Geography and Network Tie Formation

Sándor Juhász

GEOGRAPHY AND NETWORK TIE FORMATION

Geografie en netwerkverbinding vorming

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 5 december 2019 des middags te 4.15 uur

door

Sándor Juhász

geboren op 7 september 1989
te Kecskemét, Hongarije

Promotor: Prof. dr. R. A. Boschma

Copromotoren: Dr. A. Morrison
Dr. P. M. A. Balland

Committee: Prof. Dr. Marjolein Caniels
Prof. Dr. Koen Frenken
Prof. Dr. Elisa Giuliani
Dr. Anne Ter Wal
Prof. Dr. Jerome Vicente

Contents

Chapter 1

Introduction

1.1 General introduction

Society can be seen as a very complicated system, where billions of individuals move, communicate, compete and cooperate continuously. It is difficult to understand how this very complex system works by simply analysing its components. By now, networks turned out to be necessary tools to encode the interactions of individual components, in order to understand the patterns and rules of these systems (Barabási, 2018a). Ties of friendship, family or professional collaboration construct the different networks of society that determine the diffusion of resources, information and knowledge inside it.

Over the past decades, scholars from various fields acknowledged that networks play a crucial role in the organization of economic activities (Granovetter, 1985; Burt, 1992; Jackson, 2008). Networks represent the structure of social interactions that enables the exchanges of resources that otherwise do not spread easily through markets. Such resources can be strategic information or tacit knowledge, which largely contribute to the performance and innovativeness of individuals and organizations. As tools to map and analyse the structure of even large-scale real-world networks became available, considerable attention has been given to study the relationship between economic prosperity, innovation, technological change and network structures.

The social and economic connections of individuals, organizations or technologies also have spatial aspects. Geographical proximity facilitates face-to-face communication and personal interaction and therefore the sharing of information and knowledge through network ties (Saxenian, 1994; Storper and Venables, 2004; Boschma, 2005; Bresci and Lissoni, 2009). Economic geographers also turn their attention towards networks to better understand central questions of the field, such as how networks work in agglomerations or how do they contribute to the uneven distribution of economic activities (Bathelt and Glückler, 2003; Ter Wal and Boschma, 2009).

By now, it is well established that dense social networks play a crucial role in the exploitation of positive agglomeration externalities, as they facilitate the efficient sharing of ideas, information and knowledge (Almeida and Kogut, 1999; Gordon and McCann, 2000; Duranton and Puga, 2004; Eriksson and Lengyel, 2019). Participation in local networks of knowledge exchange is key to benefit from industrial concentration (Owen-Smith and Powel, 2004; Giuliani and Bell 2005). Besides dense local connections, distant ties are realized to be important to fuel agglomerations with new, non-redundant knowledge, which is necessary for the renewal and long-term prosperity of regional economies (Bathelt et al., 2004; Boschma and Ter Wal, 2007; Morrison 2008). Moreover, mapping the relationships between technologies also helped to understand the structure of regional economies and their possible development paths (Neffke et al., 2011; Boschma et al., 2015; Rigby 2015).

Studies that pioneered network application in relation to classic questions of economic geography mainly used a static approach to map the structure of knowledge exchange (e.g. Giuliani and Bell, 2005; Boschma and Ter Wal, 2007; Morrison 2008) or the connections between available technologies in regional economies (e.g. Hidalgo et al., 2007). However, social and economic interaction is a dynamic process by nature, as individuals, firms or technology fields constantly vary their relationships. Later on, methodological and technical development also allowed researchers to study the evolution of these networks over time. Discovering the dynamics of knowledge sharing and interactive learning in local networks helped researchers to understand the development of industry clusters (Giuliani, 2013; Balland et al., 2016). Mapping collaboration patterns through the life-cycles of industries enabled to capture how the influence of geography on cooperation shifts over time (Balland, 2012; Ter Wal, 2014, Balland et al., 2013). Tracking how connections of technologies change by time also allowed to capture the appearance and disappearance of technologies in regional economies (Kogler et al., 2013; Rigby 2015). The present thesis wants to contribute to this research direction in economic geography, namely on how do networks form and change and what role geography plays in this process. In the following, four case studies are provided to gain further insights on how geography influences the transfer, combination and recombination of knowledge through network ties.

1.2 Aim of the thesis and research questions

Geographical distance often forms a barrier against interaction and it is more likely that co-located actors begin to collaborate. However, geographical proximity does not necessarily make all actors connected. Evolutionary economic geography suggests that agents are heterogenous in their capabilities (Boschma and Frenken, 2010, 2018), therefore they do not connect easily, nor do they easily learn from each other. This is exactly why some actors in collaboration networks form more linkages than others (Powell et al., 1996; Giuliani, 2007). This evolutionary perspective on the geography of networks has led to additional insights in the literature on industry clusters (see Giuliani, 2013; Balland et al., 2016), on the development of industries (see Balland, 2012; Ter Wal, 2014; Balland et al., 2013) and on the development of technological connections (see Rigby, 2015; Balland and Rigby, 2017). All together related studies found that the unique characteristics of network nodes, their similarity or proximity and network structural effects influence the formation of spatial networks over time. This dissertation contributes to the above literature on the geography of networks by taking this evolutionary approach to study network formation.

To do so, two objectives are outlined. First, the thesis aims to extend the literature on network tie formation in spatial networks by introducing new factors, approaches and contexts to consider. More precisely, the dissertation argues that entrepreneurial background is an unexamined driver of network tie formation in industry clusters. It also suggests a novel

approach to study tie formation processes through the separation of tie creation from tie persistence. Moreover, the thesis also seeks to extend the literature by searching for the determinants of ties in context of the complex network of technologies. Second, the thesis searches for a novel way to connect the performance of network nodes to their ties inside agglomerations. In sum:

The general aim of the thesis is to extend the literature on the drivers and outcomes of tie formation in spatial networks.

To accomplish these general objectives, empirical exercises are presented in the following from different network settings, with different size, complexity and time dimensions. All of the chapters focus on the determinants and outcomes of tie formation in networks through cases where geography influences this process.

Previous research on the formation of networks in clusters found that besides network structural effects and the proximity of firms, individual characteristics such as the absorptive capacity (Giuliani and Bell, 2005), the external openness (Broekel and Hartog, 2013a) or the knowledge base of companies (Giuliani, 2013) determine their ability to form collaboration ties. So far, studies that explain the formation of cluster knowledge networks do not consider the role of entrepreneurial background as an individual characteristic. Therefore, the aim of Chapter 2 is to bridge two streams of literature in relation to industry clusters: the one that emphasizes the importance of social networks in the diffusion of knowledge in clusters (Giuliani and Bell, 2005; Boschma and Ter Wal, 2007; Morrison, 2008; Vicente et al., 2011), and the other that highlights the importance of spinoff companies in the development of clusters (Klepper, 2007; 2010; Buenstorf and Klepper, 2009). To do so, the following question is raised:

RQ 1) Do spinoff background matters for tie formation in cluster knowledge networks?

Chapter 2 focuses on a mature cluster of printing and paper product industry in Kecskemét, Hungary. The analysis rests on relational data that represents the informal technological knowledge exchange of firms and was collected through face-to-face interviews in 2012. The application of exponential random graph models (ERGMs) allows to test for the influence of spinoff background in knowledge tie formation, besides controlling for other node level firm characteristics, dyad level proximities of companies and network structural properties.

This chapter demonstrates that besides geographical proximity, ownership similarity and the triadic formation of linkages, being a spinoff company enhances tie formation in the local network. These results lead to the conclusion that the pre-entry experiences and inherited routines of spinoffs make them more capable to take advantage of local knowledge concentration through collaboration.

This first study shows that the formation of knowledge ties in local industry clusters is very much influenced by spinoff companies. This work connects two influential theories on industry clusters, however, its' arguments are based only on a static network of knowledge sharing between companies. Chapter 3 enters the discussion on the dynamics of cluster knowledge networks. While previous studies on the evolution of cluster networks identified the general, micro-level drivers of tie formation, this chapter claims that tie creation and tie persistence have to be analysed separately in order to better understand the development of clusters. Therefore, the following research question is formulated:

RQ 2) What mechanisms drive the creation and persistence of ties in cluster knowledge networks?

Giuliani (2013) in her pioneering study finds that cohesion effects – namely reciprocity and triadic closure – and the different knowledge bases of firms determine changes in the network structure over time. Balland et al. (2016) finds that besides the equal importance of network embeddedness, the structural evolution of business knowledge networks is driven by status effect, while the proximity of firms is more important for the technical knowledge networks behind clusters. However, there still is little evidence on how networks in clusters change, to what extent network dynamics follow path dependence (Glückler, 2007), and how network dynamics in general influence the evolution of a cluster (Cantner and Graf, 2006; Balland, 2012; Crespo et al., 2016).

Chapter 3 suggests a novel approach and argues that the distinction of tie creation from tie persistence is important because the micro-motivations of creating and maintaining collaboration ties are different and they also have different consequences for the evolution of clusters themselves. A novel perspective is proposed to study how endogenous network effects, geographical and cognitive proximity of actors and their interplay influence the evolutionary process of network formation in clusters.

