm competitiveness and overall economy growth (European Cluster Observatory 2009). According to Operational Program for Cohesion Policy Funds in Estonia (2014-2020), focus on encouragement of networks and cooperation between entrepreneurs and R&D projects is part of the policy.

Despite the wide discussion of economic networks and clusters, the literature still lacks studies that would concentrate on regional aspects of within country interactions that could give an insight into linkage patterns between territorial units and form a clear understanding of regional specialization, firms' interactions and value chain length across country regions. Moreover, limited or restricted access to the data poses a problem for researchers and can significantly influence analysis results.

This thesis intends to establish the structure and impact of networks and regional interconnections on the productivity and performance of the firms in Estonia. Network analysis applied to VAT tax declaration data. As it was outlined in the research conducted by Dhyne, Duprez (2015) on Belgian economy, by using between-firm transaction data from annual VAT tax declaration data it is possible to draw existing interactions of companies and spatial dependencies. The length of production chain was calculated at the firm-level by regional units. For example, the fragmentation of the value chains is found to be different between Belgian regions. Thus, in the Flemish region enterprises operate in sectors that are part of a more fragmented chains. Moreover, it is underlined in their research that on average firms with more connections, the ones that have a more pronounced product specialization, operate more efficiently while it was previously argued in the literature that concentration of different production stages is more efficient. Controversial results reflect a possible research gap and the conclusion that "one size fits all" may not be applicable to all countries.

The thesis is part of the research on global value chains (GVC) for Estonia "Eesti ettevõtete osalemine ja positsioon globaalsetes ja lokaalsetes väärtusahelates" conducted by Võrk, Unt, Varblane (2018). The main conclusions of previous research can be summarized as follows. Exporting firms are more productive than non-exporting and, contrary to previous findings, higher value added is generated by firms that are further upstream in the value chain. The results are partially explained by export orientation of a large share of Estonian firms. Nevertheless, there exist grounds for an assumption that firms' position in the value chain and overall length of the chain depends on the geographical location (Dhyne, Duprez 2015) and this thesis aims to fill in the research gap on possible spatial effects. The econometric analysis includes regression analysis of the tie formation between the firms with the inclusion of inter-regional effects. The main questions that are answered in the thesis are "Do Estonian firms tend to form local (within-region) networks?", "Does the fragmentation and value chain position of the firm depends on the region it is located in?", "Is it possible to increase firm's productivity by forming regional or global networks?". Due to the controversy of previous conclusions, the paper also aims to establish regional networks effects on the productivity of Estonian firms.

This thesis investigates the case of Estonian economy and it is important to give a brief overview of the economic conditions in the country. Estonia presents an interesting case of a small open economy. After the break of the Soviet Union, Estonia took on an individual development, establishing liberal economic policies and successfully integrating itself in global value chains. At this moment, Estonia offers beneficial and attractive conditions for various businesses and start-ups (for example, e-residency, etc.). The investment climate is also favorable. Estonia is an active participant of international trade exporting intermediate goods to Finland, Sweden, Russia, Germany, Latvia, Lithuania, etc. According to the statistics for 2011, approximately 69% of Estonian exports are intermediate goods and around 66% of Estonian imports are intermediate goods. The values are also noted to have risen compared to 1995 (Yrkkö, Mattila, Seppälä 2017). According to OECD Trade and Investment Statistical Note, around 44% of value added is directed to foreign final demand and around 41% constitute foreign value added in Estonian final demand. The reason behind this data can be partially explained by the share of the foreign-owned companies in Estonia (approximately 38% of private sector employment in 2013[1]). These companies tend to be more export-orientated and contribute a significant part to country's export values. The abovementioned facts show that Estonian economy presents a case of a small open economy actively participating in value chains.

The within-country level shows that Estonian economy can also be researched at the cross-regional level. There are 15 regional units (counties) in Estonia: Harju county, Hiiu county, Ida-Viru county, Jõgeva county, Järva county, Lääne county, Lääne-Viru county, Põlva county, Pärnu county, Rapla county, Saare county, Tartu county, Valga county, Viljandi county, Võru county. These units differ in the economic activity concentration, specialization, etc. For example, in Harju county (Northern Estonia) the vast majority of enterprises belongs to "Wholesale trade and retail; repair of motor vehicles, motorcycles" sector, followed by "Professional, scientific and technical activities" and "Conctruction" Same structure of economic activity is in Tartu county, while the rest of Southern Estonia specilizes mostly in "Agriculture, forestry and fishing sector". The number of enterprises from wholesale trade industry is also the largest for Ida-Viru county (Northeastern Estonia)[2]. The specialization of imports and exports differ across counties as well (Statistics Estonia bulletin, 2018). For example, the largest share of exporting commodity for Tallinn and Harju county was electrical equipment, for Viljandi and Valga counties - wood and products of wood, for Ida-Viru county – mineral products. The largest numbers

[1] OECD Estonia. Trade and Investment statistical note, 2017.
[2] Statistics Estonia database, data on the enterprises' number by county.

of operating enteprises are in Harju, Tartu, Parnu and Ida-Viru counties (Appendix 1). Same regional units are categorized by high presence of foreign-owned enterprises that can significantly influence the export orientation and productivity of the region (Javorsik 2004). Estonian economy characteristics provide grounds for an assumption that firms' interconnections and regional patterns may influence overall value chain position and productivity of enterprises across regions. Similar analysis was conducted by Dhyne, Duprez (2015), who used trade data between the firms to explain the value chain position of the company. By using VAT tax declarations data set to establish between-firm interactions the paper aims to fill in the research gap on regional networks for Estonia.

The paper is structured as follows. Section 2 gives an overview of the existing literature on firms' networks and clusters. Section 3 describes hypotheses that are researched in the thesis. Section 4 gives an overview of the methodology used. Section 5 reports the data used in the study and data preparation stages. Section 6 describes main findings for Estonian data and econometric analysis results and provides possible explanations behind the results and further research opportunities.

2. LITERATURE OVERVIEW

This study focuses on the firms' networks, clusters, value chains and productivity effects. The volume of literature on these topics is quite large and can be divided into three main parts: i) research on networks; ii) studies on cluster formation and cluster externalities; iii) cross-sectional studies on networks, clusters, value chains.

The academic literature published on networks is quite wide. A few theoretical papers introduce main definitions and establish structural differences of networks. One of the fundamental sources of network theory is Goyal (2007), who defines nodes (actors in the network), degrees of nodes (the number of links between nodes), structure types (star, core-periphery, etc.). Also, the literature on networks is not limited by purely economic interactions but introduces social, media and institutional networks as well. Economic between-firm networks (also can be referred as inter-organizational networks) can include input exchange, joint venture or R&D projects, strategic alliances, etc. Previous research directions include cooperation motives, learning, trust and conditions for network formation as well as network consequences such as innovation, firm survival, performance, etc. (Brass et al. 2004).

Much of the current economic literature investigates the effects of firm networks on knowledge and innovation spillovers, firm's performance and economy. Main directions of research can be summarized in three main groups: explanation for the motivation behind the formation of networks; research of the network effects on firms' performance and spatial interdependencies; structural differences of networks (Ozman 2009). Since the focus of this research is on regional dependencies, the literature we are particularly interested in is spatial interdependencies and network effects on firms' performance.

Networks are assumed to produce positive externalities for all participants. Access to the same knowledge, inter-reliance of the network participants on each other, location in the same geographic region and local competition effects can serve as incentives for future growth (Porter 1996). A study by Basant, Chandra, Upadhyayula (2008) on a case of Indian IT sector shows that networks, in fact, help to develop capabilities that can increase firms' performance indicators. Another research by Beckman, Haunschild (2002) has shown that a company with the portfolio of diverse network partners exhibits higher productivity. The benefits of networks can be clear but there are a few crucial questions raised in the literature about network analysis that still do not have a precise answer for.

One of the questions raised is whether the geographical proximity is pivotal for network's formation, that can be summarized as whether the location or partners matter more for firm's success. The results seem to be quite controversial. Dahl, Pedersen (2004) on the case of Northern Denmark provide evidence that geographically close regions are more likely to form knowledge and technology exchange connections. The results are supported by findings of Owen-Smith, Powell (2004) that geographical proximity matters, especially for information and knowledge flows. It is particularly crucial for scientifically oriented sectors that seek high-skilled labor and access to new introduced innovations. On the contrary, Boschma (2005) argues that proximity "in general" is essential for tie creation. Assessing spatial proximity significance individually cannot fully explain all the processes within a network. Such factors as social interactions, organizational structures should also be taken into account. In fact, it is emphasized by the author that operating in a closed, geographically bounded network can negatively affect innovation processes. Nevertheless, nowadays the development of transportation system allows the formation of efficient long-distance business connections. For example, Bernard, Moxnes, Saito (2016) investigate the introduction of high-speed train lines and its effects on the productivity of firms located near these lines. The findings confirm that enterprises located geographically closer to the train lines performed better compared to the firms located further away. Therefore, location within the access to transportation allows more effective buyer-seller search conditions and in this case the effects of geographical proximity are less pronounced. Another important finding is that the more productive firms tend to have more partners that are located farther. These findings reflect that spatial proximity does not guarantee effective buyer-seller relationship and interacting only within regional network in some cases can actually limit firms' development. Boschma, Ter Wal (2007) on a case of footwear district Barletta in Italy confirm that being "co-located" is not a sufficient condition for either knowledge or productivity spillovers. Same conclusions are also made by Kesidou and Romijn (2008), Terre and Rallet (2005). Ter Wal et al. (2014) underline that economic geography has developed beyond the limits of geographical location and the impact of spatial distances has faded over the last decades. Moreover, Weterings, Boschma (2009) on a case of Dutch software companies research the effects of geographical proximity and conclude that although geographical closeness contributes to the development of social interactions, it does not show any direct significant influence on innovation or business partnerships. Overall, the question of the role of geographical proximity is still open.

Previous studies also established that position inside the network can carry a significant impact on overall firm's performance, either positive or negative. For example, network position can either grant information access or deprive the actor from it. Similar parallels can be drawn for power over other actors and market control. One of the classic studies that analyzed the effect of firms' performance is the one conducted by Uzzi (1996) that considers firms' survival and network position correlation and finds positive dependence up to some point. A more recent study by Giuliani, Bell (2007) establishes that central position in the network can positively affect firms' innovative performance. Therefore, being central actor in the network can bring positive effects and advantages but the very fact of embeddedness in the network does not itself guarantee productivity increase or positive spillover effects. Taking all of the abovementioned into account, it can be summarized that network effects are dependent on the reason behind tie formation and the position of an actor within the network.

By establishing the structural form of networks for Estonian counties the question whether firms tend to form local or global collaborations can be answered[3]. Another case that can be researched is related to the location and university spillovers on firm's productivity. Audretsch, Lehman (2012) establish that locational and university spillovers are complementary but do not fully explain firm's performance. Therefore, it is possible to see a high productivity correlation of Estonian counties that are geographically bounded with Tartu county (location of the University of Tartu). Moreover, by comparing the production chain length it is possible to establish whether the companies located closer to research centers are more likely to have a more centralized or fragmented value chain.

Network theory and economic geography literature is quite wide but still encounters theoretical and empirical challenges. Most of the research papers apply static network analysis due to the lack of appropriate data and techniques while dynamic analysis remains undeveloped. Data collection poses another significant problem. There are a number of methods for network analysis that use different data sources. For example, one way to establish network connections among firms is a roster-recall methodology which requires direct phone questioners for firms. The analysis is feasible only in case of a high response rate and it is virtually not possible to collect enough data for dynamic network research. Another possible option introduced in the literature is the usage of primary data, for example patent data (Ter Wal, Boschma 2007). Therefore, because of the data collection issues as well as relatively few applicable methods, the field of dynamic network analysis only begins its development. Statistical models for networks that use one period observation for analysis include exponential random graph class of models (ERGM), network block models, latent network models (Kolaczyk, Csárdi 2014). The largest class is ERGM (or often referred to as p* models) that consist of several approaches based on underlying assumptions of actor interdependencies and parameter constraints. These type of models are based on an assumption that there is a stochastic process that forms respective ties in an existing network. The obtained analysis results

[3] In this paper global networks are referred to as firm collaborations outside geographic unit (according to NUTS).

allow to see network characteristics and relevant nodes' attributes (effects such as transitivity, etc.) that might have an impact on future tie formation within a network (Robins et al. 2007)[4]. Understanding the reasons behind tie formation in this case is essential for explaining network types and its spillovers. Another classic tool used to analyze tie formation between actors in the networks is binary regression model approach where existence of a tie is taken as a dependent variable. Such method has been successfully used for interorganizational and social connections analysis (Dhyne, Duprez 2016; Goyal, Fafchamps, van der Leij, 2008). The main advantages of the binary regression model approach are simplicity and minimization of computational time in the case of large networks.

A considerable amount of literature has also been published on clusters. The first study on clusters is considered to be the research of industrial districts by Marshal in 1890. Another classical research paper on clusters is "Competitive advantage, agglomeration economies, and regional policy" by Porter (1996). Commenting on cluster, Porter (1996, p.1) writes: "Cluster is a group of industries connected by specialized buyer-seller relationships or related by technology and skills". Similar definition is provided by Longhi (2017, p.2), who refers to clusters as "geographic concentration of interconnected firms". Cluster belongs to the economic agglomeration classification and the necessary condition is that firms operate in the same industry. One of the most famous examples of clusters can be Silicon Valley (Boja 2011). Within the framework of this study, clusters are referred to as communities of enterprises that do not necessarily belong to the same sector of activity but have above average probability of interaction. Analysis of clusters gives more efficient results rather than analysis of the industries since firms are influenced by same shocks, spillovers and infrastructure conditions (if located in the same geographic region). Although initially cluster formation implied geographic proximity, nowadays with the development of transport and infrastructure systems the choice of the partner in buyer-seller or cooperation relationship is no longer bounded by location but rather by optimization and management decisions. Still, the presence of a cluster may have an impact on the overall regional development and performance.

[4] The field of network analysis constantly develops, especially ERGM. New approaches introduced for data analysis can be found here: http://www.melnet.org.au/

The benefits of clusters are widely discussed in the literature. Previous research established that clustered firms show higher employment rates and wages (Audretsch, Feldman 1996). Moreover, either strong or weak ties between firms contribute to information and knowledge exchange (Rowley, Behrens, Krackhardt 2000) and clustered firms tend to attract high-skilled labor (Boja 2011). Kozovska (2010) investigates the impact of regional clusters on firm-level productivity for Poland and Romania by matching European Cluster Observatory and firm-level financial data. The results indicate that cluster effect is statistically significant and contributes to the reduction of technical inefficiency of firms. Moreover, firms with strong networks control innovation processes more and reduce the knowledge outflow risk (Alcácer, Zhao 2012). Giuliani E. (2006) argues that due to heterogeneity among firms the performance of companies highly depends on its characteristics and position in the network. Therefore, the impact of clusters on firms varies significantly and is controversial. The effects of inter-regional connections on productivity of the firm is of particular importance in this study.

Based on the research by Turkina et al. (2016), it can be concluded that one way to investigate the cluster formation is to conduct a social network analysis. Cluster itself is usually also embedded in the value chains (Bathelt, Li 2014). By transmitting some part of value chain activities to actors outside the cluster, the firms can gain access to knowledge or resources that might be unavailable inside the cluster (Sturgeon et al. 2008). Ter Wal and Boschma (2007) also emphasize that social network analysis is an alternative young but promising tool for analyzing clusters and regional performance. Clustered firms also try to expand their network connections to benefit from different spillover effects. Network research can potentially shed more light on regional performance and interaction issues and partially explain both the performance of clusters as well as value chains. Since the value chain can consist of both individual firms and clustered firms, the degree of production stages’ concentration may vary across regions. Therefore, a regional level research of networks and clusters can help explaining industry and country level chain length and firms’ productivity.

3. RESEARCH QUESTIONS AND HYPOTHESES

The network analysis is expected to reflect main tendencies of firms' interactions across regions but based on the official statistics some hypotheses and propositions can be made regarding potential results.