For empirical underpinning, the same printing and paper product cluster in Kecskemét is examined as in Chapter 2, however, a second wave of data collection in 2015 allowed to investigate the determinants of tie formation in a dynamic network setting. Stochastic actor-oriented models are applied to identify the drivers of tie creation and tie persistence from 2012 to 2015.

On the one hand, triadic closure and geographical proximity turn out to increase the probability of tie creation, but do not influence tie persistence. On the other hand, cognitive proximity is positively correlated to tie persistence, but firms create ties to cognitively proximate companies only if they are loosely connected through common third partners. Based on these results, the study argues that through the distinction of tie creation and tie persistence we can understand the evolutionary processes of network retention and network variation better in clusters.

After searching for the drivers of tie formation in social networks of industry clusters, Chapter 4 shifts the focus to the dynamics of a rather abstract and complex network, namely to the network of technologies. Two activities – such as technologies, industries, occupations or research fields – are related when they require similar knowledge, skills, capabilities or inputs (Hidalgo et al., 2018). The relatedness of different technologies is usually represented as a network, where nodes are technologies and ties sign the similarity or proximity of technologies to each other. This network – the technology space – is widely used to map the diversification opportunities of cities, regions or countries based on their available technologies and their relations to other technologies. However, there is still limited evidence about how technological relatedness comes into existence, from where it emerges, and how it develops over time. In fact, in most of the empirical studies relatedness is either treated as a constant or as an exogenously given component. Chapter 4 aims to contribute to the literature by searching for determinants that explain tie formation in the dynamically changing network of technologies. Therefore, the following research question is formulated:

RQ 3) How do co-location, complexity and collaboration determine technological relatedness ties?

Three different aspects are introduced as potential forces that influence how technologies are related to each other. First, the co-location of technologies enhances learning processes and knowledge spillover effects between them, which further contributes to the presence and strength of their relatedness. Second, the complexity of technologies is used to proximate the related re-combinatorial efforts (Sorenson and Fleming, 2004), required management (Mcnerney et al., 2011) and possible socio-economic benefits of technological combinations (Fleming and Sorenson, 2001). Therefore, the complexity of technologies can also influence the presence and strength of relatedness between them. Third, human collaboration creates a pool of ideas, skills, capabilities and knowledge that enables the combination of different technologies (Hidalgo, 2015) and therefore it can mediate the effect of co-location and complexity on tie formation in the technology space.

The relatedness of technologies – or in other words the ties between technologies – is operationalized by the co-occurrence of 4-digit technology classes on patents. To verify the above arguments, dynamically changing non-overlapping weighted networks of technologies are constructed for every 5-year in the period of 1980-2010 on the basis of the OECD REGPAT 2018 patent dataset. Co-location of technologies is captured by the co-agglomeration measure of Ellison et al. (2010), while technological complexity is operationalized by the recently developed structural complexity index by Broekel (2019). Results of zero-inflated negative binomial regressions in gravity model settings show that co-location and complexity of technology pairs indeed determine the presence of technological relatedness ties and also impact the strength of them. Moreover, the chapter provides evidence that these determinants differ in case collective efforts are involved.

The fourth and final empirical exercise extends the framework of the thesis by shifting the focus from the drivers of tie formation to the possible outcomes of forming linkages and taking positions in collaboration networks. Chapter 5 searches for the relationship between tie formation and individual success by asking which ties do matter for exceptional performances in urban creative industries? Building on central concepts from the literature on social networks, the following research question is raised in relation:

RQ 4) How do bridging ties in core/periphery networks of creative production influence individual success?

Creative industries typically rely on project-based production systems that require intense collaboration between individuals. Local buzz, or the geographic concentration of dense social networks can make information and knowledge sharing more efficient, which is crucial for such industries (Grabher, 2002; Storper and Venables, 2004; Stark, 2009). So far, empirical studies in economic geography on the formation of networks in creative production established that collaboration between firms are driven by different dimension of proximities and their importance – such as the role of geographical proximity – changes during the life-cycle of industries (Balland et al., 2013). Chapter 5 takes a different approach to contribute to this literature. First, it focuses on individual, creator level cooperation instead of firm level collaborations (in a similar fashion to De Vaan et al., 2015). Second, it searches for the relationship between network position and the success of individuals inside an agglomeration.

In project-based, creative industries, such as film production, relations to prestigious central creators can facilitate individual success; while bridging otherwise unconnected communities provide access to diverse pools of ideas and knowledge. Chapter 4 combines two classic social networks concepts, namely brokerage and core/periphery structure to analyse how network position of individuals can contribute to success in urban creative industries. The study argues that links to peripheral creators provide additional and complementary benefits for core members and consequently, those creators who broker the core of the network with the periphery enjoy both type of benefits and thus are more likely to achieve success.

To verify the argument, the exercise focuses on the film industry of Budapest and a dynamic network of movie creators based on a unique dataset of Hungarian feature films for the 1990-2009 period is constructed. Findings confirm that being in the core and brokering the network at the same time induce individual success together. Moreover, a new way to capture brokers' role in core/periphery networks is also proposed and findings suggest that those core creators who also act as gatekeepers and form ties to bridge the core with the periphery have an increased likelihood of award winning. The chapter highlights that tie formation to different parts of the network also influences the performance of individuals in agglomerations.

In summary, the present thesis deals with a variety of networks with spatial dimensions, all of which contributes to our understanding on the micro-processes of network tie formation

and its outcomes. The findings of the different chapters add pieces to the discussion on how does geography shape networks of knowledge sharing, collaboration and technologies.

1.3 Contribution and outline

The following outlines describes the structure of the thesis and also highlights the main contribution of each part. Chapter 2 introduces a new factor to consider while studying cluster networks and illustrates the influence of spinoff companies on tie formation in cluster knowledge networks. It provides evidence that spinoffs have higher impact on network formation in clusters than other companies. Chapter 3 takes a dynamic perspective on cluster knowledge networks and searches for the determinants of tie creation and tie persistence to better understand cluster evolution. Chapter 4 discusses how technological relatedness or the ties in the technology space depend on the co-location and complexity of technologies. Moreover, it demonstrates that human collaboration mediates the influence of co-location and complexity on both the presence and strength of relatedness. Chapter 5 presents how individual success in an urban creative industry depends on network ties. It shows that forming ties that bridge the core and the periphery of a collaboration network significantly help individual creative success. Finally, Chapter 6 summarises the findings, draws conclusions outlines the limitations of this research, as well as discusses the possible directions for future research.

Chapter 2

Spinoffs and tie formation in cluster knowledge networks

This chapter is based on a single authored research paper of the candidate, which is accepted for publication in the journal Small Business Economics.

2.1 Introduction

Industry clusters, the geographic concentrations of economic activities that operate in the same or interconnected sectors (Gordon and McCann, 2000), foster higher innovation and economic performance of firms (Krugman, 1991; Porter, 1990; Cooke, 2002). Their success is usually explained by agglomeration externalities (Rosenthal and Strange, 2004; Tallman et al., 2004) that arise from labour market pooling, specialized suppliers and knowledge spillovers as Marshall (1920) described in his influential early works. Following this seminal contribution, scholars have emphasized the importance of localized knowledge spillovers on innovation and argued that it is mainly the geographical and social proximity of actors that help the circulation of new ideas from one firm to another, promoting the processes of incremental innovation and collective learning (Asheim, 1996; Saxenian, 1994; Audretsch and Feldman, 1996; Maskell and Malmberg, 1999).

The realization that knowledge is not 'in the air' available for every actor in industrial concentration in contrast to the original idea of Marshall (1920) has increased the interest on social networks in clusters (Gordon and McCann, 2000; Cooke, 2002; Fornahl and Brenner, 2003; Giuliani, 2007). Giuliani and Bell (2005) show that knowledge is not diffused evenly in industry clusters, but circulates in local networks to which only a core group of firms characterized by advanced absorptive capabilities have access. This selective nature of local knowledge networks implies that firms are likely to differ in their ability to exploit and benefit from locally accumulated knowledge (Rigby and Brown, 2015). Moreover, recent studies on cluster knowledge network formation demonstrated that along with individual capabilities and skills of firms, the establishment of local network ties is also influenced by the proximity of firms and the network structure itself (e.g. Broekel and Boschma, 2012; Giuliani, 2013; Balland et al., 2016).

Besides positive externalities and knowledge spillovers, clustering has also been explained through the formation of spinoff companies (Klepper, 2007; 2010; Buenstorf and Klepper, 2009; Boschma, 2015; Qian, 2018). Spinoffs, the new companies founded by employees of incumbent firms in the same industry tend to locate close to their parents (Dahl and Sorenson, 2012), which helps the reproduction of clusters over time. Moreover, their superior inherited capabilities help them to perform better than other firms (Klepper, 2009). In recent years, the special role and importance of spinoffs in clusters have been investigated by several empirical studies (e.g. Feldman et al., 2005; Klepper, 2007; 2010; Heebels and Boschma, 2011; Wenting, 2008; Morrison and Boschma, 2019), and most of these works have highlighted the importance of pre-entry background in order to benefit from clustering and locally accumulated knowledge (Buenstorf and Klepper, 2009; 2010; Boschma, 2015).

However, there is still a lack of evidence as to whether the acquired capabilities, routines or relationships of spinoffs matter in order for firms to collaborate and form ties in cluster networks.

The main aim of this chapter is to emphasize the role of spinoff companies in clusters by showing their outstanding ability to form local knowledge ties. I argue that because of their pre-entry experience and inherited capabilities, spinoffs form knowledge ties more efficiently and therefore gain more from locally accumulated industrial knowledge. It is important to stress these questions as it is still unclear whether spinoffs are generally more capable to collaborate and embed in cluster networks than other firms. Furthermore, the general aim of this chapter is to combine the emerging literature on cluster knowledge networks and the literature on spinoffs. In order to meet these expectations, I focus on the mature cluster of printing and paper product industry in Kecskemét, Hungary. The analysis rests on relational data on the informal technological knowledge exchange of firms, collected through face-to-face interviews in 2012. Besides testing for the influence of spinoff background in knowledge tie formation, I control for other node level firm characteristics, dyad level proximities of companies and network structural properties by applying exponential random graph models (ERGMs).

The empirical results show that spinoff companies are more likely to be connected to other firms in the cluster than non-spinoffs. I also found that there is more collaboration inside domestic and foreign ownership groups than there is between them and the geographical proximity of actors also enhances networking. Controlling for the structure of the network, I found that the formation of triads and mutual relationships had a significant influence on knowledge sharing. The findings support the arguments that spinoff companies are more likely to form knowledge ties in clusters and that their inherited capabilities enable them to get better access to local industrial knowledge. To further strengthen these arguments and explain the possible reasons behind the superior capabilities of spinoffs to form network ties, a detailed discussion is provided on the formation of the specific cluster.

The structure of the chapter is as follows. Section 2.2 reviews the literature on knowledge networks and spinoffs in clusters. In Section 2.3, I present the context of the analysis and the data collection process, while in Section 2.4 I detail the applied statistical method to analyse network formation. Section 2.5 presents the results of the applied ERGMs. The main conclusions, the discussion over the implications and promising future research directions are presented in Section 2.6.

2.2 Knowledge sharing, networks and spinoffs in clusters

2.2.1 Knowledge networks in clusters

A crucial contribution to understanding the success of clusters was the realization that geographical proximity in itself does not necessarily help firms in specialised industries, but it is rather the case that social networks play an important role in innovation and learning (Sorenson and Audia, 2000; Caniels and Romijn, 2003; Owen-Smith and Powell, 2004). As Giuliani (2007) shows in her studies, knowledge is 'not in the air' available for everyone in industry clusters, as in Marshall's original impressions, but it circulates only in selective local knowledge networks. Knowledge networks link firms through the transfer of innovation-related knowledge, by the joint solution of complex technical problems (Giuliani, 2010). In order to benefit from the locally accumulated industrial knowledge, firms must take part in these networks, which requires skills and capabilities to transfer and exploit knowledge through collaboration. Therefore, knowledge in clusters flows mainly within a core group of firms characterized by strong knowledge bases and advanced absorptive capacity (Giuliani and Bell, 2005).

Being strongly connected to firms in the local knowledge network tends to increase innovative performance, but so do connections to extra-regional knowledge sources (Boschma and Ter Wal, 2007; Morrison et al., 2013). Firms who build linkages to actors outside the region with the purpose of learning and knowledge sharing can bring new, non-redundant knowledge to the cluster, increase international competitiveness and help to avoid the technological lock-in of the cluster (Bathelt et al., 2004). However, geographical proximity and face-to-face connections are still key to exchange, combine and re-combine tacit knowledge for innovation and learning (Audretsch and Feldman, 1996; Gertler, 2003). Empirical evidence supports these theories by illustrating that the geographical proximity of actors influences the establishment of collaboration in clusters, as physical closeness facilitates frequent face-to-face contacts and reduces the costs of creating relationships (Broekel and Hartog, 2013a; Balland et al., 2016; Juhász and Lengyel, 2018).

Besides the geographical closeness of actors, other forms of proximities can enhance collaboration and facilitate knowledge transfer in clusters (Knoben and Oerlemans, 2006; Ter Wal and Boschma, 2009; Boschma and Frenken, 2010; Caniels et al., 2014). Cognitive proximity, the similarity between the technological profile of firms makes knowledge transfer accessible as they understand each other better and can expect more accurate and useful advice for their technical problems. Empirical studies also showed that institutional proximity, the similarity in the legal forms of organisations could help collaboration and knowledge

exchange as related routines and incentive mechanisms might influence the willingness to cooperate (Broekel and Hartog, 2013a; Balland et al., 2016).

Furthermore, the actors' decision to collaborate and form knowledge ties in clusters is influenced by the network structure of existing relationships (Ter Wal and Boschma, 2009; Boschma and Frenken, 2010). Social tie formation, in general, is often path-dependent and depends on the structure of relationships itself, which applies to cluster networks as well (Glückler, 2007). Triadic closure, the notion that partners of partners become partners, enhances the formation of collaboration ties in the cluster (Giuliani, 2013; Broekel and Hartog, 2013a; Balland et al., 2016). As actors become more embedded in the network through closed triads, cohesion and the level of trust increases, which further facilitates knowledge sharing and collaboration inside the clusters. Reciprocity, the mutuality of technical advice in cluster knowledge networks, also increases the level of trust, stabilises relationships and improves the quality of interaction (Giuliani, 2013; Balland et al., 2016).

In summary, firms in clusters have to collaborate in local knowledge networks to take advantage of co-location and knowledge externalities. Most of the above literature is focused on how and why firms form linkages to exploit locally accumulated knowledge. The participation in these knowledge networks requires advanced capabilities and skills, but besides firm-level characteristics, the formation of network ties is influenced by both the similarity of firms and the network structure itself. In the next subsection, I describe the role of spinoff companies in cluster knowledge networks. Due to their inherited competences, spinoffs could possibly participate in cluster networks more easily; however, their role in knowledge networks is still not clear.

2.2.2 Spinoffs in clusters

Until recently, the explanation of industry clustering was dominated by the Marshallian view based on positive local externalities (Rosenthal and Strange, 2004; Tallman et al., 2004). As the industry started to develop in a region, local externalities were believed to support growth, further firm entries and overall prosperity in the cluster (Boschma, 2015). These arguments were challenged by Steven Klepper and his collaborators, who argued that organisational reproduction and inheritance through spinoff formation are behind the strength and success of clusters (Klepper, 2007; 2010; Buenstorf and Klepper, 2009; 2010).

As companies try to enhance their own performance through continuous technological development and organisational process improvements, successful incumbents inevitably function as training grounds for their employees, allowing them to learn the skills needed to start their own venture. Therefore, the competence of new entrants depends on their heritage, their pre-entry experiences and the accumulated industrial and organisational knowledge. In practice, these could originate from experience within existing markets and established routines, or from the tacit knowledge of employees which is transferred through

the spinoff process (Agarwal et al., 2004; Nelson and Winter, 1982). Spinoffs also tend to locate close to their geographic roots where their founders worked and collected knowledge and experiences before (Buenstorf and Klepper, 2010). The reasons why spinoffs often stay near to their parents could be the need to gain access to specialized local suppliers and services (Boschma, 2015) or the high relocation costs which include the 'opportunity cost' of losing local networks (Dahl and Sorenson, 2012). As a result, spinoffs fuel the concentration of industries through an endogenous process governed by the supply of capable entrants.

Several empirical studies have shown that spinoffs tend to perform better and have higher survival rate than other firms in clusters (e.g. Carias and Klepper, 2010; Klepper, 2010; Boschma and Wenting, 2007; Buenstorf and Guenther, 2011; Wenting, 2008). The common explanation of why spinoffs stand out in clusters is that their superior inherited capabilities make them more able to benefit from Marshallian externalities (Klepper, 2007; 2010; Buenstorf and Klepper, 2010; Cusmano and Morrison, 2015). However, there is still limited evidence on how spinoffs behave and form relationships in cluster knowledge networks to access to the accumulated industrial knowledge. Ter Wal (2013) shows, on a comparative basis, that small local spinoffs perform a key role in the establishment of dense collaboration networks in clusters. Furthermore, Bagley (2019) demonstrates that the personal ties of founders to parent companies influence the performance of spinoffs in clusters.

As previous studies highlighted that only firms with advanced capabilities are able to collaborate and share knowledge in industry clusters, I suggest that spinoff companies are more likely to form relationships in cluster networks. I assume that their inherited routines and capabilities enable them to cooperate and establish ties of knowledge sharing more easily. Because of their pre-entry experience they can be more familiar with the capabilities and knowledge of other local actors. This can help them to search for technical knowledge and find a solution in a critical situation more efficiently. Their previously acquired skills can also enable them to interact, communicate and exchange technical knowledge in a more effective way. As a consequence, spinoffs could gain more from industrial concentration and positive externalities by exploiting more knowledge through collaboration. Even though the actual capabilities and routines that spinoffs inherited are hardly observable, their significance can be inferred from looking at the effect of spinoff background on tie formation, while controlling for node level, dyad level and structural properties at the same time.

2.3 Context and data

2.3.1 A printing and paper product cluster in Hungary

The focus of this study is on a printing and paper product cluster in Kecskemét, Hungary. Kecskemét is a middle-sized town with approximately 112,000 inhabitants about 80 km south

of Budapest, the capital city of Hungary. Its economy is rooted in agriculture, processing and manufacturing industries. The printing industry has a long tradition in the region as the first printing-house, the *Petőfi Press*, was established in the 1840s and is still working under this name. After the collapse of the planned economy in Hungary, it became possible to found privately owned enterprises. As a result, by the early 1990s, numerous small and medium printing presses were founded around the town of Kecskemét.