The results of previous studies show that the roots of firms' decision to collaborate with each other may be referred to as a combination of resource based, cost based and spatial proximity reasons (Bathelt, Li 2014; Bernard, Moxnes, Saito 2016). There are 5 major geographical units in Estonia according to NUTS (Regional Classification in European Union) – Northern, Central, Northeastern, Western and Southern Estonia. Nothern and Northeastern parts constitute only of one county – Harju and Ida-Viru counties respectively. Based on Statistics Estonia data results, it can be concluded that these regional units differ in operating enterprises' volume, specialization and concentration. Harju county (Northern Estonia) is characterized by high concentration of economic activity. Due to the developed infrastucture, the presence of transport routes (Tallinn Port) and the concentration of research and innovation activity, Tallinn as a city-region and Harju county as a whole unit is expected to well operate inside its regional network. The presence of different enterprises with diverse specializations and innovation opportunities can partially influence the structure of the network inside Harju county.

Hypothesis 1. Firms located in Harju county (Northen Estonia), which is characterized by the large number of registered enterprises, are more likely to form local connections while companies registered in Southern or Central parts have a more diversified spatial pattern.

Based on the research by Audretsch, Lehman (2006) there exist grounds to assume that university spillovers can effect the network pattern inside the region and have an impact on production chain division. Tartu county, located in Southern Estonia, is another research and innovation centre of Estonia due to the presence of Tartu University. The volume of research activity is higher than in other regions and it can partially effect overall firms' performance and network structure in the geographical unit.

Hypothesis 2. Tartu county is likely to have a similar network structure as Harju county and firms inside this regional unit are on average more productive.

Value chain position takes into account particular actors of production process. The methodology of value chain length calculation though does not reflect whether some value chain members specilize only in certain activity while other participants might have a less concentrated production.

Hypothesis 3. The presence of communities can partially explain value chain position. The choice of whether the production is more fragmented or more centralized is dependent on the network structure and its externalities in the region.

Descriptive social network analysis introduces several important characteristics that are common for most networks – transitivity (how likely two actors that have a common third partner to coooperate among themselves as well), homophily (actors that are similar are more likely to cooperate), assortativity or assortative mixing (measure of correlation of connected vertices or in other words tendency of nodes to be connected with other nodes of the same degree) (Kolaczyk, Csárdi 2014). Organizational networks are similar to networks of individuals in many ways and therefore same patterns are expected to be present in between-firm data. For example, transitivity pattern that can be described as "if company A is connected to company B that interacts with company C, then the likelihood that firm A and firm C cooperate is high". Each company within a network carefully chooses a partner based on reliability criteria. Reliability can be defined through observed firm characteristics such as size or productivity as well as through unobserved factors such as management interactions between the firms or sharing a common partner. Therefore, it is reasonable to assume that the presence of "common partner" effect in the firms' data in Estonia.

Hypothesis 4. Interorganizational network exhibits typically observed patterns such as transitivity, homophily.

To summarize, the paper examines three main aspects – network perspective on regional economy, value chain length based on geographical location and possible spatial spillovers on firms' productivity. The main questions to be answered are:

1. Do Estonian firms tend to form local (within-region) networks?;

2. Does the fragmentation and value chain position of the firm depends on the region it is located in and are there any effects of geographical location on firm's productivity?

4.METHODOLOGY

As it was already mentioned, the main difficulty in conducting network analysis is the data collection. Longitudinal data is often unavailable and researcher usually work with a static list of existing directed or undirected interactions between either individuals or institutions. Another problem poses the reliability of the data. If roller-recall methodology is applied, there is still a high degree of uncertainty to whether the obtained data is correct. Moreover, individuals and companies might refuse to answer the questionnaires and in this case the absence of potentially significant actors in the network may limit the correctness of the network characteristics.

There exist many approaches to network analysis in the literature. The models can be classified into several categories – network evolution models (NEMs), nodal (node) attribute models (NAMs), exponential random graph models (Toivonen et al. 2009). The first class, NEMs, analyze evolution mechanisms. The main idea behind the model is that network evolution is dependent on a set of stochastic rules. These rules are determined by network structure (for example, tie strength) and define new nodes and links at each time step. The process continues until the desired number of nodes is reached or when statistical parameters in the model no longer change. The model simulates actor's behavior and how new links are formed in the network with an introduction of new actors. Another set of models, NAMs or sometimes referred to as spatial models, uses node attributes as determinants of tie existence. The concepts such as homophily, transitivity or location parameters are assumed to be the main drivers of tie formation.

Recently, the most widely discussed model has been p*class models, ERGM in particular, and its developments. The logic behind ERGM is quite simple. The existing network data (all ties between actors) is just one realization of many other potential ones. In other words, the choice of the partner actor is the decision based on certain principles (for example, a big company prefers to partner with other big companies) but theoretically there exists a set of possible actors that match the decision criteria. Therefore, the choice of one particular actor is, to some extent, random. Robins et al. (2007) describe this choice as being guided by an unknown stochastic process guided by the presence or absence of node attributes and ERG model builds all other potential connections that could have emerged from the data. Therefore, by modelling all possible likely outcomes, conclusions on the underlying criteria behind firms' decision to collaborate can be made. The idea behind ERGM is to assign each potential outcome a probability of realization and draw graphs according to the assigned probabilities. The model is considered to be a good representation of the data if a random graph drawn from the model is similar to the real one. The parameters in the ERGM are usually determined by maximum likelihood estimate from Monte Carlo Markov Chain sample that allows to account for the network state in the previous time step. The model also can include node attributes to explain the global network structure (Toivonen et al. 2009).

In the literature, many empirical studies incorporated recent developments and new approaches in ERGM.[5] Although this model seems to account for both node attributes and structural components, the problem of near degeneracy may arise while applying this class of models, especially to large networks with many actors. The term degeneracy implies that only few graphs in the distribution are assigned non-zero probabilities, which means high instability in the model and may signal that the model does not fit the data well – or in other words effects that are incorporated in the model do not explain network structure. Toivonen et al. (2009) compare the results of NEMs, NAMs and ERGM on friendship and university mail data sets. The results show that NAM produce assortative networks but non-realistic cluster coefficients. NEMs show reasonable clustering and degree distribution effects and also reflect network structure closely to the real data. ERGM is found to produce weak community structure and encounters near degeneracy problem. Proponents of the ERGM (Robins et al. 2007, Snijders et al. 2006, 2010) also emphasize the necessity to first assess whether the model is degenerate. New developments and specifications are constantly introduced in the model to mitigate the goodness-of-fit issue – alternating k-stars, alternating k-triangles, alternating independent two-paths (Robins et al. 2007).

[5] Recent network studies on ERGM can be found at: http://www.melnet.org.au/ergm

Another tool used to analyze inter-organizational networks is Multiple Regression Quadratic Assignment Procedure (MRQAP) that estimates the relation variables between two nodes and characteristics of the two nodes that share a tie. Broekel, Hartog (2011) incorporate ERGM to analyze inter-organizational network of Dutch aviation companies and compares results to Multiple Regression Quadratic Assignment Procedure (MRQAP). MRQAP can only reflect dyadic level determinants whereas ERGM account for structural level and node level determinants. As a result, while taking into account node level characteristics some of the dyad level ones become insignificant. Therefore, MRQAP can only describe pair ties but fails to account for individual node features and cannot describe the whole network structure. Again, one of the major problems with ERGM is goodness-of-fit. Therefore, it can be concluded that ERGM has the potential to fully explain the network structure at node, dyad and triad level but the ability of the model to fit the data can be a problem.

The data set used in this study is large and involves around 40 000 companies (nodes) after cleaning the data for possible misreporting. In this case, one of the problems that arises in the case of ERGM is computational time. The main aim is to establish the main inter-regional reasons for tie formation and one potential solution to this problem is to estimate a binary regression model with the probability of a tie between companies as a dependent variable. Based on previous research papers (Dhyne, Duprez 2016; Goyal, Fafchamps, Van der Leij 2008), binary model (logit) is used for estimation of connection formation probability. The results for Belgian regional structures show that distance, in fact, matters for establishing a trading relationship as well as embeddedness in the same sub-network (Dhyne, Duprez 2016).

The VAT data for Estonia is available for two consecutive years (2016,2017) and includes transactions that exceed 1000 euros. Dependent variable in the model is a tie between two nodes in 2017. Independent variables include sizes of seller/buyer, between region interaction directions, a dummy variable indicating being located in the same county and operating in the same sector, being embedded in the same sub-network, productivity, network characteristics (node degree, clustering coefficient for 2016).

$$P\left(y_{ij} = 1 \middle| x_i, x_j\right) = \Lambda\,(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{k-1,j} + \beta_k x_{k,j} + \cdots + \beta_n x_n) \quad (1)$$

where y indicates existence of a tie between two companies (i and j) and x_i includes size of firm i and j, a variable direction (as "company i is located in Central Estonia, company j is located in Southern Estonia").

The data includes only firms that shared a tie at least in one of the years and, therefore, a random sample of non-linked actors is added to the data to allow binary model estimation. Following the methodology of Goyal, Fafchamps, Van der Leij (2008) of co-authorship network analysis, the analysis is divided to account for tie formation and tie persistence. The data is divided into two estimation samples based on the following principle - companies that do not share a tie in 2016 and do not have any partnerhip in 2017 as well as companies that do not have a tie in 2016 but form it in 2017 constitute the first sample of the data. Enterprises that either have a tie both in 2016 and 2017 and those that share a tie in 2016 but do not have any partnership in 2017 constitute the second sample. In this case first data set is used for tie formation analysis and second one is used for tie persistence analysis. All node attributes (firm characteristics) for 2016 are used since it is assumed that each firm takes into account previous period information and makes partnership decision before submitting VAT declaration in 2017.

Additionally to the regression analysis of trading relationship, main descriptive statistics for networks are also analyzed. The impact of geographical proximity for firms' interaction is of particular importance. As it was mentioned previously, there are contradicting views on this issue. Although one could argue that with the development of transportation system and openness to trade geographical location no longer poses restrictions on choosing a trading partner, the importance of social and knowledge connections and better buyer-seller relationship conditions in the case of spatial proximity should not be underestimated (Dahl, Pedersen 2004; Owen-Smith, Powell 2004). By constructing a table of between-region interactions and accounting for the locations of seller and buyer in the regression model the main conclusions on the significance of geographical proximity for Estonian regions are made. Moreover, such characteristics of networks as homophily and transitivity are researched at the country level (when Estonia is viewed as a separate network). These statistics allow to make conclusions on the main tendencies of Estonian enterprise networks.

In the second part of the thesis network effects on productivity are analyzed. Firstly, regional differences in terms of value chain length and productivity are analyzed. Based on firm-level data available for 2016 the aggregated average values for each county are calculated. The length of value chain is calculated based on the methodology introduced by Antras et al. (2012) as a sum of upstreamness and downstreamness values of the firm. Value chain length is defined as follows

$$L_i = U_i + D_i - 1 \quad (2)$$

where U_i is upstreamness measure (weighted average distance to final consumer) and D_i is downstreamness measure (average number of processing operations conducted before firm *i* and firm *i* itself). The concepts of upstreamness/downstreamness and methodology of the calculations are also introduced by Antras et al. (2012). The upstreamness variable is defined as follows

$$U_i = 1 * \frac{F_i}{Y_i} + 2 * \sum_{j=1}^{N} \frac{d_{ij} F_j}{Y_i} + 3 * \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{d_{ij} d_{ik} F_k}{Y_i} + 4 * \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \frac{d_{ij} d_{ijk} d_{kl} F_l}{Y_i} + \cdots \quad (3)$$

where F_i indicates final use, Y_i is the total output and the value d_{ij} indicates the dollar amount of firm's i output needed to produce one dollar amount of firm's j output - $d_{ij} = \frac{z_{ij}}{Y_j}$. The downstreamness variable is defined as follows

$$D_i = 1 * \frac{VA_i}{Y_i} + 2 * \sum_{j=1}^{N} \frac{d_{ij} VA_j}{Y_i} + 3 * \sum_{j=1}^{N} \sum_{k=1}^{N} \frac{d_{ij} d_{ik} VA_k}{Y_i} + 4 * \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{l=1}^{N} \frac{d_{ij} d_{ijk} d_{kl} VA_l}{Y_i} + \cdots \quad (4)$$

where VA_i is value added.

The values of the chain length at the firm level are already calculated by Võrk, Unt, Varblane (2018) and incorporated in the data.

Since productivity can be measured using several approaches (total factor productivity (TFP), value added per employee, etc.) and the data set used for analysis includes value added and number of employees per company, then for the sake of simplicity in this paper the term productivity is referred to as log of value added per employee.

Since networks have become an inevitable part of the economy, the question of whether the very fact of embeddedness in the network or a cluster can increase firms' performance has been discussed (Giuliani 2006, Chang 2018). As results indicate, individual firms' characteristics play a more significant role in firm's success. Regardless of how significant results are at individual firm level, the impact of external conditions (such as networks, for example) has not been investigated much. Nevertheless, firm's innovation performance is found to be affected by embeddedness in local networks (Zaheer, Bell 2005).

There are many techniques used to analyze sub-network or community embeddedness. One of the most widely used and comparatively easy tools for community detection in large networks is Louvain algorithm (Blondel et al. 2008). Same methodology was previously used by Dhyne, Duprez (2016) on Belgium firm network. The idea behind the method lies in modularity maximization.[6] The algorithm starts from N communities where each node in the network presents a separate community and for each actor i the modularity gain of joining actors i and j is estimated. Eventually, the algorithm outputs communities where combinations of nodes yell the highest modularity gain. Mathematically, modularity can be presented as follows

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta(c_i, c_j) \quad (5)$$

where A is an adjacency matrix of interactions in a network; $A_{i,j} - \frac{k_i k_j}{2m}$ shows how strongly nodes are connected within a network compared to how they are connected in alternative random network; k_i, k_j show total weight of the links of node i and node j ($k_i = \sum_j A_{i,j}$, $k_j = \sum_i A_{i,j}$) ; $m = \frac{1}{2} \sum_{i,j} A_{i,j}$ represents total weight of the links within a network; $\delta(c_i, c_j)$ equals 1 if nodes i and j belong to the same sub-network and 0 if i and j are not in the same sub-network.[7]

Communities are analyzed in terms of their geographical and sectoral heterogeneity. To evaluate the industrial heterogeneity within communities, Herfindahl-Hirschman index is applied (Rhoades, 1993):

---

[6] The term modularity was introduced by Newman and Girvan (2004).
[7] For more information on Louvain method refer to Blondel et al. (2008).

$$HHI_j = \sum_{i=1}^{n} (S_i)^2 \quad (6)$$

where S_i is a share of sector i in community j.

Louvain algorithm is rather fast with large networks and according to robustness checks exhibits the most stable and high-quality results (Blondel et al. 2008). Moreover, the comparison of the results of Louvain method and other community detection methodologies such as Hierarchical cluster, Fast Greedy, X-Means on large Twitter data set prove the superiority of Louvain method (Deepak, Jurek-Loughrey 2018). In order to evaluate each community significance Wilcoxon rank sum test is used.

5.DATA

The data used in this thesis is Estonian VAT tax declaration data for 2016-2017. The data set includes firm-to-firm transactions that exceed 1000 EUR.

The main benefit of this data set for network analysis is the high degree of reliability. Research conducted on the basis of questionnaires methodology (Weterings, Boschma 2009) allows to distinguish between knowledge or business ties but in turn involves such risks as the absence of several important actors. In this case, VAT data provides clear firm-to-firm sales amounts and all existing ties are certain. Submitting tax declarations is required by law, therefore, all firms are obliged to report transactions over a year. Although it can be argued that the absence of transactions less than 1000 EUR poses restrictions on the set of actors as well, in this paper we assume that inter-organizational networks that are based on higher amounts of transactions are more likely to be long-term impactful interactions that are crucial for regional analysis. Additionally, business registry data is used to account for ownership type, official location, etc. Overall, for further productivity and a more thorough network analysis data on the number of employees, value added, length of value chain (based on the results by Võrk, Unt, Varblane 2018), assets, exports, sales volume to other companies, ownership type (foreign, municipality, domestic, state-owned), enterprise type (state-owned or educational), company’s location (county location) and industry (EMTAK2 classification) are added to the data.