As the old Petőfi Press functioned as a training ground for their employees, many spinoff companies emerged in the early 2000s. Combined with international companies that also located their facilities there (e.g. Ringier Axel Springer Media AG), nearly forty companies are now operating in the sector throughout the town. In 2012, the location quotient based on the number of employees showed a significant relative concentration of both the printing and service activities related to printing (LQ = 1.059) and the manufacture of articles of paper and paperboard (LQ = 4.602)[1]. The intense local competition requires flexible specialization of firms and the local industry as such. Most of the companies base their main activity on specialized technological solutions for creating unique paper products (such as specifically printed and folded paper products; stickers, labels and other packaging materials). Firms typically do not carry out R&D activities; they mainly build on customer-driven process-oriented innovations and mature technological knowledge.

As I discovered during my interviews, there is a strong informal network behind the cluster, which is characterized by the personal interactions of technicians that search for advice on technical issues. For example, they may ask for advice on specific paper types or experience with new printing machines. Altogether, the local industry can be characterized as an old social network-based cluster (Iammarino and McCann, 2006), and it provides suitable conditions for analyzing the social networks behind it.

2.3.2 Data collection

The sample selection is based on The Company Code Register (2011) by the Hungarian Central Statistical Office. This is a firm-level, nation-wide database with basic information and statistics on companies, including seat addresses and classification of main activities. All the firms with a minimum of two employees, who have company location within the urban agglomeration of Kecskemét and whose main activities are classified under the industry code 17 (Manufacture of paper and paper products) or 18 (Printing and reproduction of printed media) in the Statistical Classification of Economic Activities of Eurostat (2008) were selected. Of these, 38 firms met the above criteria and some companies with identical addresses and similar names had also merged. This resulted in a sample of 35 firms with which to begin the data collection.

[1] Detailed explanations of the location quotient and the concentration of manufacturing subsectors can be found in the Appendix (section 2.A).

The necessary relational data was collected at the firm level on the basis of face-to-face interviews. All the interviews were conducted with skilled workers (mostly with co-founders, operational managers or foremen), who were able to accurately answer the questions. The interviews were structured by a questionnaire in order to get detailed information on firms and their relationships. The data on cooperation ties was collected by the roster recall method (Wasserman and Faust, 1994) where each firm was presented with a complete list (roster) of the other firms and was asked to report about their relations to all the other firms. The following question was used to collect relational data on knowledge sharing in the cluster:

If you are in a critical situation and need technical advice, to which of the local firms mentioned in the roster do you turn?

This question is formed to collect relational data on knowledge exchange and has been used previously by numerous studies on cluster knowledge networks (e.g. Giuliani and Bell, 2005; Morrison and Rabellotti, 2009). The question targets the transfer of technical and innovation-related knowledge, and it only reveals collaboration-based problem solving and technical assistance inside the cluster (Giuliani and Bell, 2005). It is meant to capture the transfer of complex contextualised knowledge and not simply the transfer of basic information. The roster was left open in the sense that respondents could also add further contacts that did not already appear on the list. As many firms indicated knowledge exchange in relation to a company that mainly engineers knives and heavy machinery for paper product creation, I also interviewed this frequently-mentioned company.

Besides capturing interfirm knowledge ties, additional questions were formed to collect data on firm size, interregional knowledge ties, ownership and spinoff background of firms. It is possible to characterize three types of entrants into an industry in terms of their pre-entry experience and inherited competences: startups, diversifiers and spinoffs (Buenstorf and Klepper, 2010). In this study, I focus only on spinoff companies and distinguish firms only as spinoffs or non-spinoffs.

2.3.3 Descriptive statistics of the sample

I acquired 26 responses from companies, which resulted in an above 70% coverage of the original selection. As the list of firms was compiled on the basis of a dataset from 2011, it turned out that eight of the companies had closed down or had temporarily suspended their business activities. Only one actor refused to answer the questions and all of the non-respondents were domestic small- and medium-sized enterprises (SMEs). I also compared the average years in the industry and the average size of firms between non-respondents and the final sample and obtained very similar values for both groups. As a result, a ‘non-respondent bias’ was not detected (Armstrong and Overton, 1977; Lambert and Harrington, 1990).

Moreover, I encouraged firms to mention knowledge exchange with any other companies in the region not presented in the roster. As a result, I also interviewed the one frequently mentioned firm. I believe that I captured all the companies in the printing and paper product industry around Kecskemét and the data collection was inclusive.

As Table 2.1 shows, the majority of the companies in the final sample are domestic SMEs. There is only one firm with more than 100 employees and only a minority of them are foreign-owned (less than 25%). The average number of extra-regional knowledge ties are 7.4 and apart from three companies, all firms exchange knowledge with other firms outside the region. The average amount of time firms had spent in the industry was above fourteen years, and more than 75% of the companies had above ten years' experience in the printing and paper product industry. In addition to this, 11 out of 26 of the actors interviewed said that their company was a spinoff and the majority of them verified themselves as spinoffs of Petőfi Press, which is the oldest and still the most prominent printing press in the region. Table 2.1 also shows that the group of spinoffs and non-spinoffs have relatively similar characteristics at the firm level.

Table 2.1 Firms' characteristics

	Spinoffs (n=11)	Non-spinoffs (n=15)	Full sample (n=26)
Ownership			
Foreign	2	4	6
Domestic	9	11	20
Extra-regional knowledge ties			
0	2	1	3
1-10	8	10	18
11-50	1	4	5
Years in the industry (Age)			
1-10	3	5	8
11-20	6	7	13
21-25	2	3	5
Employees			
2-5	5	7	12
6-50	4	6	10
51-400	2	2	4

Source: Author's own data.

Based on the question about knowledge transfer between firms, I construct a directed adjacency matrix with *n* rows and *n* columns (where *n* stands for the number of respondents). In this matrix, each cell indicates the transfer of knowledge from the firm *i* in the row to firm *j* in the column. The cell (*i, j*) contains the value of 1 if the company *i* have given advice to company *j* and contains the value of 0 in cases where no knowledge has been transferred. Based on this adjacency matrix I created a directed, unweighted graph which represents the knowledge network behind the cluster.

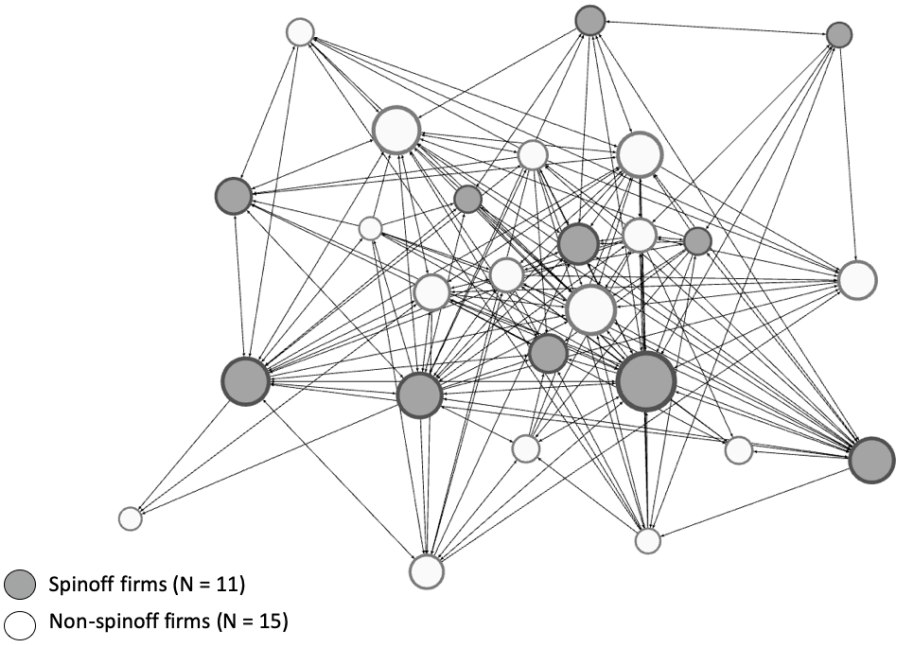
Table 2.2 Descriptive statistics of the knowledge network

Attribute	Value
Nodes	26
Edges	223
Density	0.343
Average degree	17.154
Average distance	1.72

Source: Author's own data.

The knowledge network of the printing and paper product cluster in Kecskemét consists of 26 nodes and 223 edges. In this relatively dense network, actors exchange knowledge with around 17 other firms and the distances between companies are very short. The visual representation also suggests (Figure 2.1) that the cluster is based on an intensive, informal collaboration network where some of the most influential actors are spinoffs.

Figure 2.1 The local knowledge network of the printing and paper product industry in Kecskemét in 2012



Source: Author's own data.

Note: The size of the nodes is proportional to in-degree.

2.4 Methodology and variables

2.4.1 ERGMs – Exponential Random Graph Models

In order to better understand the role of spinoffs in the formation of cluster knowledge networks, exponential random graph models (ERGMs) are applied (Snijders et al., 2006; Lusher et al., 2013). ERGMs are stochastic models that approach tie formation as a time-continuous process. They are built on an observed network at one point in time, which is a particular realization out of a set of hypothetical networks with similar properties. The aim of ERGMs is to identify the factors that maximize the probability of the emergence of a network with the same properties as the structure of the observed network (Broekel et al., 2014). The general form of exponential random graph models is as follows (after Robins et al., 2007a):

$$\Pr(X = x) = \left(\frac{1}{k}\right) \exp\left\{\sum_{A} \eta_A g_A(x)\right\} \tag{2.1}$$

where the summation is over all configurations *A*. η_A is the parameter corresponding to configuration *A* (and is non-zero only if all pairs of variables in *A* are assumed to be conditionally dependent on the rest of the graph). These configurations can contain factors related to the node level, dyad level and structural level. $g_A(x) = \prod_{x_{ij} \in A} x_{ij}$ is the network statistic corresponding to configuration *A*. $g_A(x) = 1$ if the configuration is observed in the network x, and is 0 otherwise. k is a normalizing constant ensuring that the equation is a proper probability distribution (summing up to 1). It is defined as

$$k = \sum_{X} \exp\left\{\sum_{A} \eta_A g_A(x)\right\} \tag{2.2}$$

where $X(n)$ represents all the possible networks with *n* nodes. Consequently, the probability of observing any particular graph x in this distribution is given by the equation, and this probability $\Pr(X = x)$ depends on the network statistics $g_A(x)$ in the network x and on the parameters represented by η_A for all considered configurations *A*. The value of η_A indicates the impact of the configuration on the log-odds of the appearance of a tie between two nodes.

In an ERGM estimation, the equation is solved such that parameter values are identified for each configuration that maximizes the probability that the simulated network is identical

to the empirically observed one. This is achieved by Markov chain Monte Carlo maximum likelihood estimation (for more details see Snijders, 2002; van Duin et al., 2009). The procedure is based on the generation of random graphs by stochastic simulation from a starting set of parameter values. These parameter values are subsequently refined through the comparison of the obtained random graphs against the observed graph. The process is repeated until the parameter estimates stabilize. In cases where they do not, the model might prove to be unstable and fail to demonstrate convergence (for more technical details see Hunter et al. 2008). The above-described procedure is implemented in the *statnet* R package (see Robins et al., 2007b; Goodreau et al., 2008).

In order to check whether the given parameters predict the observed network well, the 'goodness of fit' test is carried out in order to compare the structure of the simulated networks to the structure of the observed network. Suggested by Hunter et al. (2008) the comparison is usually made on the basis of degree distribution, distribution of edgewise shared partners (the number of links in which two actors have exactly k partners in common, for each value of k) and the minimum geodesic distance (the number of pairs for which the shortest path between them is of length k, for each value of k). The more these statistics are similar for the estimated and observed networks, the more accurate and reliable estimated parameters of the ERGMs are. Additionally, along with the iterations, the simulated parameter values should be relatively stable and vary more or less around the mean value (Goodreau et al., 2008)[2]. In case of a successful estimation, the given parameters of the ERGM can be interpreted as non-standardized coefficients obtained from logistic regression analysis, which can be transformed into odds ratios.

Through ERGMs I can examine how node level, dyad level and network level characteristics influence tie formation in a network, observed only at a certain point in time. Because of these relatively low-level barriers, the application of ERGMs became popular in social sciences. However, only a handful of studies use ERGMs on networks of innovation and knowledge (Broekel and Hartog, 2013b; Capone and Lazzeretti, 2016; De Stefano and Zaccarin, 2013) and only a minority of them focuses on cluster knowledge networks (Broekel and Hartog, 2013a; Molina-Morales et al., 2015, Capone and Lazzeretti, 2018).

2.4.2 Variables

It is possible to include variables to ERGMs in three different levels. Node level variables capture how individual properties influence tie formation. Dyad level variables capture how similarity of actors influences the probability that actors form ties. Structural level variables try to capture the influence of network structure on tie formation. In order to observe how

[2] All the suggested goodness-of-fit statistics in relation to the final model are presented in the Appendix (section 2.G).

spinoffs influence learning and knowledge sharing in clusters, I do not only use node level variables but also controls for several dyadic and structural effects as well.

The most important node level factor in this study is whether a company is established as a spinoff or not. Through a node level spinoff dummy, I examine how spinoff background influence tie formation in the cluster knowledge network. I expect that spinoffs form significantly more ties in the cluster than they would in a random setting. Additionally, firm-level control variables such as ownership, age, external knowledge ties and number of employees are included. By adding a foreign company dummy I control for group homophily, or in other words, whether firms tend to share knowledge in their own ownership group. Firms are foreign-owned in my sample if they are at least partly owned by foreign companies. Age refers to the years of experience in the printing industry. External knowledge ties as possible sources of new knowledge and novel technological solutions are key in the cluster literature (Bathelt et al., 2004; Glückler, 2007; Morrison, 2008). To measure the importance of extra-regional relationships as a node level characteristic the number of external knowledge ties (meaning all the links to other regions in Hungary or abroad) are included. The number of employees in a logarithmic form is used to provide a control for the size of the firms, as it could determine firms' ability to acquire knowledge in clusters (Parra-Requena et al., 2010).

Two control variables are also included at the dyad level. The geographical proximity of firms is measured as the distance of the selected pair of firms subtracted from the maximum physical distance between firms in the cluster. As a result, the variable takes a higher value as the distance between firms diminishes (Juhász and Lengyel, 2018). Cognitive proximity is made operational as the number of digits the two firms have in common in their 4-digit NACE codes (Balland et al., 2016). This measure assumes that the technological profiles of firms have greater similarity and therefore that they are cognitively closer if they operate in the same sector category (Frenken et al., 2007)[3].

To control for the structural dependencies, three variables at the network level are included. Triadic closure is captured by the geometrically weighted edgewise shared partner statistics (GWESP). They measure the number of triangles in the network while taking into account the number of ties that are involved in multiple triangles and hence connect the same neighbours in multiple triads (Hunter et al., 2008). In cases when the parameter is positive and significant, there is a tendency toward triadic closure in the network. Geometrically weighted dyad-wise shared partners (GWDSP) are also included. They measure the extent to which nodes are not directly linked to each other, being at least indirectly linked (Snijders et al., 2006). In other words, they capture the multi-connectivity of nodes that are not directly linked through the approximation of indirect existing paths between such nodes. Geometrically weighted in-degree statistics (GWIDEGREE) are also included to model the observed network's in-degree distribution. They allow us to model the preferential attachment processes.

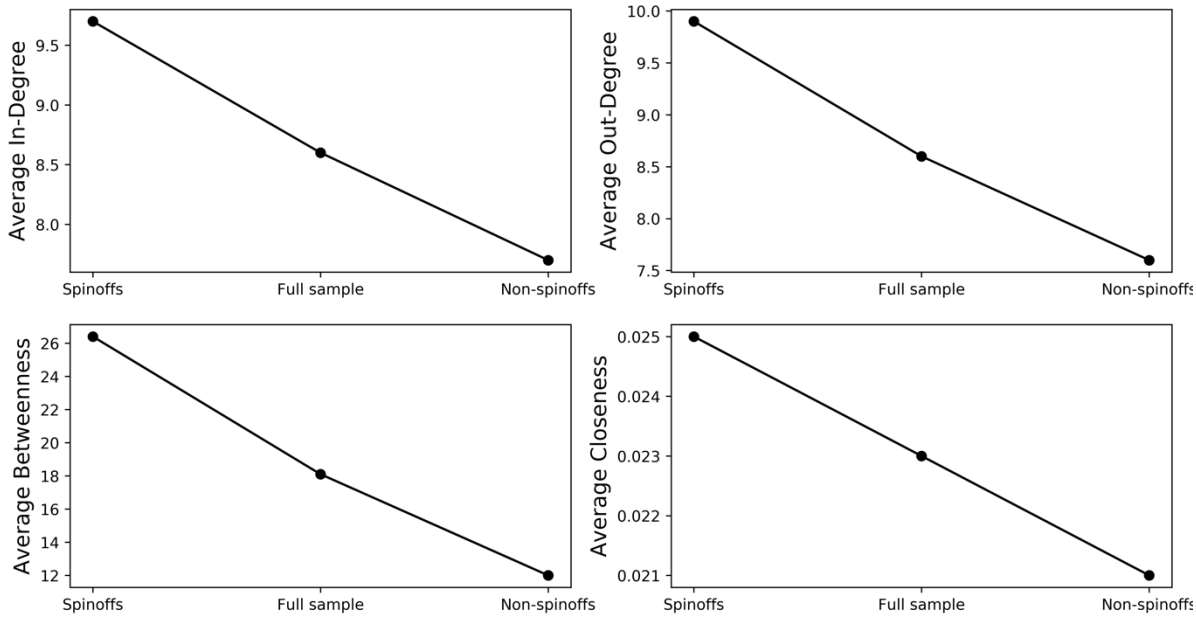
[3] The descriptive statistics of the applied geographical and cognitive proximity measures are provided in the Appendix (section 2.B).

Moreover, two baseline structural control variables are also included. EDGES equal the number of links in the network and are used as overall network structural controls. They compare the density of the observed network to the density of the simulated exponential random graphs. MUTUAL ties provide controls for reciprocity in tie formation. Both of these effects are necessary to control for in every (directed) ERGM.