The data set is informative but involves several potential drawbacks and therefore preliminary data cleaning is done in order to avoid any data misspecification issues. First of all, the industry division is done by EMTAK2 classification. One potential issue that can pose obstacles with regional interaction research is the location (address) information. If the enterprise is officially located in Tallinn but has subsidiries all over Estonia, then the analysis of regional connections can be biased. To mitigate potential drawbacks of legal/actual firm address misspecification, the data is filtered to include only those industries that are highly unlikely to change their legal address very often and are more likely to have a more concentrated production. The following industries are excluded from the analysis: Electricity, gas, steam and air conditionning supply (D); Wholesale trade, except for motor vehicles and motorcycles (G46); Retail trade, except for motor vehicles and motorcycles (G47); Postal and courier activities (H53); Activities of head offices, management consultancy activities (M70); Rental and leasing activities (N77). Other sectors may also experience such issue but to a smaller extent that is acceptible for this study.

Second, the official NUTS (Regional Classification in European Union) Estonia is divided into 5 geographical units. Nevertheless, the structure of economic activity varies across counties. For example, the number of industial enterprises for Tartu differs from the other Southern Estonia counties and resembles Harju county. Moreover, Tallinn and Tartu are major R&D and educational centres (Tallinn Technology University, University of Tartu) and Tallinn also acts as a major connection centre (Tallinn Port). To separate the effects of these two centres, the division is done as follows – Tallinn and Tartu are separate geographical units, other counties are analyzed as whole geographic unit.

Third, as Dhyne, Magerman, Rubinova (2015) construct Belgian inter-firm network 2002-2012 and emphasize that due to the usage of raw declarations certain missrepoting issues should be accounted for. In this paper, we also account for some of these potential issues. First of all, transactions value are non-negative, therefore any negative values in the data are due to wrongly reported data. Moreover, 19 547 observations do not contain any information on value added in accounts data and therefore are not included. Second, observations where transaction value exceeds total sales of seller and total input of buyer are excluded from the data set due to possible missreporting mistakes. Approximately 29 397 observations for sellers and 91 827 observations for buyers are excluded due to possible missreporting issue. Around 516 observations also do not include any information on the number of employees or reported zero number of employees. For these observations this measure is taken as 1, assuming that at least 1 employee must work at the firm and fill in the declaration data.

And forth, annual accounts, VAT declarations and business registry data together present a so-called unbalanced data set. When the data is merged into one full data set, some observations do not have all information (for example, there is no data from business registry on location, etc.). Again, as Dhyne, Magerman, Rubinova (2015) point out the possible reasons for this might be that these observations present micro-enterprises or companies that do not have to report documentation. Overall, once the transaction data is cleaned there are approximately 40 000 – 45 000 distinct companies each year and about 300 000 connections of firms. Since a fraction of the observations is deleted from the data due to possibly missreported or incorrect data, the analyzed network is smaller and more concentrated. Nevertheless, the data cleaning is essential to avoid possible shifts in the results because of the incorrect declaration data.

Additionally, for the analysis a dummy variable reflecting whether the seller and buyer operate in the same sector, a dummy variable indicating being located in the same county and being embedded in the same sub-network (according to Louvain algorithm results) are added to the data. Network characteristics are also included in the data set at the node level (node degree, clustering coefficients, etc.).

6.RESULTS

As it has been described previously, the main data set used includes transaction data between firms for 2016 and 2017. Table 1 presents data on the number of nodes (companies) and edges (transactions) between actors in the network for 2016 and 2017. Approximately 56% of the firms are located in Northern region, 21% in Southern, 11% in Western, 8% in Central and 4% in Northeastern Estonia.

Table 1. Summary statistics on the number of nodes and edges in Estonian network for 2016-2017.

2016	
Nodes	43 524
Edges	308 645
2017	
Nodes	48 488
Edges	349 780

Source: own calculations

By aggregating observations based on the region of location, the data of interactions between geographic units in Estonia is presented in Table 2 and Appendix 2.

Table 2. Summary statistics on the frequency of interactions between geographic units in Estonia, 2016.

	Customer's location						
Supplier's location	Central	North-eastern	Northern	Southern	Tallinn	Tartu	Western
Central	**37%**	3%	12%	8%	30%	4%	6%
North-eastern	6%	**46%**	8%	5%	28%	4%	3%
Northern	7%	2%	**24%**	8%	47%	5%	7%
Southern	6%	2%	7%	**42%**	21%	16%	6%
Tallinn	5%	2%	15%	6%	**60%**	5%	6%
Tartu	4%	2%	7%	26%	25%	**33%**	4%
Western	6%	1%	10%	8%	28%	4%	**43%**

Source: own calculations

It is notable from the results that Central, Northeastern, Western and Southern (except for Tartu) geographic units have a similar pattern of interacting mostly inside its own region and with Tallinn and other Northern counties. Tartu as a separate unit interacts mostly within itself and with other counties located in Southern Estonia which can be partially explained by Tartu being a center of research in Southern Estonia (the location of the University of Tartu) as well as a southern biggest city in Estonia. As it was expected, Tartu county has its own developed network of enterprises and therefore the number of interactions with Tallinn is lower than within its own geographical unit. Enterprises located in Tallinn also tend to have a more closed network system and interact mostly inside Tallinn and other Northern Estonia counties. Overall, the results are in line with expectations and Estonian economic statistics. The main conclusion is that geographical proximity is important for firms' interactions that is in line with previous findings by Dahl and Pedersen (2004); Owen-Smith and Powell (2004).

Table 3. Summary statistics on the frequency of interactions between companies of different ownerships, 2016.

	Customer's ownership			
Supplier's ownership	State-owned	Municipality	Domestic	Foreign
State-owned	**5%**	3%	66%	26%
Municipality	3%	**6%**	71%	20%
Domestic	1%	1%	**83%**	15%
Foreign	1%	1%	79%	**20%**

Source: own calculations

According to Table 3, companies of all ownership types interact mostly with domestic companies or foreign while state-owned or municipality enterprises are less involved in the network. Such results show the homogeneity of the network interactions in terms of ownership type of firms.

Previous findings by Javorsik (2004) reveal that the presence of foreign affiliates might have possible implications on the performance of domestic enterprises due to possible spillover effects. Based on foreign direct investment data (FDI) for Lithuania it has been established that productivity is positively correlated with the number of international contracts. The analysis of VAT data for Estonia shows that there is a small positive correlation between the number of international suppliers and firm level productivity and virtually no correlation between productivity and the number of foreign buyers. Therefore, productivity spillover effects from foreign enterprises are not pronounced in Estonia based on VAT data.

The statistics on average performance of different ownership types shows that the most productive are state-owned enterprises followed by foreign companies. Domestic and municipality companies are comparatively less productive. A more detailed regional distribution of productivity of different ownership types of companies can be found in Appendix 3. As expected, in Tallinn, Northern Estonia and Tartu firms of all ownership types have higher productivity. In Northeastern region municipality enterprises are comparatively more productive than in other regions.

Next, network specific characteristics are analyzed. Figure 1 summarizes node degrees in the network. Node degree value shows how many partners a firm has on average each year. Overall, the degree distribution shows that most firms in Estonia have less than 100 distinct connections per year and approximately 50% of companies have less than 25 reported partnerships. There exists a certain fraction that has up to and above 1000 ties per year but it is comparatively small and is mostly present in Tallinn, Tartu and Nothern region. Overall, network node distribution is rather fat-tailed, which is a common characteristic of large networks (Blondel et al. 2008). Country and regional node degree distribution can be found in Appendix 4, Appendix 5 and Appendix 6 respectively.

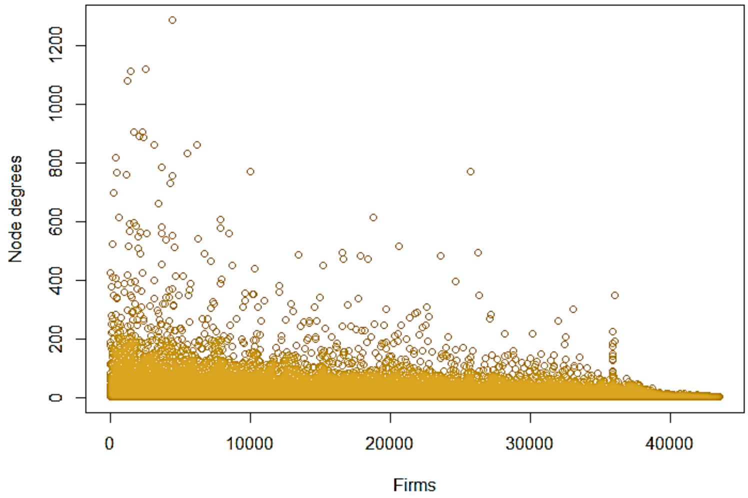


Figure 1. Statistics on node degree distribution in Estonian enterprises' network, 2016.

Source: own calculations

Based on the industrial classification, sectors with highest numbers of connections per year are "Media services", "Financial activities", "Insurance activities", with the lowest number – "Social work activities without accommodation".

Beckman, Haunschild (2002) have shown that there is a positive correlation between a diverse portfolio of partners and productivity of the firm. Figure 2 illustrates the case of Estonian enterprises based on available transaction and annual accounts data for 2016.

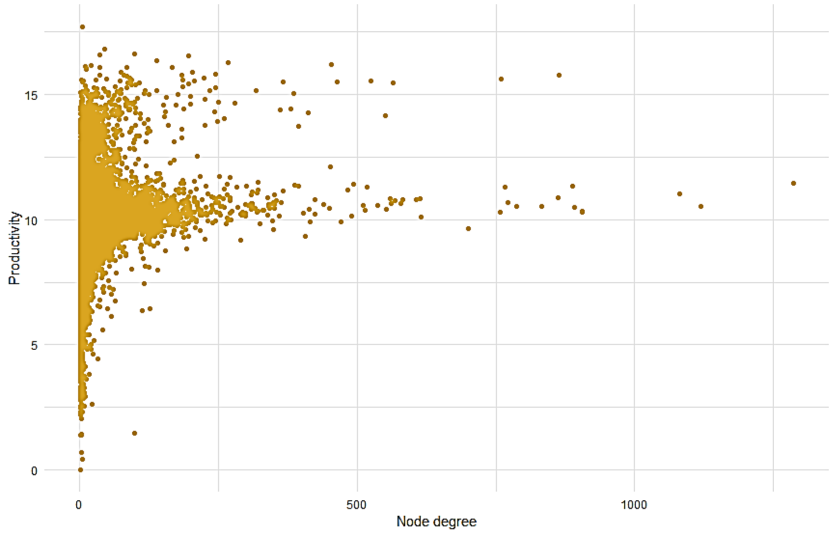


Figure 2. Relationship between productivity (log scale, y-axis) and number of connections per year (x-axis), 2016.

Source: own calculations

It is noticable that there is no exact pattern of productivity and node degree dependence. There exist highly productive firms with few connections and firms of average productivity with many ties per year. The same pattern is present and stable across all industries and regions. The analysis for all industries shows that the correlation is rather weak and not well-pronounced but, on average, the variation of productivity is higher for companies with comparatively fewer number of partners. It shows that, although, having a diverse portfolio of partners does not directly improve firm productivity but is still associated with better firm performance.

Clustering coefficient and directed density for the whole country network are rather low – 0.022 and 0.0001 respectively. Firm level clustering coefficient distribution is skewed and there exists a large fraction of enterprises with zero clustering, which can be caused by missing values or weak interconnection within a network. Such low values indicate that Estonian network of enterprises is less concentrated for the whole country - firms tend to form smaller concentrated close networks with a comparatively small number of companies that are spatially proximate to them and have similar characteristics or structure.

Additionally, by taking a closer look at assortativity statistics[8] of Estonian network presented in Figure 4, it can be inferred that Estonian enterprises have quite diverse patterns. Horizontal axis shows the number of connections a certain actor in the network has while the number of connections of its nearest neighbor within a network is shown on a vertical axis. There exists a fraction of vertices of low node degree (i.e. those that do not have many interactions in the network) that tend to interact with higher degree vertices. In other words, firms that do not have a very extended network tend to link with companies that have many partners. At the same time, most firms with higher node degree tend to choose partners with similar or average nodes degrees. All in all, results are economically reasonable. One possible explanation for the links between low and high degree vertices can be in the difference between the sectors the enterprises are operating in. Currently, there is no distinction by industries, therefore, it can be assumed that some large enterprises with concentrated production chain may outsource non-core activities from other firms that specialize in one particular operation field and thus have many other ties inside the network. Overall, the results show that Estonian firms exhibit a homophily pattern and interact with the actors that have similar characteristics (size of the firm, node degree, etc.).

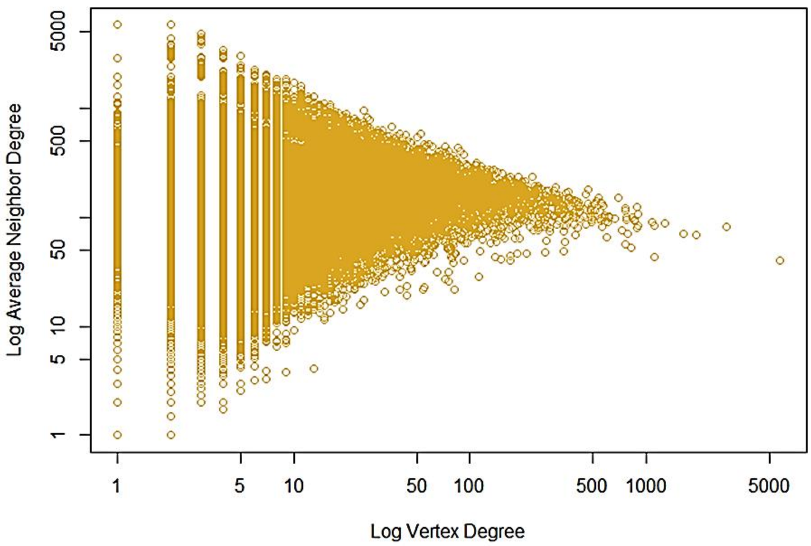


Figure 3. Degree assortativity of Estonian enterprises' network (log-log scale).

Source: own calculations

[8] Same as homophily principle introduced previously in the methodological part.

For community detection within the whole Estonian network Louvain algorithm is applied. The choice of the methodology is due to the computational time and applicability reasons. Previously this algorithm has beed used by Dhyne and Duprez (2016) for Belgian network. One important notice is that the Louvain community detection method defines firms that have mutual connections and also enterprises that are indirectly connected through mutual partners. In other words, the concept of close triangles ("friends of my friends are my friends") is also taken into account. Firms that are identified as a part of the same community have above average probability of partnership. Companies that are less interconnected within a network are treated as separate communities - each firm as an individual community.

Louvain algorithm is applied to the network data set of Estonia for 2016. According to the results, there are 216 detected communities of firms. The hierarchy and structure of these communities vary greatly. To sort the most significant of the subnetworks Wilcoxon rank sum test is applied. The result shows that 23 of those communities are statistically significant at the 0.01 confidence level. Further, the communities are sorted to include only those with a size at least 5. It is done in order to exlude single nodes and pairs of firms that were identified by the algorithm as separate communities. The remaining 12 subnetworks are included in the transaction data set as a dummy variable in order to indicate whether the interaction of firms occurs within the same community. Overall, firms that are part of identified communities generate 97% of average value added in the node data set (Appendix 7). Communities are ordered from 1 to 12 by Louvain algorithm and are presented in Table 4.

Table 4. Communities of Estonian firms in 2016.

Community	Size	Geographic distribution
Community 1	26	Tallinn – 30% Southern – 27% Tartu – 27% Central – 8% Northern – 8%
Community 2	6543	Tallinn – 42% Tartu – 7% Nothern – 17% Northeastern – 4% Southern – 12% Western – 10%

		Central – 8%
Community 3	653	Tallinn - 42% Western - 15% Southern - 12% Northern - 9% Tartu - 9% Central - 6% Northeastern - 6%
Community 4	4556	Tallinn - 43% Northern - 19% Northeastern - 13% Southern - 7% Western - 7% Central - 7% Tartu - 4%
Community 5	8712	Tallinn – 37% Tartu – 11% Nothern – 16% Northeastern – 2% Southern – 16% Western – 9% Central – 8%
Community 6	11 648	Tallinn – 64% Tartu – 7% Nothern – 12% Northeastern – 2% Southern – 5% Western – 6% Central – 3%
Community 7	49	Tallinn - 76% Northern - 8% Southern - 8% Tartu - 4% Central - 2% Western - 2%
Community 8	1211	Western - 52% Tallinn - 24% Southern - 8% Northern - 8% Tartu - 3% Central - 3% Northeastern - 2%
Community 9	9	Tallinn – 56% Northern – 33% Northeastern – 11%
Community 10	5766	Tallinn - 25% Southern - 24% Northern - 13%

		Central - 13% Western - 12% Tartu - 7% Northeastern - 6%
Community 11	2900	Southern - 36% Central - 23% Western - 16% Tallinn - 8% Northern - 8% Tartu - 6% Northeastern - 3%
Community 12	7	Tartu – 86% Southern – 14%

Source: own calculations

The results show that the largest communities of firms in Estonia are quite geographically distributed. It shows that despite the significance of geographical proximity, firms in Estonia present one big connected network or, in other words, “small world”. Such result is reasonable since Estonia is a small economy and companies have a high likelihood of being directly or indirectly connected via mutual partners. The graphic geographic distribution of the three largest communities and geographical concentration of communities by regions can be found in Appendix 8 and Appendix 9 respectively. Moreover, the fraction of domestic firms inside communities is the highest in all three clusters (above 90%). Although such results might be a consequence of domestic firms being the largest proportion of the firms’ VAT declaration data set, another possible explanation can be that foreign enterprises are simply less intergrated in the local network. Moreover, foreign-owned companies are more likely to be a part of a foreign network system and, therefore, interact mostly with enterprises located outside Estonia.

The heterogeneity of communities can be described by the industrial presence. The sectoral distribution of the communities is extensive. Each community includes many industries (based on EMTAK2 classification) and it can be assumed to be a consequence of embeddedness in value chains. One of the main reasons behind the business interactions is the fragmentation of production. If parts of a sub-network are firms that are involved in the production chain and they additionally outsource some non-core activities from small specialized companies then all these actors in the community are indirectly connected to each other. As a consequence, there is a large diversification of sectors within each community. The largest three communities are Community 2, Community 5, Community 6. Community 2 consists mostly of enterprises operating in “Wholesale and retail trade and repair of motor vehicles and motorcycles” sector (27% of the community), Community 5 – “Specialised construction activities” (27% of the community), Community 6 – “Real estate activities” (11% of the community). Partially due to the large sizes, the diversification of industries is high in these communities. Community 2 combines comparatively large firms in Tallinn and Northeastern region and smaller companies located in Southern and Western regions. In Community 5 larger firms are located in Central region and smaller enterprises are in Northeastern region. Community 6 unites companies with approximately equal sizes (log of the number of employees). Table 5 shows Herfindahl-Hirschman indices of industrial concentration for three largest communities. The second column shows indices based on the number of enterprises operating in each industry. The indices in the third column show concentration of each industry in communities based on the total sales of the companies. The results differ based on the methodology used for calculations. The values for Communities 2 and 6 show low concentration of sectoral presence in communities and Community 5 has moderate degree of industrial concentration based on the number of enterprises operating in each industry within communities (overall concentration of industries in the data is approximately 0.043 or 430). The results based on total sales of firms show that industrial distribution is more concentrated in Community 6 and Community 5 rather than in Community 2. Such results demonstrate that in Community 6 (the largest one in the data) there is a more diversed industrial distribution – there are more industries present with approximately equal number of operating companies, but the volume of total sales for these enterprises is more concentrated.

Table 5. Herfindahl-Hirschman index of industrial concentration within three largest communities.

Community	HHI (number of firms)	HHI (total sales)
Community 2	0.10 (1000)	0.05 (500)
Community 5	0.13 (1300)	0.154 (1540)
Community 6	0.05 (500)	0.303 (3030)

Source: own calculations

Smaller communities are less geographically distributed and more sectorally concentrated. Three communities involve less than 100 firms. For example, Community 12 consists of only 7 firms located in either Tartu or Southern Estonia and involves “Computer programming, consultancy and related activities” and “Information service activities” sectors. Community 1 with a size of 26 members mostly involves firms operating in “Education” sector (46% of firms) and located in Tallinn, Tartu and Southern Estonia. Community 7 with 49 members consists mostly of enterprises from “Transportation” sector (69% of firms) and is largely present in Tallinn (75% of firms). Community 9 is comprised of enterprises located in Tallinn, Northern or Northeastern regions and combines firms operating in “Security and investigation activities’ (44% of firms) and “Employment activities” (33% of firms).

Productivity distribution of firms shows that overall, most of firms in communities are equally productive, except for Community 9 and Community 12 where distribution is rather unequal. Community 12 is small and includes only 6 members, therefore, the distribution is possibly due to the presence of a few comparatively less productive firms in community. Appendix 7 shows that Community 4 and Community 12 have comparatively higher weighted average value added generated within community, while Community 1 has the lowest.

Analysis of communities at regional level shows that above 90% of companies in each geographical region can be classified as separate units. In other words, these companies do not belong to any of the detected communities, which implies that the firms simply have, on average, a small number of connections and are not tied to other enterprises within a network. The most integrated sector across all regions is “Construction” and its related activities. In Southern and Central regions there are communities with large concentration of “Manufacture of wood and products of wood”. Tallinn and Tartu have small communities that involve only “Education” and “Scientific research and development” sectors.

All in all, community detection algorithm shows that most of the firms are more or less connected to each other through the network system. Nevertheless, inside the whole Estonian country there are small concentrated geographic units that have its own developed structures and interact mostly within regional boundaries.

The probability of tie formation and tie persistence between enterprises in Estonian network are analyzed by logit regression. The dependent variable is a tie between firms in 2017 (0 when there is no tie in 2017 and 1 if there is a tie in 2017). First model analyzes tie formation process. The data set includes only observations where companies either do not share a tie in 2016-2017 or formed a tie in 2017. Since the data only includes companies that actually have a tie at least in one of the years, then a sample data set with all possible combinations of firms is generated. A random sample of firms' combinations is included in the transaction data with zero as a tie value. Independent variables in the regression include sizes of supplier and buyer (log of the number of employees), node degree of supplier and buyer, clustering coefficient of supplier and buyer, variable indicating the direction of the tie (for example, firm *i* in Nothern Estonia, firm *j* in Southern Estonia) and dummy variables such as being located in the same county, operating in the same sector, embeddedness in the same sub-network. The values used for independent variables are for 2016 since it is assumed that firms' decision to cooperate is based on previously observed characteristics of potential partners and is made before the VAT declaration for 2017 is submitted. Second model analyzes tie persistence between companies. The data includes firms that either share a tie both in 2016 and 2017 and those that share a tie in 2016 but no longer have a tie in 2017. Same dependent variables are used. Independent variables are consistently included in the model. Full regression tables can be found in Appendix 11 and 12.[9]

First, the model that estimates the effects of firms' sizes is analyzed. The results show that sizes of both supplier and buyer have positive correlation with the probability of establishing a trading relationship. Then such effects as operating in the same industry, the direction of the connection and being located in the same county are analyzed. The results prove that regional distances are statistically significant for the emergence of a tie at 0.01 level. Being located in different counties decreases the probability of a trading relationship while being located in the same county increases the chances of firms' interaction. The case of Tartu and Southern Estonian counties differs a little. If firm *i* is located in Tartu and firm *j* in Southern counties (or *i* in Southern, *j* in Tartu), the

[9] The model is checked for autocorrelation; the effects of location are separated for seller and buyer; model performance is estimated by AIC and McFadden R^2.

probability of trading is higher, which means that firms in this region have a more developed and well connected network.

Next, network characteristics such as node degree and embeddedness in the same sub-network are added to the model. The results show that firms with higher number of connections have higher probability of tie formation. In other words, for enterprises with many connections it might be the result of outsourcing non-core activities from other companies or, on the contrary, performing these activities for other firms. Being part of the same community also increases the probability of a partnership almost by 20 percentage points. Such result can be explained by the effects of social and management interactions. If firms are even indirectly connected through a mutual partner, then it is likely that managers of these companies have mutual acquaintances or colleagues. In such case, the effects of such intangible asset as "trust" can influence tie formation decision for firms but can not be observed directly in the data. Models that include geographical location of seller and buyer perform comparatively better than models that only account for firm specific characteristics (based on AIC and McFadden R^2 criteria) that leads to a conclusion that differences in seller-buyer locations are important for establishing a trading relationship between companies.

Second analysis introduces tie persistence evaluation. In this case, it is noticeable that the results are slightly different. The effects of a geographical location are less significant while firm specific characteristics (such as size, industry, etc.) are still important. Such outcome is also expected since the role of already established connections and the effects of partner's reliability can not be underestimated. Low values of McFadden R^2 also show that only observed characteristics such as size or location do not explain the chances of continuing a trading relationship. Apart from geographical proximity and mutual partners, social and trust aspects should be accounted for while evaluating the probability of continuing trading relationship. The history of previous successful interactions positively influences the chances of partnership and an established connection between firms is less affected by any geographical proximity changes. Overall, the results seem to be in line with research hypotheses and expectations.

Since it is established within the framework of this thesis that spatial proximity matters for enterprises' interaction, next issue to be considered is current regional performance and regional differences in Estonia. By taking a closer look at Estonian regional characteritics regarding value chain length as well as productivity it can be concluded that there exist several significant differences across Estonian counties. Figure 4 shows regional differences in productivity in Estonia. It is noticeable that Tallinn and Tartu reasonably have the highest productivity, while Ida-Viru county has the lowest. A more detailed information about productivity and value chain length across regions can be found in Appendix 13.

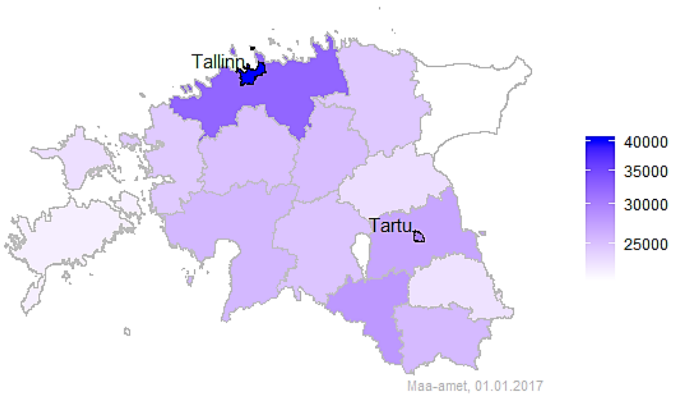


Figure 4. Average value added per employee by Estonian counties, 2016-2017.

Source: own calculations

Figure 5 shows regional distribution of value chain length. Value chain length is calculated based on the methodology introduced by Antras et al. (2012).

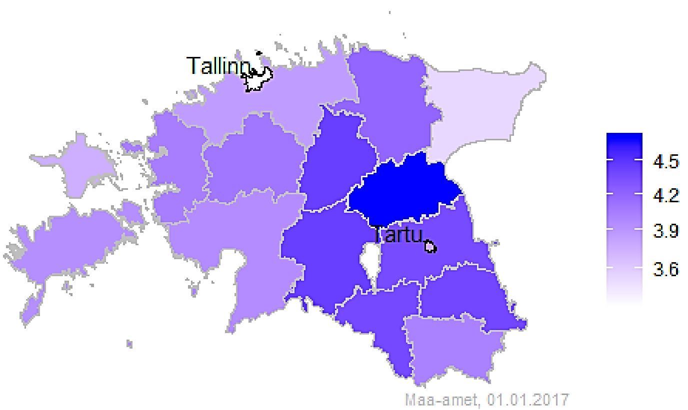


Figure 5. Average value chain length by Estonian counties (weighted by value added)[10].

Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

[10] The table with exact regional values is in Appendix 13.

The lowest values of chain length are in Tallinn, Tartu and Ida-Viru county while the longest chain length is in Jõgeva county. Overall, average value chain length is comparatively longer in Southern and Central regions, which implies that production is somewhat more fragmented in these counties.

Specialization of Tallinn and Tartu are “Wholesale and retail trade, including repair of motor vehicles”, “Professional, scientific and technical activities” and “Construction” that on average have smaller value chain length. Also, it might be a consequence of a more concentrated production chain due to the developed scientific and technical activities sector. Tallinn shows higher degree of non-local interactions (Table 2, Appendix 2). The highest value chain length is in Jõgeva county that specializes mostly in “Agriculture, forestry and fishing” sector. “Forestry” sector has comparatively higher value chain length and therefore can partially explain the average value for the region.

Nevertheless, the results indicate that in Ida-Viru county there is the lowest value of productivity and the lowest value of chain length. Such result can be due to the industrial structure in the county and due to low network embeddedness. According to Statistics Estonia data, the highest proportions of enterprises in the county belong to “Wholesale and retail trade, including repair of motor vehicles”, “Construction” and “Other service activities” sectors that on average have less fragmented production chain, therefore the average value for the region is also lower than for other counties.[11] Moreover, this region shows also low network embeddedness (firms have comparatively lower numbers of connections with other firms per year) and it can be assumed to be a consequence of more concentrated production stages that also influences overall value chain length. Appendix 2 illustrates that Northeastern region is less embedded in between-region interactions - most of the input is provided by local sellers. The average value chain length differs for each industry across counties that leads to an assumption of the presence of spatial effects.

[11] Conclusions on the value chain length are made based on upstreamness values calculated with Antras et al. (2012) methodology using input-output tables from Statistics Estonia website. The table can be found in Appendix 14.

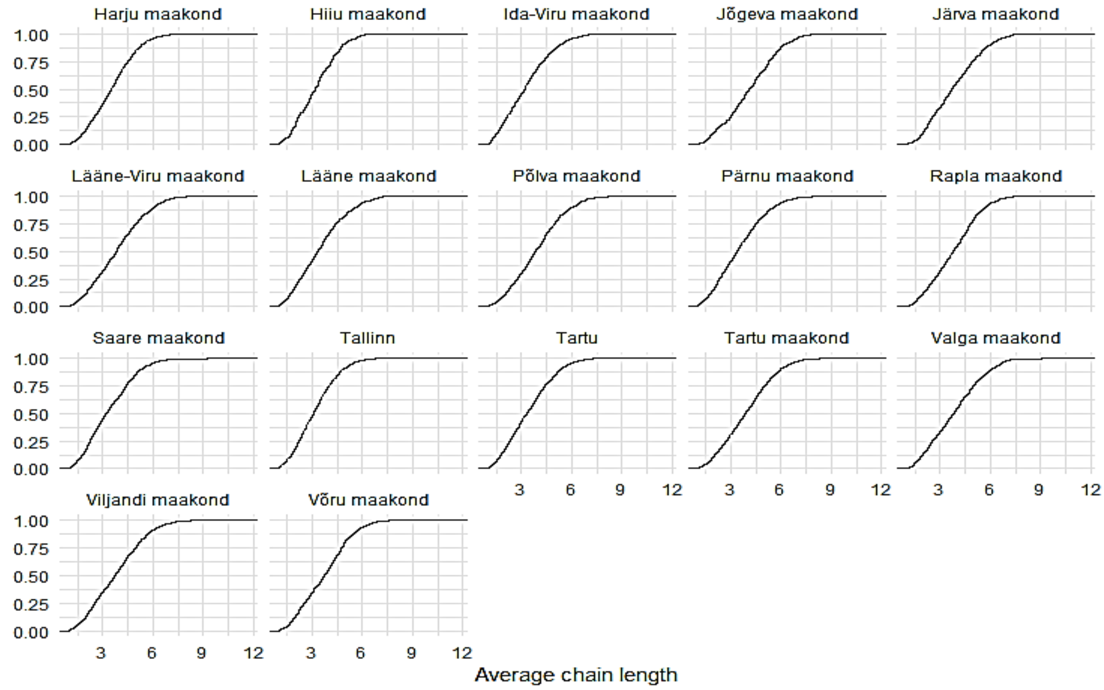


Figure 6. Value chain length cumulative distribution by Estonian counties, 2016-2017.[12]

Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Regional distribution of chain length (Figure 6) shows that the number of production stages varies greatly within each county. For example, Lääne-Viru, Pärnu or Jõgeva counties have comparatively lower variation of the number of production stages, but overall average chain length is higher than in Tallinn or Harju county. Partially such result can be explained by regional specialization. Lääne-Viru, Pärnu and Jõgeva counties have similar industrial distribution with most firms operating in sectors with comparatively high value chain length, while the situation in Tallinn and Harju county is the opposite. Another possible option can be that the variation of value chain length within one industry is higher in Tallinn and Harju, while firms in Lääne-Viru, Pärnu and Jõgeva counties have approximately stable value chain structure within industries. Nevertheless, it is possible to track the dynamics of value chain length across different counties with more historical data to determine the trends in regional development over time.