2.5 Results

Figure 2.2 presents some general descriptive statistics on the position of spinoff companies in the knowledge network. It shows that spinoffs on average have more incoming and outgoing ties than other actors. This suggests that spinoffs tend to form more ties to exchange technical knowledge within the cluster. Spinoff companies also have higher than the average value of 'betweenness'. The measure is based on the number of shortest paths a firm is involved in between two randomly chosen actors. The higher than average betweenness value of spinoff companies suggest that they are more important in terms of connecting other firms in the cluster knowledge network. Closeness is calculated as the reciprocal of the sum of the length of the shortest paths between the firm and all other actors in the network. The higher average closeness value of spinoffs suggests that they can have faster, more direct access to other companies in the knowledge network.

Figure 2.2 Descriptive statistics on the network position of spinoff companies



Source: Author's own data.

Note: Dots represent group average of in-degree, out-degree, betweenness centrality and closeness centrality for spinoff companies, for the full sample and for non-spinoff companies.

These descriptive statistics suggest that spinoffs are on average more embedded in the cluster knowledge network. However, by the application of ERGMs, I was able to test whether spinoff background matters for knowledge tie formation, while I also provide controls for structural, dyadic and node level characteristics at the same time.

Table 2.3 Results of the exponential random graph model

	Main model		Refined model	
	Coefficient	(SE)	Coefficient	(SE)
Spinoff	0.3699 **	(0.124)	0.4272 ***	(0.1286)
Ownership group	0.5120 ***	(0.1536)	0.4939 ***	(0.1442)
Extra-regional knowledge ties	0.0128 *	(0.0063)	0.0149 *	(0.0062)
Age (experience)	0.0204	(0.0434)		
Employment (log)	0.0204	(0.0434)	0.0133	(0.0474)
Geographical proximity	0.1059 **	(0.0355)	0.1214 ***	(0.0361)
Cognitive proximity	0.0121	(0.0436)	0.0171	(0.0422)
GWESP (fixed 0.32)	1.7159 ***	(0.4722)	1.9596 ***	(0.4640)
GWDSP (fixed 1.725)	-0.1769 ***	(0.0487)	-0.1858 ***	(0.0491)
GWIDEGREE (fixed 0.1325)	-0.7690	(0.5415)		
MUTUAL ties	1.6632 ***	(0.2737)	1.5988 ***	(0.2681)
EDGES	-4.2661 ***	(0.7913)	-4.6813 ***	(0.7526)
AIC	733.6		731.9	
BIC	787.3		776.7	

Source: Author's own data.

Note: Significance codes: *** p<0.001, ** p<0.01, * p<0.05

To begin at the node level, the main variable on spinoff companies turned out to be positive and significant. It means that spinoff companies form significantly more ties than non-spinoffs while controlling for other factors. This is in line with my expectation and suggests that spinoffs are more capable to collaborate and form linkages to exchange knowledge in the cluster knowledge network. The wider consequences and possible reasons for this result are discussed in the conclusions section (Section 2.6). The similarity of ownership has a positive and significant effect too. This indicates that knowledge ties are more easily formed within the group of domestic or foreign firms than across these groups. The reason behind this notion could be the barriers of language or the technological gap between foreign and domestic companies. This finding underlines previous results related to the importance of ownership structure in knowledge spillover effects (Elekes and Lengyel, 2016). External knowledge ties turned out to have positive, but barely significant effect on local tie formation. It means that extra-regional knowledge ties are important, but do not necessarily influence local knowledge sharing. The age and size of firms do not determine their abilities to form local knowledge ties. To provide a more accurate estimation, years of experience were excluded from my final model, which further refined the results as smaller AIC and BIC statistics indicate.

At the dyad level, geographical proximity is characterised by a positive and significant coefficient. This finding suggests that physical proximity helps the formation of knowledge ties in clusters and further emphasises the importance of micro-level geography on knowledge sharing (e.g. Broekel and Boschma, 2012; Balland et al., 2016). Cognitive proximity, on the other hand, does not influence tie formation in this case. This suggests that firms with a similar technological profile are less likely to collaborate and exchange knowledge. Similar effects were found for cognitive proximity previously by Broekel and Hartog (2013a) who also applied ERGMs to the knowledge network of the Dutch aviation industry. However, the results are not in line with the growing literature that finds the positive influence of cognitive proximity for tie formation of cluster networks (e.g. Broekel and Boschma, 2012; Balland et al., 2016; Lazzeretti and Capone, 2016).

At the structural level, the formation of triads measured by GWESP statistics turned out to be positive and significant, confirming the fact that a relatively large number of triangles exist in the network. In other words, firms that are directly linked are also more likely to link through indirect connections. The negative and significant GWDSP statistics indicate that two firms without a direct link are less likely to be indirectly connected. The coefficient of GWIDEGREE is negative and not significant which suggests that there is no preferential attachment mechanism in this case. To improve the main model, I removed GWIDEGREE from my model setting, which resulted in a much more accurate estimation. Furthermore, the positive and significant effect of MUTUAL ties suggests that firms reciprocate advice and knowledge sharing. The negative and significant coefficient of EDGES indicates that the network is less dense than exponential random networks, which is a common feature of networks representing social interactions (Snijders et al., 2006). In summary, the models suggest that controlling for node level, dyad level and structural level interdependencies, spinoff companies form significantly more ties than other firms.

2.6 Conclusions

The main aim of this study was to present the influence of spinoff companies on knowledge tie formation in clusters. Applying ERGMs to capture how node level, dyad level and structural level factors influence collaboration in the cluster, I show that spinoffs form more local knowledge ties than other actors. As spinoffs collaborate and share knowledge more easily, they might as well gain more from locally accumulated knowledge. This suggests that the pre-entry experience, inherited routines and capabilities of spinoffs indeed influence their ability to cooperate and exchange innovation-related knowledge, and thus are better able to exploit the positive externalities of co-location. This study has also contributed to the emerging empirical studies on knowledge network formation in clusters (e.g. Giuliani, 2013; Broekel and Hartog, 2013a; Balland et al., 2016; Capone and Lazzeretti, 2018) and has provided further

insights on the influence of company ownership, geographical closeness and common third partners on knowledge tie formation between firms.

The study also aimed to connect the literature on informal knowledge networks to the literature on spinoffs in clusters. Pioneer studies on cluster networks found that firms' ability to absorb knowledge is important for collaboration and knowledge-sharing (Giuliani and Bell, 2005; Giuliani, 2007). The present chapter would extend this line of argument by showing that inherited capabilities and previous experiences could also increase the abilities of firms to participate in cluster knowledge networks. Even though the results suggest that there is a strong relationship between spinoff background and knowledge sharing, further research is needed to understand the particular capabilities that enable spinoffs to be more capable of collaboration and knowledge sharing in cluster networks.

Despite the fact that the current study was unable to include variables on the specific inherited capabilities of companies that might influence their ability to form ties, the face-to-face interviews with firm representatives provided several points to discuss. Through these conversations, I got to know more about the history of the local printing industry and the role of spinoff companies on informal network formation. An interesting conclusion resulting from the conversations is that since most of the actors know the other firms' representatives personally, there is a strong informal network behind the cluster. This shed light on the importance of historical aspects. I am becoming increasingly aware that a great number of companies started as a spinoff of the old Petőfi Press where their founders learned the profession and most of them know each other personally from those formative learning years. Many of the interviewed representatives explained that they accumulated relevant experience, up-to-date professional knowledge and personal relationships as an employee, which enabled them to start their firm as a spinoff in an industry with which they were already familiar. They also explained that the confidence in their own technological capabilities helped them to bravely interact and ask for technical help at the beginning of their career. This reinforced my impression that pre-entry experience and inherited capabilities are indeed essential for collaboration and also led me to the conclusion that spinoffs inherit personal relationships too, which might well influence their ability to collaborate and share knowledge in the future.

Since this exercise is based on a printing and paper product cluster in its later life-cycle stage, conclusions might be seen as being limited to clusters in traditional manufacturing industries. In addition, an important issue for future research is the extent of access to detailed, longitudinal data on knowledge transfer. The analysis here is based on relational data from only one point in time, however,more detailed, longitudinal datasets on knowledge exchange in clusters could help to answer several, still open, questions. First of all, the incorporation of firm entry, exit or the differentiation in generations of spinoffs, could help to change our understanding of network formation drastically. It could also make clear, whether spinoff formation is a result of informal collaborations or rather of a more intense knowledge sharing environment emerging from spinoff foundations. Detailed information on parent

companies would make it possible to test whether firms that gave birth to spinoffs are more capable of collaborating in the local knowledge network. Secondly, by using relational data on the individual level rather than firm-level I might understand more accurately the motivations behind tie formation. Moreover, the direct measurement of how inherited network ties matter for industry clusters is also a promising area for future research (Bagley 2019). Further insights on the extent to which firms and entrepreneurs build on inherited capabilities, routines and previous experiences would help us to understand better why spinoffs form more ties than other companies in clusters.