[12] Overall, value chain distribution can be found in Appendix 15. Based on the results, around 90% of firms have value chain length lower than 6.

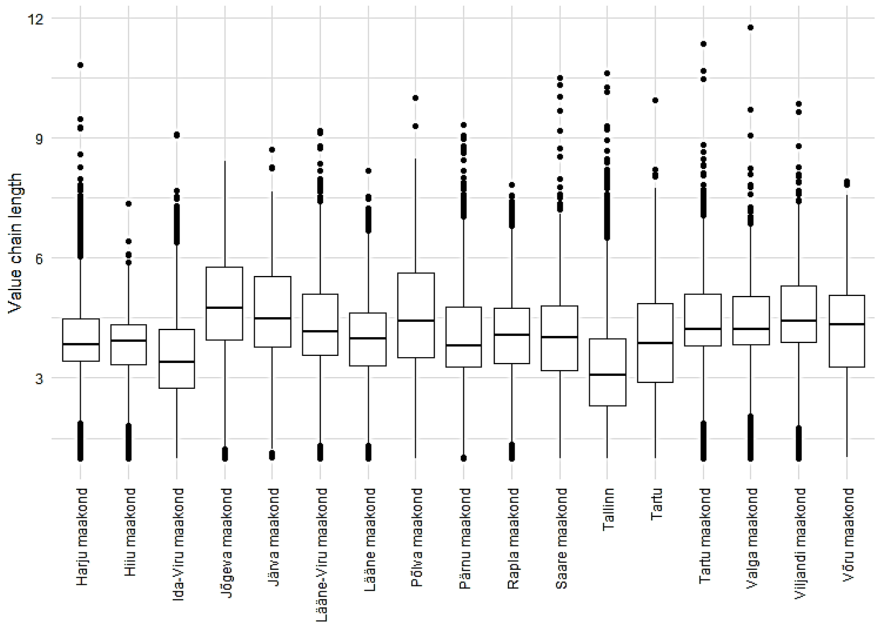


Figure 7. Regional distribution of value chain length, 2016-2017.

Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Based on Figure 7, regions with comparatively higher average value chain lengths are Jõgeva, Jarva and Põlva counties. Overall, it can be concluded that Northern and Western parts of Estonia have lower values of chain length than Southern (except for Tartu) and Central parts. Regional effects on value chain length are not pronounced and might be caused by industrial and productivity differences across counties.

The most productive geographical units are Tallinn and Tartu that have smaller chain length values and production in these regions is less fragmented. The structure of detected sub-networks shows that the largest proportion of each community is located in Tallinn and overall volume of community concentration is the highest in Tallinn. Productivity analysis shows that embeddedness in communities is positively correlated with productivity within a region.[13] Regional productivity and value chain length relationship is not well pronounced. There is a small positive correlation between value added per employee and value chain length in Central, Southern and Western regions. For Northern and Northeastern regions, the relationship between chain length and productivity is either not pronounced or negative.

[13] Productivity analysis includes linear regression analysis (with robust standard errors and autocorrelation check). Control variables such as firm size, industry, number of steps to exports are taken into account. Effects of variable "region" are viewed only in terms of correlation with productivity.

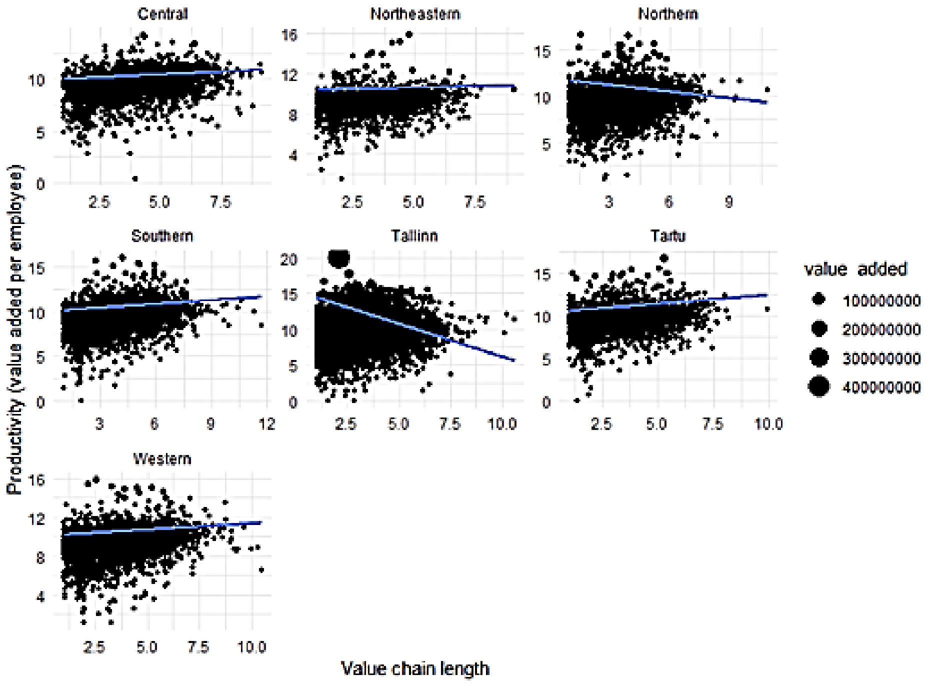


Figure 8. Regional relationship between value added per employee and value chain length, 2016-2017.

Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Tallinn and Harju county have seemingly negative relationship between value added per employee and value chain length. Such trend can be a consequence of the presence of companies with low value chain length but significantly higher value added (most likely firms operating in service sector). Other regions have a weak positive relationship between value chain length and value added per employee. Overall, the results show that most of the variance of productivity in regions cannot be explained purely by value chain length and firm specific characteristics.

There is no clear consensus in the literature about the effects of fragmentation or concentration of value chains on economic performance. The balance between effective outsourcing strategy and minimizing risks from spatial fragmentation or strong interdependence still presents a research gap. Estonian case shows that production fragmentation is different across counties and industries but still is high for the whole economy. The influence of geographical distance shows that seller-buyer search conditions right now are limited by regional boundaries (or possibly regional trade barriers) and cost-minimizing outsourcing strategy is dependent on spatial proximity of potential trading partners. Such implications might limit knowledge or information flows across regions and existing differences in productivity might cause undesired labor movements within the country. At a country level value chains are also linked to a phenomenon called “middle income trap” (Gill, Kharas 2007). The vulnerability of Estonia to fall into such “middle income trap” can be associated with spatial economic disparities, export of mostly medium-skilled sector products, heterogeneity of education system across regions, etc. (Staehr 2015). Current results show that there are differences in the development of urban and rural regions in Estonia that should be addressed. The involvement in communities, knowledge-intensive regional specialization, low value chain length and high productivity are mostly observed in Tallinn, Tartu and Northern region, while the performance of other regions is comparatively worse. The analysis of cooperation formation processes shows that geographical distance, being a member of the same community and firm size are significant for establishing a trading relationship. Potential policy implications in this case might involve financing sectors such as “Professional, scientific and technical activities” and “Education” at cross-regional level (Staehr 2015), focusing on both innovation and specialization in high-tech industries (Gill, Kharas 2007), separating policy effects for state-owned and non-state-owned companies, encouraging joint projects and network embeddedness of high-tech enterprises as well as better integration of foreign companies within domestic network, encouraging cooperation between small businesses and encouraging trade of goods and capital across regions.

7.CONCLUSIONS AND IMPLICATIONS

This paper has illustrated main characteristics and tendencies for Estonian firms’ network and analyzed main factors that influence the emergence and persistence of trading relationship between firms. One of the important results is that spatial proximity is significant for the formation of partnership among firms. Estonian enterprises tend to interact mostly within its geographic unit and location in the same county significantly increases the chances of between-firm ties. Also, all geographic units are connected to Tallinn and Northern Estonia that have a high volume of operating enterprises and are major connection centers for firms that are more embedded in import and export operations. Tartu, as expected, has its own developed network and has fewer connections to Tallinn than to its own geographic unit. Regional differences in terms of productivity and value chain length are partially explained by sectoral structures of the regions and network characteristics. Also, network characteristics may influence productivity.

The analysis of tie formation and tie persistence shows that geographical distance plays an important role in tie formation process but is less significant when a tie between firms has already been established. Individual firm characteristics are significant for both processes and social aspects (such as attending same network events by employees, previous history of successful interactions, etc.) in tie persistence analysis might be taken into account in the future research. Embeddedness in sub-networks (or communities) also shows positive impact on average performance of enterprises and further initiatives to encourage domestic firms' cooperation should be considered.

The influence of geographical distance shows that seller-buyer search conditions right now are limited by regional boundaries and cost-minimizing outsourcing strategy is likely to be dependent on spatial proximity of potential trading partners. The encouragement of distant interactions may allow better seller-buyer conditions and more efficient partnerships and joint projects. Another issue is the impact of export orientation on the value chain length and possible spatial shocks that are associated with it. The productivity analysis shows that there is a small negative correlation between the number of foreign buyers and productivity. Possible alternatives of concentrating more high-tech production stages within economy should be considered to avoid falling into "middle income trap".

Foreign or state-owned companies perform comparatively better than domestic firms and productivity analysis shows that enterprises are more likely to have higher productivity when they have more foreign suppliers. Numerous studies focus on learning effects from interactions with foreign companies but the effects of such cooperation should be viewed at regional scale. Based on previous research and current analysis results, it can be concluded that the effects of geographical location of foreign and domestic companies as well as the efficiency of integration into local networks should be taken into account in policy making.

The scope of the research can be further extended to a more thorough analysis of regional differences and application of additional data sets. Particular issues that emerged during the research are data availability and data cleaning. VAT tax declaration is more reliable source of firms' interactions than data from questionnaires but the data set is constructed based on raw declarations that are subjects to misreporting and absence of particular data. All such data has been excluded from the analysis in order to avoid any misspecification issues but the introduction of more data provides opportunities for further research. The problem of official and actual addresses has been solved by excluding several industries from the data. This issue can also be solved by comparing employees' working addresses from TÖR (Töötamise registrisse or working registry) and comparing it to business registry data on firm location. It can allow accounting for larger network that includes all industries and companies that can have several locations within the country with more pronounced regional effects. Another important issue that can be reseached is the performance of domestic and foreign companies. Due to favorable investment and business opportunities there is quite a large fraction of foreign companies in the economy. Presence of a succesfull multinational or foreign enterprise can significantly influence the regional performance and create competitive environment for domestic firms. By using foreign direct investment for the analysis of innovation flows, the influence of foreign enterprise in the market can be assessed. Thus, there are perspectives for further regional analysis.

8.REFERENCES

Ali-Yrkkö, J., Mattila, J., Seppälä, T. (2017). Estonia in Global Value Chains (No. 69). The Research Institute of the Finnish Economy.

Amighini, A., Rabellotti, R. (2003). The effects of globalization on Italian Industrial districts: evidence from the footwear sector. Università degli studi del Piemonte orientale, Facoltà di economia, Dipartimento di scienze economiche e metodi quantitativi.

Alcácer, J., Zhao, M. (2012). Local R&D strategies and multilocation firms: The role of internal linkages. Management Science, 58(4), 734-753.

Antràs, P., Chor, D., Fally, T., Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review*, *102*(3), 412-16.

Asheim, B. T., Isaksen, A. (2002). Regional innovation systems: the integration of local ‘sticky’and global ‘ubiquitous’ knowledge. The Journal of Technology Transfer, 27(1), 77-86.

Audretsch, D. B., Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. The American economic review, 86(3), 630-640.

Audretsch, D. B., Hülsbeck, M., Lehmann, E. E. (2012). Regional competitiveness, university spillovers, and entrepreneurial activity. Small Business Economics, 39(3), 587-601.

Backer, K. and Miroudot, S. (2013). Mapping Global Value Chains. OECD Trade Policy Papers, 159.

Baldwin, R. (2016). The great convergence. Harvard University Press.

Basant, R., Chandra, P., Upadhyayula, R. S. (2011). Knowledge flows and capability building in the Indian IT sector: A comparative analysis of cluster and non-cluster locations. Indian Institute of Management.

Bathelt, H., Li, P. F. (2014). Evolutionary economic geography and relational geography. In Handbook of Regional Science (pp. 591-607). Springer, Berlin, Heidelberg.

Beckman, C. M., Haunschild, P. R. (2002). Network learning: The effects of partners' heterogeneity of experience on corporate acquisitions. Administrative science quarterly, 47(1), 92-124.

Bell, M., Giuliani, E. (2007). Catching up in the global wine industry: innovation systems, cluster knowledge networks and firm-level capabilities in Italy and Chile. International Journal of Technology and Globalisation, 3(2-3), 197-223.

Bernard, A. B., Moxnes, A., Saito, Y. U. (2015). Production networks, geography and firm performance (No. w21082). National Bureau of Economic Research.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.

Boschma, R. (2005). Proximity and innovation: a critical assessment. Regional studies, 39(1), 61-74.

Boschma, R. A., Ter Wal, A. L. (2007). Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy. Industry and Innovation, 14(2), 177-199.

Boja, C. (2011). Clusters models, factors and characteristics. International Journal of Economic Practices and Theories, 1(1).

Brass, D. J., Galaskiewicz, J., Greve, H. R., Tsai, W. (2004). Taking stock of networks and organizations: A multilevel perspective. Academy of management journal, 47(6), 795-817.

Broekel, T., Hartog, M. (2011). Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter. Papers in Evolutionary Geography, (11.07)

Coen, D., Thatcher, M. (2008). Network governance and multi-level delegation: European networks of regulatory agencies. Journal of Public Policy, 28(1), 49-71.

Dahl, M. S., Pedersen, C. Ø. (2004). Knowledge flows through informal contacts in industrial clusters: myth or reality?. Research policy, 33(10), 1673-1686.

De Backer, K., D. Flaig (2017), "The future of global value chains: Business as usual or “a new normal”?", OECD Science, Technology and Industry Policy Papers, No. 41, OECD Publishing, Paris, https://doi.org/10.1787/d8da8760-en.

Deepak, P., Jurek-Loughrey, A. (Eds.). (2018). Linking and Mining Heterogeneous and Multi-view Data. Springer.

Dhyne, E., Duprez, C. (2015). Has the crisis altered the Belgian economy’s DNA? Economic Review, (ii), 31-43.

Dhyne, E., Duprez, C. (2016). Three Regions, three economies?. Economic Review, (iii), 59-73.

Dhyne, E., Magerman, G., Rubínová, S. (2015). The Belgian production network 2002-2012 (No. 288). NBB Working Paper.

Economy review (2017). Estonian Ministry of Foreign Affairs, vm.ee/en/estonian-economy-overview (accessed 18 May 2019).

European Comission (2008). Convergence report (2008b), www.ecb.europa.eu/pub/pdf/conrep/cr200805en.pdf (accessed 18 May 2019).

Fafchamps, M., Van der Leij, M. J., Goyal, S. (2010). Matching and network effects. Journal of the European Economic Association, 8(1), 203-231.

Gill, I. S., Kharas, H. (Eds.). (2007). An East Asian renaissance: ideas for economic growth. The World Bank.

Giuliani, E. (2006). The selective nature of knowledge networks in clusters: evidence from the wine industry. Journal of economic geography, 7(2), 139-168.

Giuliani, E. (2007). Networks and heterogeneous performance of cluster firms. Applied evolutionary economics and economic geography, 161-179.

Goyal, S. (2007). Introduction to connections: An introduction to the economics of networks. Introductory Chapters.

Grossman, G. M., Rossi-Hansberg, E. (2008). Trading tasks: A simple theory of offshoring. American Economic Review, 98(5), 1978-97.

Hardeman, S., Frenken, K., Nomaler, Ö., Ter Wal, A. L. (2014). Characterizing and comparing innovation systems by different ‘modes’ of knowledge production: A proximity approach. Science and Public Policy, 42(4), 530-548.

Kesidou, E., Romijn, H. (2008). Do local knowledge spillovers matter for development? An empirical study of Uruguay’s software cluster. World development, 36(10), 2004-2028.

Kolaczyk, E. D., Csárdi, G. (2014). Statistical analysis of network data with R (Vol. 65). New York: Springer.