In this study, I have only emphasized the role of spinoff companies in a cluster knowledge network. As recent studies have highlighted, structural, dyadic and actor level characteristics might play a distinct role in the formation of different formal and informal cluster network types (Capone and Lazzeretti, 2018; Balland et al., 2016; Ferriani et al., 2013). Therefore, it would be worth investigating the importance of spinoff companies in networks of business interaction, information flow, friendship and innovation. Further, I was unable to make any differentiation between the transmitted content of collaboration ties. However, the volume, diversity and depth of transferred information content could allow us to investigate how the value of advice influences tie formation (Aral and Van Alstyne, 2011). Moreover, it would be interesting to see how the network position of spinoffs influence both their survival and the long-term prosperity they experience in cluster environments.

2.7 Appendix

2.A Concentration of manufacturing industries in the urban agglomeration of Kecskemét, Hungary

Concentration of economic activities in the urban agglomeration of Kecskemét was measured by the location quotient (LQ). The basic equation for LQ is the following:

$$LQ = \frac{\frac{E_{i,j}}{E_j}}{\frac{E_{i,n}}{E_n}} \tag{2.3}$$

, where $E_{i,j}$ stands for the number of employees in industry *i*, in region *j*. E_j stand for the overall employment in region *j*. $E_{i,n}$ represents the number of employees in the industry *i* countrywide and E_n stand for the number of employees in every sector in the country. In other words, the numerator is the share of a given industry in the region and the denominator is the share of this industry in the overall country. This index is also known as the index of Revealed Comparative Advantage (RCA) or the Hoover-Balassa index following Balassa (1965).

Concentration can be measured on different basis, such as on employment, export or the number of companies. In this case I choose to measure the concentration of employment in different sectors. I excluded firms with 1 or less then 1 employee, similarly to Lengyel et al. (2018). Higher than 1 values of the index represent relative high concentration. This study focuses on printing and paper product creation in the urban agglomeration of Kecskemét. The sectors have relative high concentration (see Table 2.4 and Table 2.5), tradition and their value chain is typically related (see EC 2013).

Table 2.4 Location quotients (LQ) based on the number of employees in all the manufacturing industries in the urban agglomeration of Kecskemét, Hungary

Code	Manufacturing (subsections)	LQ
CA	Manufacture of food products, beverages and tobacco products	1.884
CB	Manufacture of textiles, apparel, leather and related products	0.610
CC	Manufacture of wood and paper products, and printing	1.661
CD	Manufacture of coke, and refined petroleum products	0.000
CE	Manufacture of chemicals and chemical products	0.085
CF	Manufacture of pharmaceuticals, medicinal chemical and botanical products	0.000
CG	Manufacture of rubber and plastics products, and other non-metallic mineral products	0.848
CH	Manufacture of basic metals and fabricated metal products, except machinery and equipment	0.990
CI	Manufacture of computer, electronic and optical products	0.610
CJ	Manufacture of electrical equipment	1.682
CK	Manufacture of machinery and equipment n.e.c.	0.863
CL	Manufacture of transport equipment	1.179
CM	Other manufacturing, and repair and installation of machinery and equipment	0.558

Source: Author's own calculation based on the database of Hungarian Central Statistical Office (CompanyCodeRegister 2010)

Table 2.5 Location quotients (LQ) based on the number of employees in printing and paper product industry in the urban agglomeration of Kecskemét, Hungary

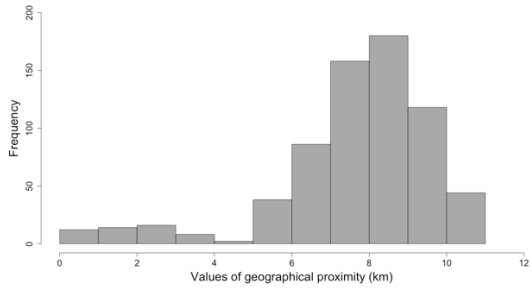
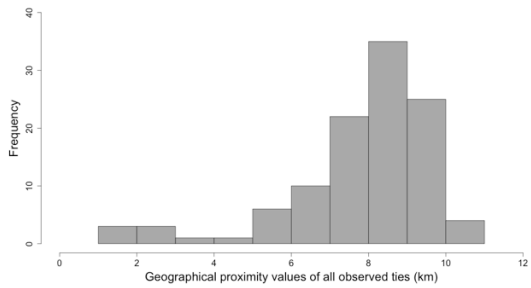
	Sectors	LQ
16	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials	2.348
16.1	Sawmilling and planing of wood	1.515
16.2	Manufacture of products of wood, cork, straw and plaiting materials	2.699
17	Manufacture of paper and paper products	3.777
17.1	Manufacture of pulp, paper and paperboard	0.000
17.2	Manufacture of articles of paper and paperboard	4.602
18	Printing and reproduction of recorded media	1.048
18.1	Printing and service activities related to printing	1.059
18.2	Reproduction of recorded media	0.393

Source: Author's own calculation based on Nace Rev 2 and the database of Hungarian Central Statistical Office (CompanyCodeRegister 2010)

2.B Descriptive statistics of the applied dyadic variables

Both of the presented ERGMs contain two dyad level control variables, the geographical and cognitive proximity of firms. Geographical proximity is measured as the distance of the selected pair of firms subtracted from the maximum distance between any firms in the cluster. In this setting the variables take a higher value as the physical distance between companies diminishes. The distribution of all the values of geographical proximity and the distribution of geographical proximity values in case of observed knowledge ties are presented in Table 2.6.

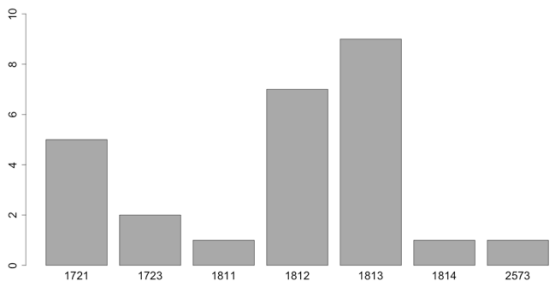
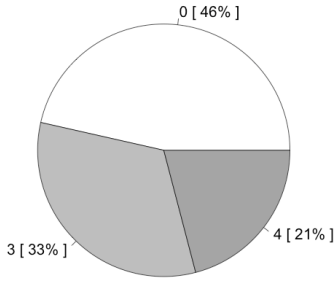
Table 2.6 Descriptive statistics of geographical proximity

All the values of geographical proximity included to the ERGMs	Geographical proximity values of observed network ties
	

Source: Author's own data.

The applied cognitive proximity measure is based on the number of digits the two firms have in common in their 4-digit NACE codes. This measure assumes that two firms have more similar technological profile and therefore are cognitively closer if they operate at the same sector category. The final sample consist of 26 companies and only 21% of all the possible firm pairs have cognitive proximity values of 4, 33% have 3 and 46% have 0 as cognitive proximity measure. As the cognitive proximity values of 1 and 2 are missing, I also tested both the main model and the refined model with the cognitive proximity values of 0-1-2 as robustness checks. Results showed no difference.

Table 2.7 Descriptive statistics of cognitive proximity

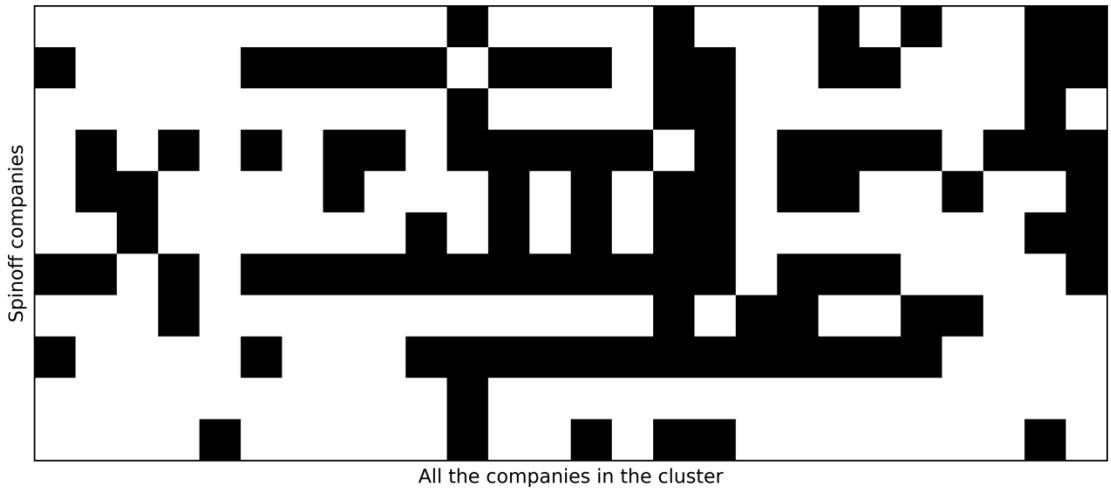
Number of firms within each sectoral category	Cognitive proximity values of all possible company pairs
	

Source: Author's own data.

2.C Patterns in advice seeking of spinoff companies

Figure 2.3 is a slice of the adjacency matrix where rows represent how spinoff companies ask for technical knowledge and columns represent all the firms as potential knowledge sources in the cluster. The aim of the figure is to visualize that spinoff companies do not follow the same pattern to seek for advice. Even though the old Petőfi Press is the parent company of the majority of the spinoffs, it seems the they do not have identical relationships.

Figure 2.3 Visual representation of how spinoffs seek for advice in the cluster



Source: Author's own data.