Kozovska, K. (2010). The role of regional clusters and firm size for firm efficiency. International Journal of Globalisation and Small Business, 4(1), 41-60.

Kraemer, K. L., Linden, G., Dedrick, J. (2011). Capturing value in Global Networks: Apple’s iPad and iPhone. University of California, Irvine, University of California, Berkeley, Syracuse University, NY. pcic. merage. uci. edu/papers/2011/value_iPad_iPhone. pdf. Consultado el, 15.

Longhi C. (2017). Cluster dynamics: learning from competitiveness cluster policy. The case of “secure communicating solutions” in the French Provence-Alpes-Cote d’Azur region. GREDEG Working Papers 2017-42, Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis.

Maskell, P., Lorenzen, M. (2003, April). The cluster and other current forms of market organization. In Regional Studies Association International Conference on ‘Reinventing Regions in a Global economy’in Pisa (pp. 12-15).

Ministry of Finance of Republic Estonia (2014). Operational Program for Cohesion Policy Funds in Estonia 2014-2020.

Nardone, G., Lopolito, A., Muscio, A. A Methodology to Evaluate Technological Cluster Policies through the Impact on Social Capital.

Newman, M. E., Girvan, M. (2004). Finding and evaluating community structure in networks. Physical review E, 69(2), 026113.

OECD (2017). Estonia: Trade and Investment Statistical Note, www.oecd.org/investment/trade-investment-gvc.htm (accessed 18 May, 2019).

Owen-Smith, J., Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. Organization science, 15(1), 5-21.

Ozman, M. (2009). Inter-firm networks and innovation: a survey of literature. Economics of Innovation and New Technology, 18(1), 39-67.

Porter, M. E. (1996). Competitive advantage, agglomeration economies, and regional policy. International regional science review, 19(1-2), 85-90.

Rallet, A., Terre, A. (2005). Proximity and localization u, Régional Studios, vol. 39.

Rhoades, S. A. (1993). The herfindahl-hirschman index. Fed. Res. Bull., 79, 188.

Robins, G., Pattison, P., Kalish, Y., Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. Social networks, 29(2), 173-191.

Robins, G., Snijders, T., Wang, P., Handcock, M., Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. Social networks, 29(2), 192-215.

Rowley, T., Behrens, D., Krackhardt, D. (2000). Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries. Strategic management journal, 21(3), 369-386.

Smarzynska Javorcik, B. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. American economic review, 94(3), 605-627.

Snijders, T. A., Pattison, P. E., Robins, G. L., Handcock, M. S. (2006). New specifications for exponential random graph models. Sociological methodology, 36(1), 99-153.

Snijders, T. A., Van de Bunt, G. G., Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. Social networks, 32(1), 44-60.

Sölvell, Ö., Ketels, C., Lindqvist, G. (2009). EU cluster mapping and strengthening clusters in Europe. The European Cluster Observatory. European Communities. Luxembourg.

Staehr, K. (2015). Economic growth and convergence in the baltic states: Caught in a middle-income trap?. Intereconomics, 50(5), 274-280.

Statistics Estonia (2018). Statistics Estonia quaterly bulletin. An overview of social and economic developments in Estonia (e-publication), www.stat.ee/publication-2018_quarterly-bulletin-of-statistics-estonia-2-18 (accessed 18 May 2019).

Sturgeon, T., Van Biesebroeck, J., Gereffi, G. (2008). Value chains, networks and clusters: reframing the global automotive industry. Journal of economic geography, 8(3), 297-321.

Suganuma, K. (2016, January). Upstreamness in the Global Value Chain: Manufacturing and Services. In Meeting of the Japanese Economic Association at Nagoya University on June (Vol. 18, p. 19).

Rallet, A., Terre, A. (2005). Proximity and localization u, Régional Studios, vol. 39.

Toivonen, R., Kovanen, L., Kivelä, M., Onnela, J. P., Saramäki, J., Kaski, K. (2009). A comparative study of social network models: Network evolution models and nodal attribute models. Social networks, 31(4), 240-254.

Turkina, E., Van Assche, A., Kali, R. (2016). Structure and evolution of global cluster networks: evidence from the aerospace industry. Journal of Economic Geography, 16(6), 1211-1234.

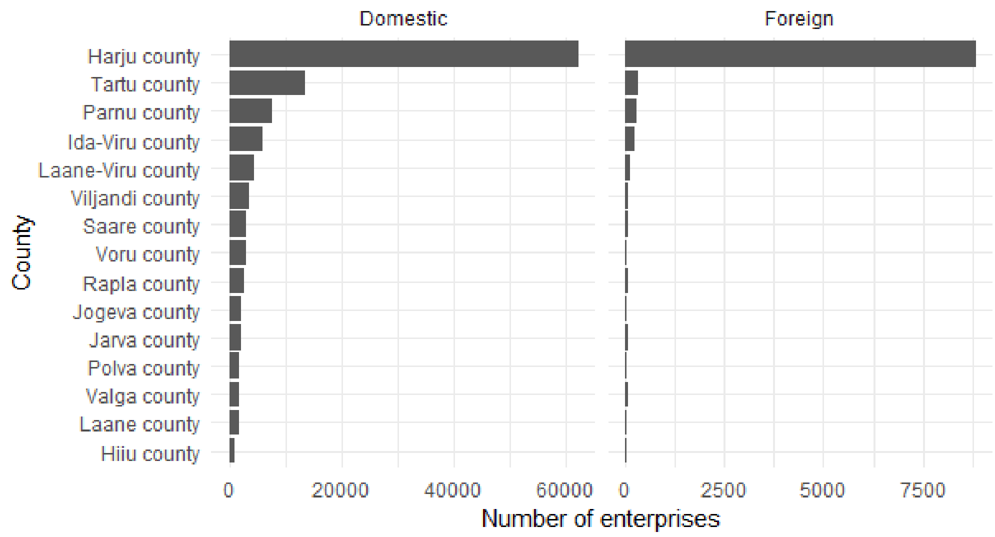
Unt T., Varblane U., Vork A. (2018). Eesti ettev˜otete osalemine ja positsioon globaalsetes ja lokaalsetes v¨a¨artusahelates. Uurimisraport.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. American sociological review, 674-698.

Weterings, A., Boschma, R. (2009). Does spatial proximity to customers matter for innovative performance?: Evidence from the Dutch software sector. Research Policy, 38(5), 746-755.

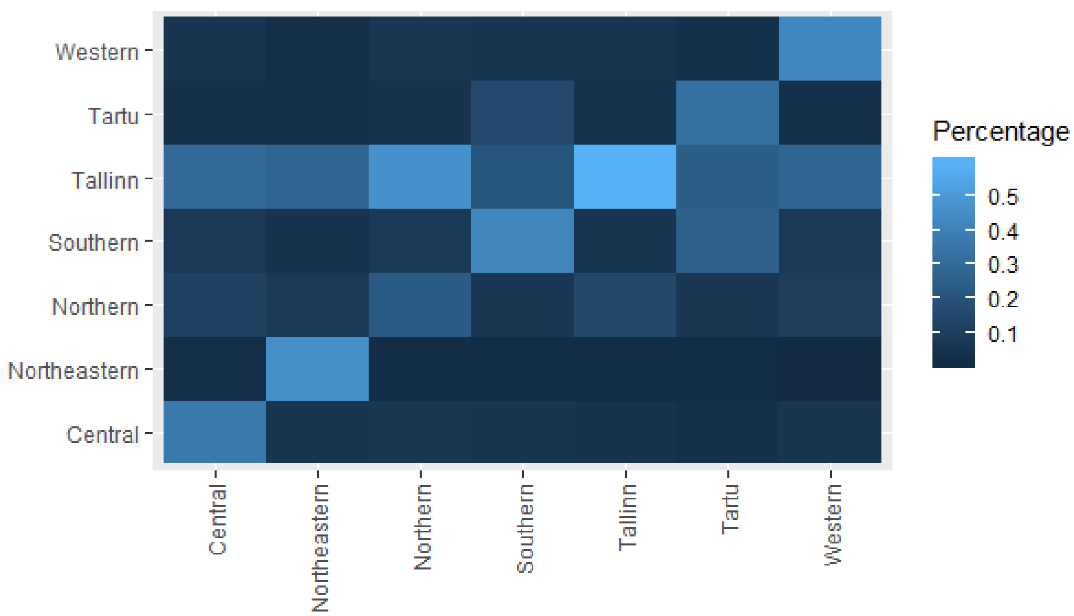
Zaheer, A., Bell, G. G. (2005). Benefiting from network position: firm capabilities, structural holes, and performance. Strategic management journal, 26(9), 809-825.

Appendix 1. Number of enterprises by county (domestic or foreign-owned), 2017.



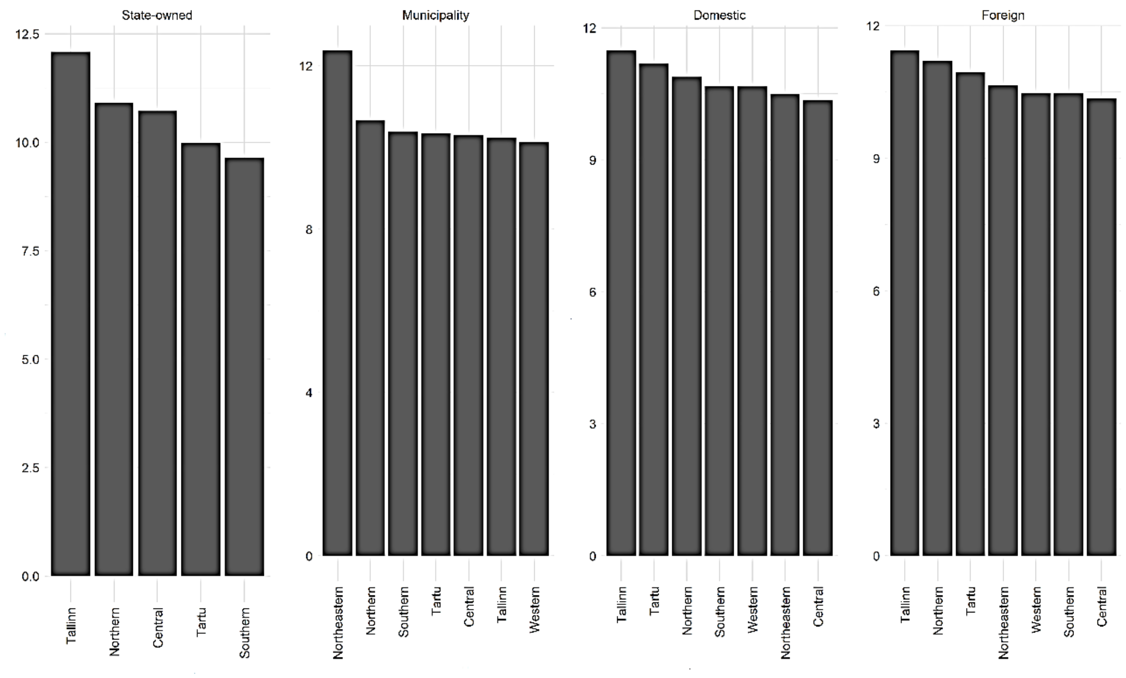
Source: own calculations based on Statistics Estonia data

Appendix 2. Summary statistics on the number of interactions between regions, 2016.



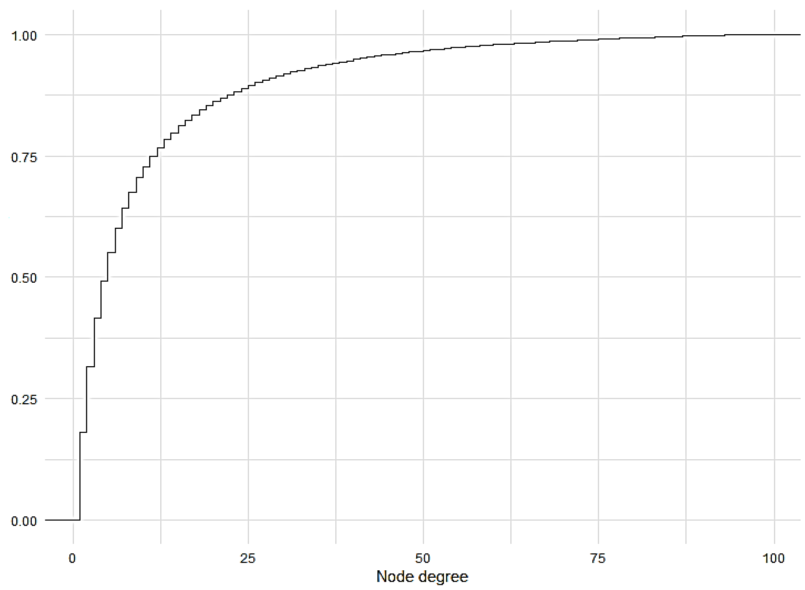
Source: own calculations

Appendix 3. Summary statistics on the productivity of companies (log scale) of different ownership types across Estonian regions, 2016.



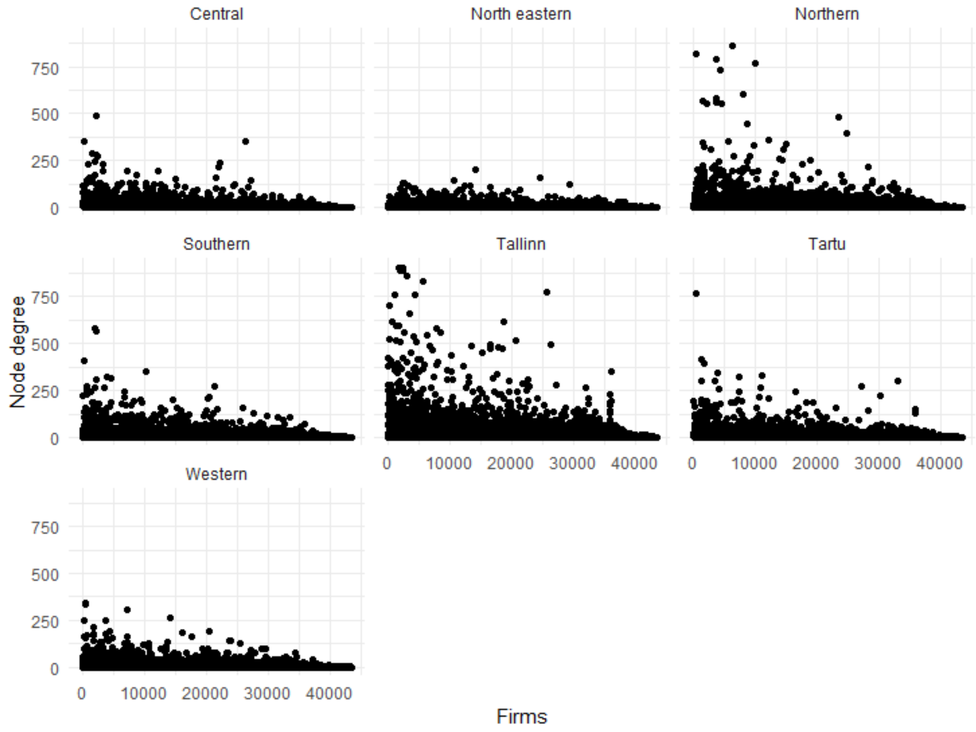
Source: own calculations

Appendix 4. Cumulative distribution of node degrees in Estonian network, 2016.



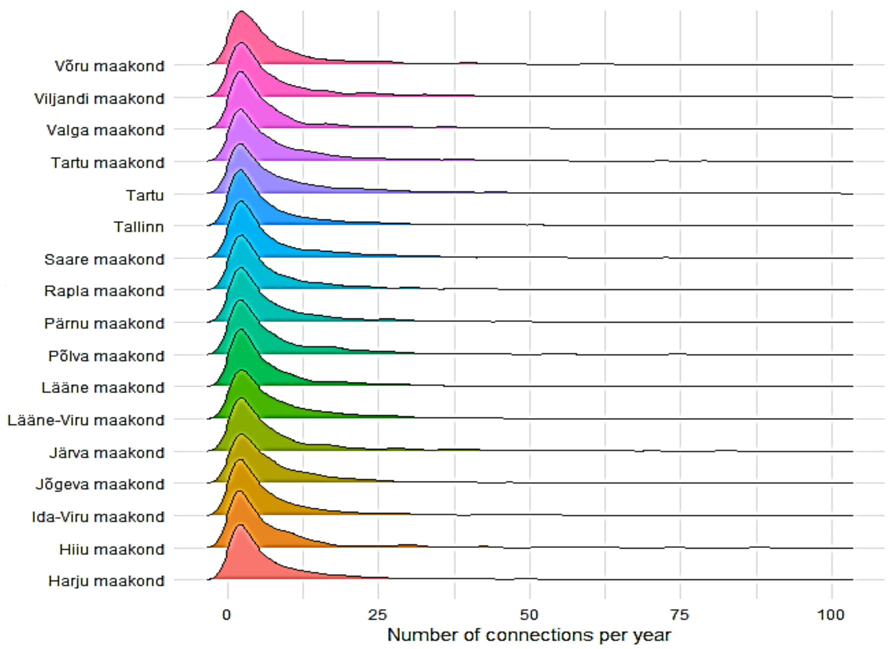
Source: own calculations

Appendix 5. Regional differences in node degree distribution, 2016.



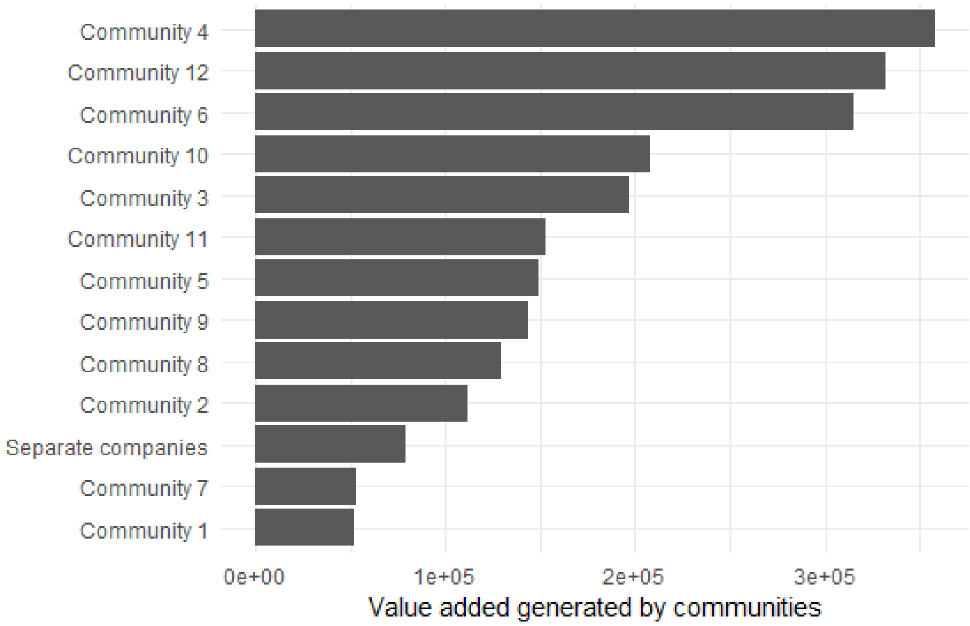
Source: own calculations

Appendix 6. Regional node degree distribution density, 2016.



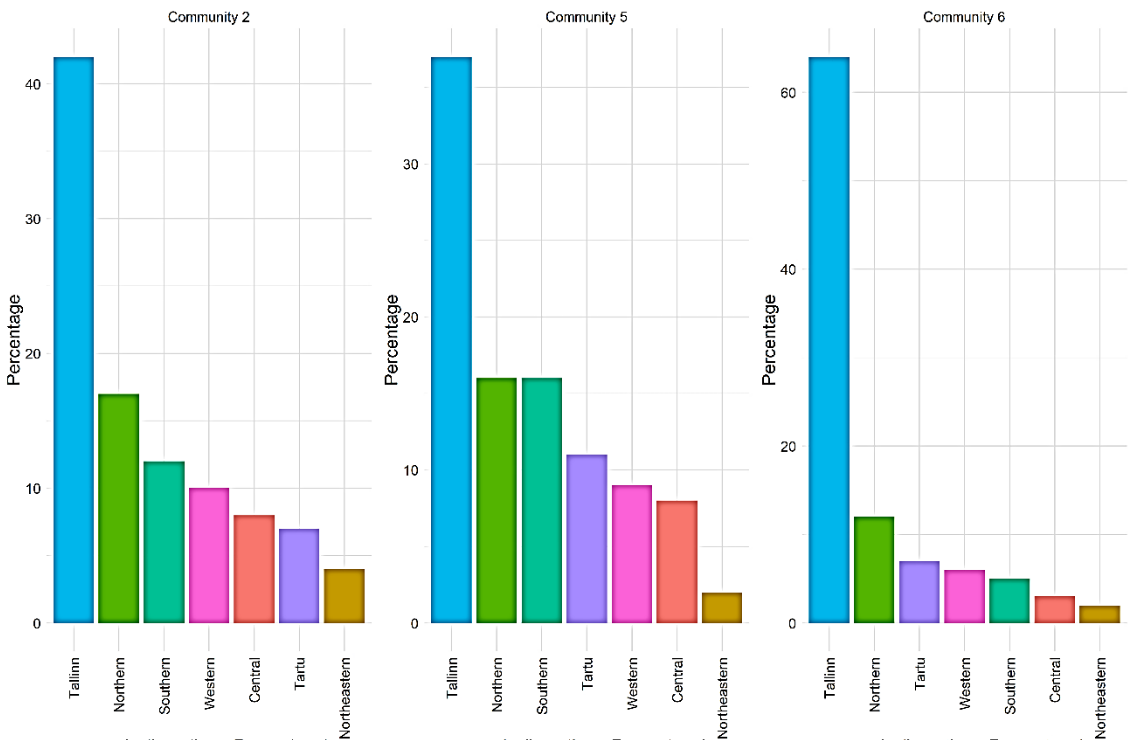
Source: own calculations

Appendix 7. Summary statistics on average value added generated by each community in the data (weighted by community size), 2016.



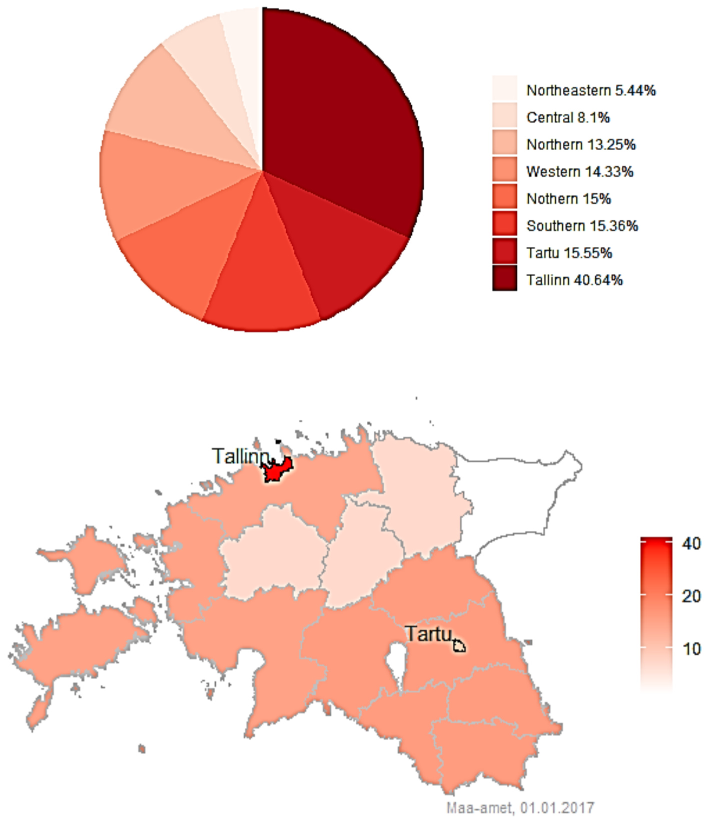
Source: own calculations

Appendix 8. Community geographic presence in Estonian network, 2016.



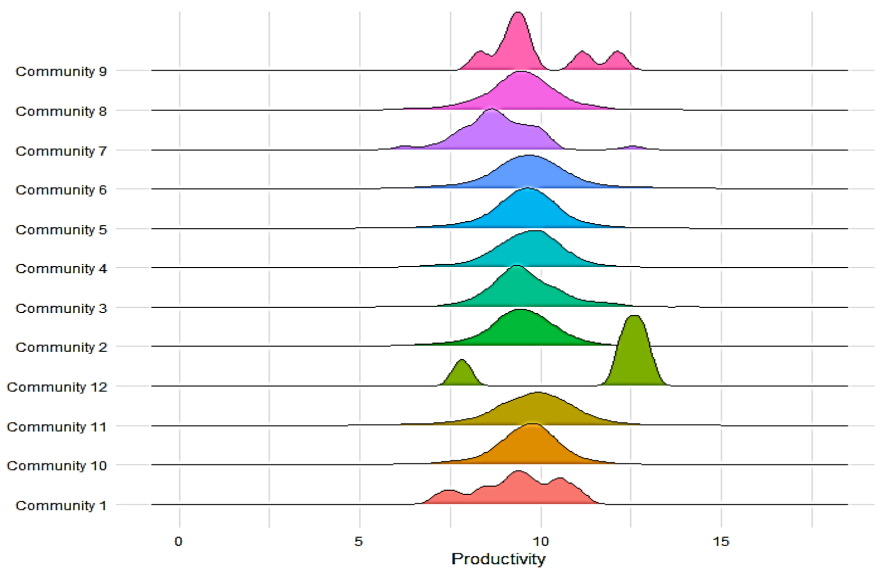
Source: own calculations

Appendix 9. Regional concentration of communities (mean share of each region in all communities), 2016.



Source: own calculations

Appendix 10. Productivity distribution within communities, 2016.



Source: own calculations

Appendix 11. Probability of trading relationship between firms (firm *i* – supplier, firm *j* – buyer), 2017 – tie formation.

	(1)	(2)	(3)	(4)	(5)
size of firm *i*	0.546***	0.558***	0.571***		0.575***
size of firm *j*	0.619***	0.625***	0.631***		0.641***
i and *j* belong to same type of ownership		0.137***	0.163***		0.088***
i and *j* in the same sector of activity		1.282***	1.257***		1.290***
i and *j* are in same sub-network		1.342***	1.304***		1.310***
i and *j* in the same county			1.290***		
node degree of *i*				0.038***	
node degree of *j*				0.022***	
clustering coefficient of *i*				-0.814***	
clustering coefficient of *j*				-0.370***	
i in Central, *j* in Central				1.387***	1.187***
i in Central, *j* in Northeastern				-0.561***	-0.844***
i in Central, *j* in Northern				-0.547***	-0.564***
i in Central, *j* in Southern				-0.709***	-0.912***
i in Central, *j* in Tallinn				-0.886***	-0.927***
i in Central, *j* in Tartu				-0.842***	-1.010***
i in Central, *j* in Western				-0.676***	-0.908***
i in Northeastern, *j* in Central				-0.648***	-1.116***
i in Northeastern, *j* in Northeastern				1.977***	1.349***
i in Northeastern, *j* in Northern				-1.360***	-1.619***
i in Northeastern, *j* in Southern				-1.358***	-1.777***
i in Northeastern, *j* in Tallinn				-1.147***	-1.473***
i in Northeastern, *j* in Tartu				-1.399***	-1.640***
i in Northeastern, *j* in Western				-1.849***	-2.416***
i in Northern, *j* in Central				-0.361***	-0.179***

i in Northern, *j* in Northeastern				-1.028***	-0.927***
i in Northern, *j* in Northern				0.308***	0.445***
i in Northern, *j* in Southern				-1.011***	-0.752***
i in Northern, *j* in Tallinn				-0.199***	-0.061**
i in Northern, *j* in Tartu				-1.139***	-0.857***
i in Northern, *j* in Western				-0.778***	-0.594***
i in Southern, *j* in Central				-0.611***	-0.629***
i in Southern, *j* in Northeastern				-1.193***	-1.285***
i in Southern, *j* in Northern				-1.059***	-0.817***
i in Southern, *j* in Southern				1.000***	0.975***
i in Southern, *j* in Tallinn				-1.257***	-1.101***
i in Southern, *j* in Tartu				0.542***	0.587***
i in Southern, *j* in Western				-0.793***	-0.791***
i in Tallinn, *j* in Central				-0.886***	-0.714***
i in Tallinn, *j* in Northeastern				-1.133***	-1.246***
i in Tallinn, *j* in Northern				-0.175***	-0.018
i in Tallinn, *j* in Southern				-1.184***	-0.904***
i in Tallinn, *j* in Tartu				-1.045***	-0.856***
i in Tallinn, *j* in Western				-0.860***	-0.694***
i in Tartu, *j* in Central				-1.084***	-0.995***
i in Tartu, *j* in Northeastern				-1.183***	-1.122***
i in Tartu, *j* in Northern				-1.310***	-0.989***
i in Tartu, *j* in Southern				0.564***	0.584***
i in Tartu, *j* in Tallinn				-1.108***	-1.090***
i in Tartu, *j* in Tartu				1.334***	1.238***
i in Tartu, *j* in Western				-1.138***	-1.197***
i in Western, *j* in Central				-0.741***	-0.800***
i in Western, *j* in Northeastern				-1.584***	-1.730***
i in Western, *j* in Northern				-0.875***	-0.747***

i in Western, *j* in Southern				-0.919***	-0.945***
i in Western, *j* in Tallinn				-0.970***	-0.908***
i in Western, *j* in Tartu				-1.234***	-1.334***
i in Western, *j* in Western				1.436***	1.219***
Observations	185,252	185,252	185,252	185,252	185,252
McFadden R^2	0.196	0.266	0.293	0.377	0.323

Source: own calculations

(1) All independent variables reflect firm *i* characteristics in previous period *t-1* (2016). Dependent variable is the existence of trading relationship in 2017. For industrial classification EMTAK2 is used. Embeddedness in community is defined by Louvain algorithm. To allow binary model estimation a random sample of firms' pairs that do not trade in 2017 is created. The symbols *, **, *** show significance at 0.1, 0.05, 0.01 level.

Appendix 12. Probability of trading relationship between firms (firm *i* – supplier, firm *j* – buyer), 2017 - tie persistence.

	(1)	(2)	(3)	(4)	(5)
size of firm *i*	0.114***	0.105***	0.107***		0.108***
size of firm *j*	0.148***	0.151***	0.153***		0.154***
i and *j* belong to same type of ownership		-0.280***	-0.283***		-0.285***
i and *j* in the same sector of activity		0.474***	0.475***		0.476***
i and *j* are in same sub-network		0.318***	0.141***		0.315***
i and *j* in the same county			0.313***		
node degree of *i*				0.001***	
node degree of *j*				0.0005***	
clustering coefficient of *i*				0.221***	
clustering coefficient of *j*				0.565***	
i in Central, *j* in Central				0.023	0.040
i in Central, *j* in Northeastern				-0.202**	-0.336***
i in Central, *j* in Northern				-0.061	-0.103**
i in Central, *j* in Southern				-0.177***	-0.253***
i in Central, *j* in Tallinn				-0.111***	-0.134***
i in Central, *j* in Tartu				-0.102	-0.209***

i in Central, *j* in Western				-0.190***	-0.224***
i in Northeastern, *j* in Central				-0.015	-0.111
i in Northeastern, *j* in Northeastern				-0.075**	-0.139***
i in Northeastern, *j* in Northern				-0.326***	-0.395***
i in Northeastern, *j* in Southern				-0.275**	-0.285**
i in Northeastern, *j* in Tallinn				-0.259***	-0.322***
i in Northeastern, *j* in Tartu				-0.117	-0.180
i in Northeastern, *j* in Western				-0.306*	-0.401**
i in Northern, *j* in Central				-0.058	-0.113***
i in Northern, *j* in Northeastern				-0.034	-0.157**
i in Northern, *j* in Northern				0.108***	0.122***
i in Northern, *j* in Southern				-0.174***	-0.193***
i in Northern, *j* in Tallinn				-0.074***	-0.064***
i in Northern, *j* in Tartu				-0.161***	-0.196***
i in Northern, *j* in Western				-0.159***	-0.175***
i in Southern, *j* in Central				-0.099*	-0.183***
i in Southern, *j* in Northeastern				-0.161*	-0.211**
i in Southern, *j* in Northern				-0.084**	-0.075*
i in Southern, *j* in Southern				-0.047**	0.004
i in Southern, *j* in Tallinn				-0.125***	-0.097***
i in Southern, *j* in Tartu				-0.098***	-0.057*
i in Southern, *j* in Western				-0.180***	-0.189***
i in Tallinn, *j* in Central				-0.069**	-0.086***

i in Tallinn, *j* in Northeastern				-0.149***	-0.246***
i in Tallinn, *j* in Northern				-0.073***	-0.047***
i in Tallinn, *j* in Southern				-0.171***	-0.129***
i in Tallinn, *j* in Tartu				-0.129***	-0.135***
i in Tallinn, *j* in Western				-0.197***	-0.176***
i in Tartu, *j* in Central				-0.064	-0.127
i in Tartu, *j* in Northeastern				-0.081	-0.116
i in Tartu, *j* in Northern				-0.021	-0.025
i in Tartu, *j* in Southern				-0.111***	-0.078**
i in Tartu, *j* in Tallinn				-0.065**	-0.075**
i in Tartu, *j* in Tartu				0.046*	0.081***
i in Tartu, *j* in Western				0.012	-0.034
i in Western, *j* in Central				0.039	0.003
i in Western, *j* in Northeastern				-0.471***	-0.548***
i in Western, *j* in Northern				-0.138***	-0.128***
i in Western, *j* in Southern				-0.063	-0.048
i in Western, *j* in Tallinn				-0.145***	-0.106***
i in Western, *j* in Tartu				-0.162**	-0.172**
i in Western, *j* in Western				0.017	0.062***
Observations	215,689	215,689	215,689	215,689	215,689
McFadden R^2	0.018	0.032	0.033	0.012	0.033

Source: own calculations

(1) All independent variables reflect firm *i* characteristics in previous period *t-1* (2016). Dependent variable is the existence of trading relationship in 2017. For industrial classification EMTAK2 is used. Embeddedness in community is defined by Louvain algorithm. The symbols *, **, *** show significance at 0.1, 0.05, 0.01 level.

Appendix 13. Average value of productivity and value chain length aggregated by regions, 2016-2017.

County	Weighted value chain length (by VA)	Average productivity (log scale)	Average productivity	Fraction of exports in total sales
Harju county	3,860	9,263	32269.51	0,646
Hiiu county	3,752	9,147	22942.29	0,605
Ida-Viru county	3,529	8,921	21017.04	0,739
Jogeva county	4,707	9,209	23018.38	0,584
Jarva county	4,429	9,197	25133.94	0,561
Laane-Viru county	4,184	9,227	24419.40	0,569
Laane county	4,038	9,147	24246.38	0,530
Polva county	4,379	9,152	22739.09	0,463
Parnu county	3,950	9,193	25609.01	0,738
Rapla county	4,081	9,214	24947.78	0,704
Saare county	3,934	9,139	21960.45	0,708
Tallinn	3,327	9,384	40118.50	0,699
Tartu	3,843	9,279	30471.26	0,525
Tartu county	4,355	9,199	26925.36	0,636
Valga county	4,367	9,244	28002.09	0,672
Viljandi county	4,434	9,328	24672.97	0,651
Voru county	4,019	9,196	25456.34	0,533

Source: own calculations based on VAT, annual accounts data, business registry data and value chain length value calculated by Võrk, Unt, Varblane (2018).

(1) Fraction of exports is calculated as a weighted average value of exports in total sales, taking into account fractions that exceed 10% and do not exceed 100%.

Appendix 14. Summary statistics on average upstreamness values (weighted average distance to final consumer) from Statistics Estonia database, 2010-2014.

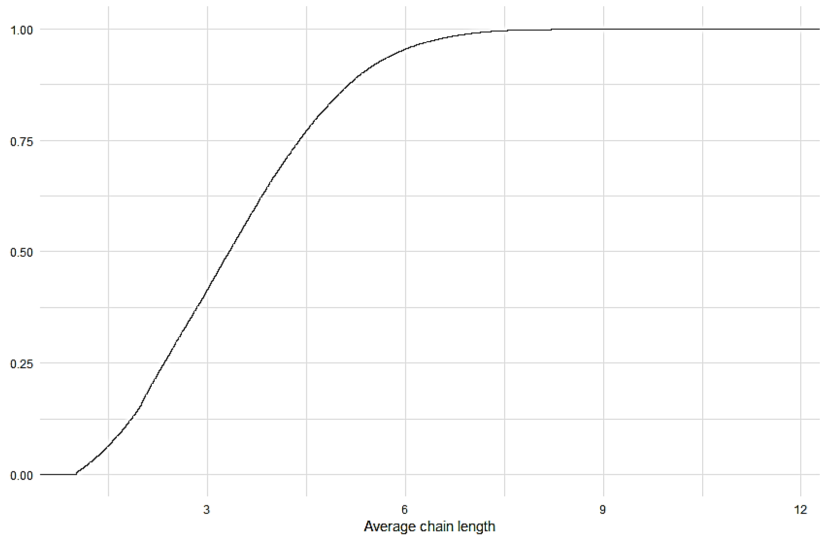
Variable	Mean
Residential care activities; social work activities without accommodation	1.022
Scientific research and development	1.042
Human health activities	1.113
Public administration and defence; compulsory social security	1.121
Education	1.143
Creative, arts, entertainment and cultural activities, libraries, museums; gambling activities	1.272
Travel agency, tour operator and other reservation service and related activities	1.330
Sports activities and amusement and recreation activities	1.387
Construction	1.392
Other personal service activities	1.405

Activities of membership organisations	1.410
Accommodation; food and beverage service activities	1.413
Retail trade	1.447
Manufacture of other transport equipment	1.475
Manufacture of food, beverages and tobacco products	1.599
Manufacture of basic pharmaceutical products and pharmaceutical preparations	1.656
Water collection, treatment and supply	1.828
Manufacture of furniture, other manufacturing	1.836
Manufacture of motor vehicles, trailers and semi-trailers	1.947
Manufacture of textiles, wearing apparel and leather products	1.958
Real estate activities	1.980
Insurance, reinsurance and pension funding, except compulsory social security	2.253
Crop and animal production	2.292
Manufacture of machinery and equipment n.e.c.	2.422
Fishing and aquaculture	2.433
Wholesale trade	2.479
Publishing activities	2.495
Computer programming, consultancy and related activities; information service activities	2.495
Movie, video, TV programme production, sound recording, music publishing, broadcasting activities	2.543
Wholesale and retail trade and repair of motor vehicles and motorcycles	2.546
Financial service activities, except insurance and pension funding	2.642
Electricity, gas, steam and air conditioning supply	2.782
Telecommunications	2.797
Manufacture of other non-metallic mineral products	2.918
Architectural and engineering activities; technical testing and analysis	2.935
Postal and courier activities	2.974
Repair of computers and personal and household goods	3.024
Other professional, scientific and technical activities; veterinary activities	3.055
Air transport	3.072
Manufacture of fabricated metal products, except machinery and equipment	3.102
Manufacture of rubber and plastic products	3.116
Manufacture of coke and refined petroleum products	3.214
Security, investigation; services to buildings and landscape; office and business support activities	3.236
Advertising and market research	3.275
Activities auxiliary to financial services and insurance activities	3.390
Manufacture of electrical equipment	3.406
Land transport and transport via pipelines	3.408
Legal and accounting activities; activities of head offices; management consultancy activities	3.481
Printing and reproduction of recorded media	3.489
Manufacture of computer, electronic and optical products	3.514

Manufacture of paper and paper products	3.568
Water transport	3.648
Rental and leasing activities	3.651
Mining and quarrying	3.661
Manufacture of chemicals and chemical products	3.674
Manufacture of wood and wood and cork products, ex furniture, articles of plaiting materials	3.761
Repair and installation of machinery and equipment	3.768
Manufacture of basic metals	3.993
Employment activities	4.290
Forestry and logging	4.554
Sewerage, waste collection, treatment and disposal; materials recovery; remediation activities	4.574
Warehousing and support activities for transportation	4.631

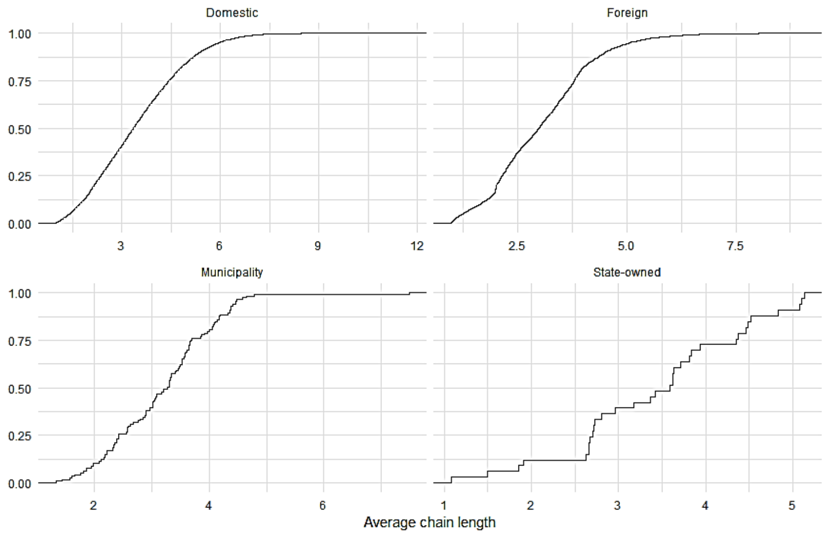
Source: own calculations based on supply, use and input-output tables from Statistics Estonia database.

Appendix 15. Value chain length cumulative distribution for Estonia, 2016-2017.



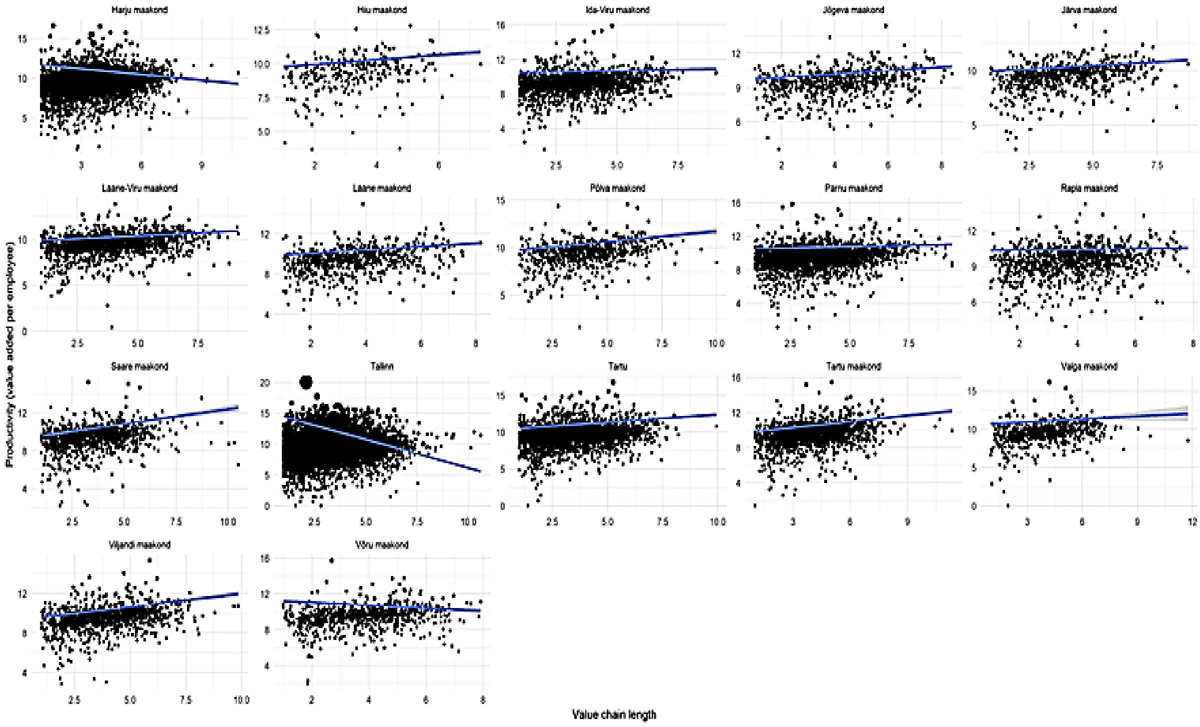
Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Appendix 16. Value chain length cumulative distribution for different ownership types, Estonia, 2016-2017.



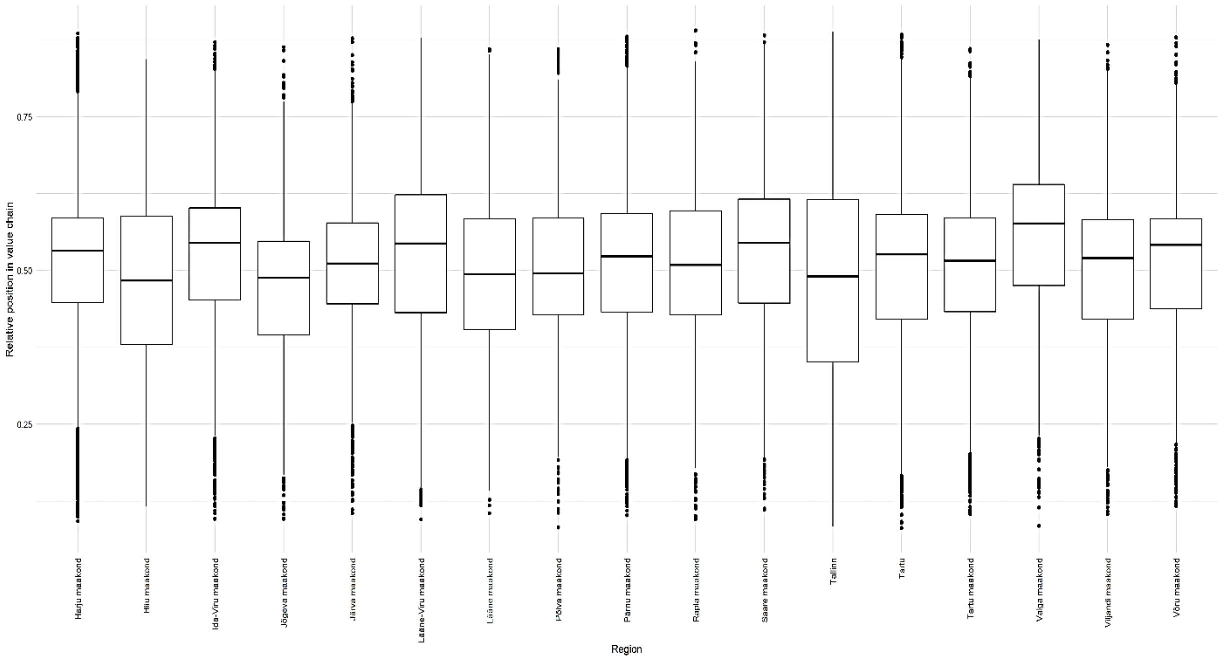
Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Appendix 17. Value chain length and productivity relationship by counties, 2016.



Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Appendix 18. Distribution of relative position of firms in the value chain across counties, 2016-2017.



Source: calculations based on the data calculated by Võrk, Unt, Varblane (2018).

Appendix 19. Productivity of firms by regions (log of value added per employee), 2016.

	Tallinn	Tartu	Northern	North-eastern	Southern	Western	Central
Length of chain	0.03 **	0.10 ***	0.07 ***	0.08 ***	0.12 ***	0.06 ***	0.12 ***
Relative position	-1.35 ***	-1.11 ***	-1.61 ***	-0.97 ***	-1.08 ***	-1.29 ***	-1.12 ***
Observations	18047	3336	5922	1901	5717	4713	3396
R^2 adjusted	0.20	0.23	0.23	0.19	0.22	0.18	0.20

Source: own calculations based on VAT, annual accounts data, business registry data and value chain length value calculated by Võrk, Unt, Varblane (2018).

(1) Effects of value chain length on firm productivity are separated by regions, number of observations is different across regions; for industrial classification EMTAK2 is used; independent variables also include firm size, total sales, assets and number of steps to exporter; robust standard errors. The symbols *, **, *** show significance at 0.1, 0.05, 0.01 level respectively.