2.D Baseline model without controlling for spinoff companies

Table 2.8 Results of the exponential random graph models – including Baseline model

	Baseline model		Main model		Refined model	
	Coefficient	(SE)	Coefficient	(SE)	Coefficient	(SE)
Spinoff			0.3699**	(0.124)	0.4272***	(0.1286)
Ownership (same)	0.5053***	(0.0000)	0.5120***	(0.1536)	0.4939***	(0.1442)
Extra-regional knowledge ties	0.0069	(0.0054)	0.0128*	0.0063)	0.0149*	(0.0062)
Age (experience)	0.0074	(0.1482)	0.0204	(0.0434)		
Employment (log)	0.0470	(0.0429)	0.0204	(0.0434)	0.0133	(0.0474)
Geographical proximity	0.0608*	(0.0311)	0.1059**	(0.0355)	0.1214***	(0.0361)
Cognitive proximity	0.0300	(0.0919)	0.0121	(0.0436)	0.0171	(0.0422)
GWESP (fixed 0.32)	1.7513***	(0.4821)	1.7159***	(0.4722)	1.9596***	(0.4640)
GWDSP (fixed 1.725)	-0.1467**	(0.0499)	-0.1769***	(0.0487)	-0.1858***	(0.0491)
GWIDEGREE (fixed 0.1325)	-1.1127*	(0.5436)	-0.7690	(0.5415)		
MUTUAL ties	1.6855***	(0.2748)	1.6632***	(0.2737)	1.5988***	(0.2681)
EDGES	-4.0153***	(0.8234)	-4.2661***	(0.7913)	-4.6813***	(0.7526)
AIC	741.7		733.6		731.9	
BIC	790.9		787.3		776.7	

Source: Author's own data.

Note: Significance codes: *** p<0.001, ** p<0.01, * p<0.05

2.E Controlling for the spinoffs of Petőfi Press

As a robustness check I run the same models, but in a setting where I only consider the spinoffs of Petőfi Press, the oldest and largest printing press in the region. Only 5 companies confirmed to be a spinoff of Petőfi Press. The results are similar to the original version, however, variables are less significant.

Table 2.9 Results of the exponential random graph models considering Petőfi spinoffs

	Main model		**Refined model**	
	Coefficient	(SE)	Coefficient	(SE)
Petőfi spinoff	0.3317*	(0.1527)	0.3558*	(0.1501)
Ownership (same)	0.4590**	(0.1558)	0.4289**	(0.1386)
Extra-regional knowledge ties	0.0085	(0.0057)	0.0110	(0.0057)
Age (experience)	0.0083	(0.0074)		
Employment (log)	0.0391	(0.0441)	0.0294	(0.0452)
Geographical proximity	0.0992**	(0.0369)	0.1194**	(0.0377)
Cognitive proximity	0.0571	(0.0951)	0.0742	(0.0921)
GWESP (fixed 0.32)	1.7636***	(1.7636)	2.0705***	(0.4822)
GWDSP (fixed 1.725)	-0.4685***	(0.4685)	-0.1724***	(0.0488)
GWIDEGREE (fixed 0.1325)	-0.1676	(0.1676)		
MUTUAL ties	1.6786***	(0.2834)	1.5973***	(0.2664)
EDGES	-4.2871***	(0.7889)	-4.7486***	(0.7811)
AIC	739.3		738.9	
BIC	793.0		783.7	

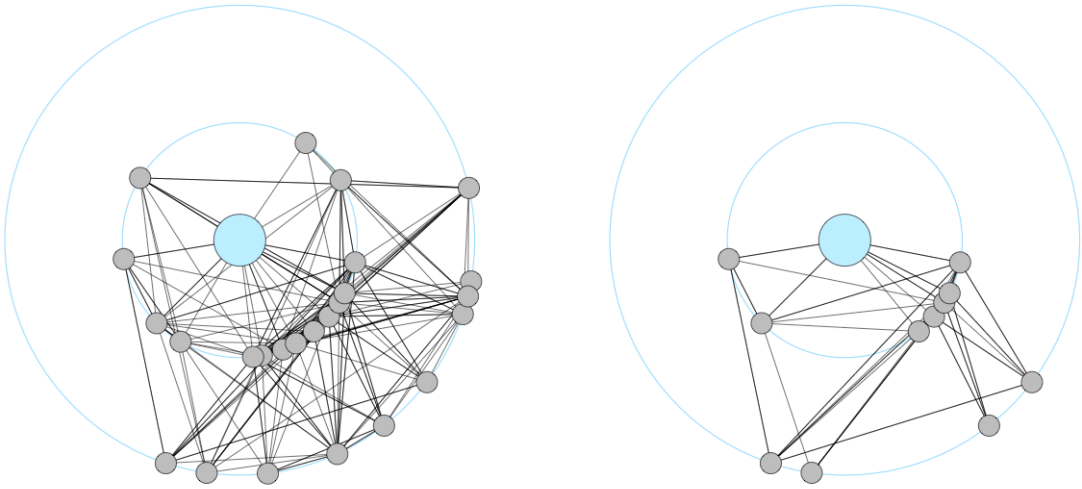
Source: Author's own data.

Note: Significance codes: *** p<0.001, ** p<0.01, * p<0.05

2.F Spinoffs and Petőfi Press in the local knowledge network

Figure 2.4 The old Petőfi Press as a focal actor and spinoff companies in the local knowledge network of the printing and paper product industry in Kecskemét in 2012

Full network	Petőfi Press and spinoff firms

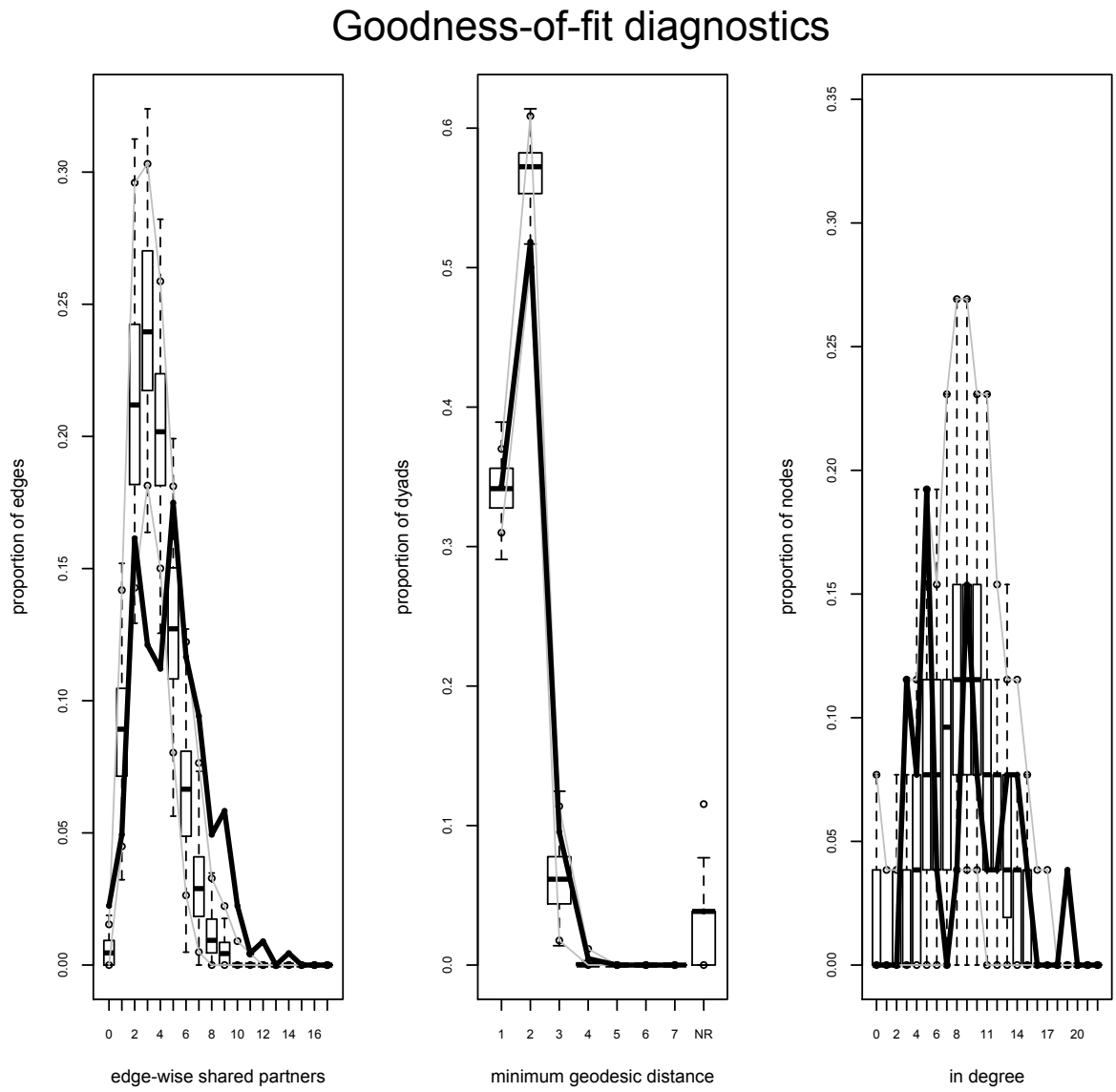


Source: Author's own data.

Note: The highlighted focal actor is the oldest press in the region, which is the parent company of the majority of spinoffs.

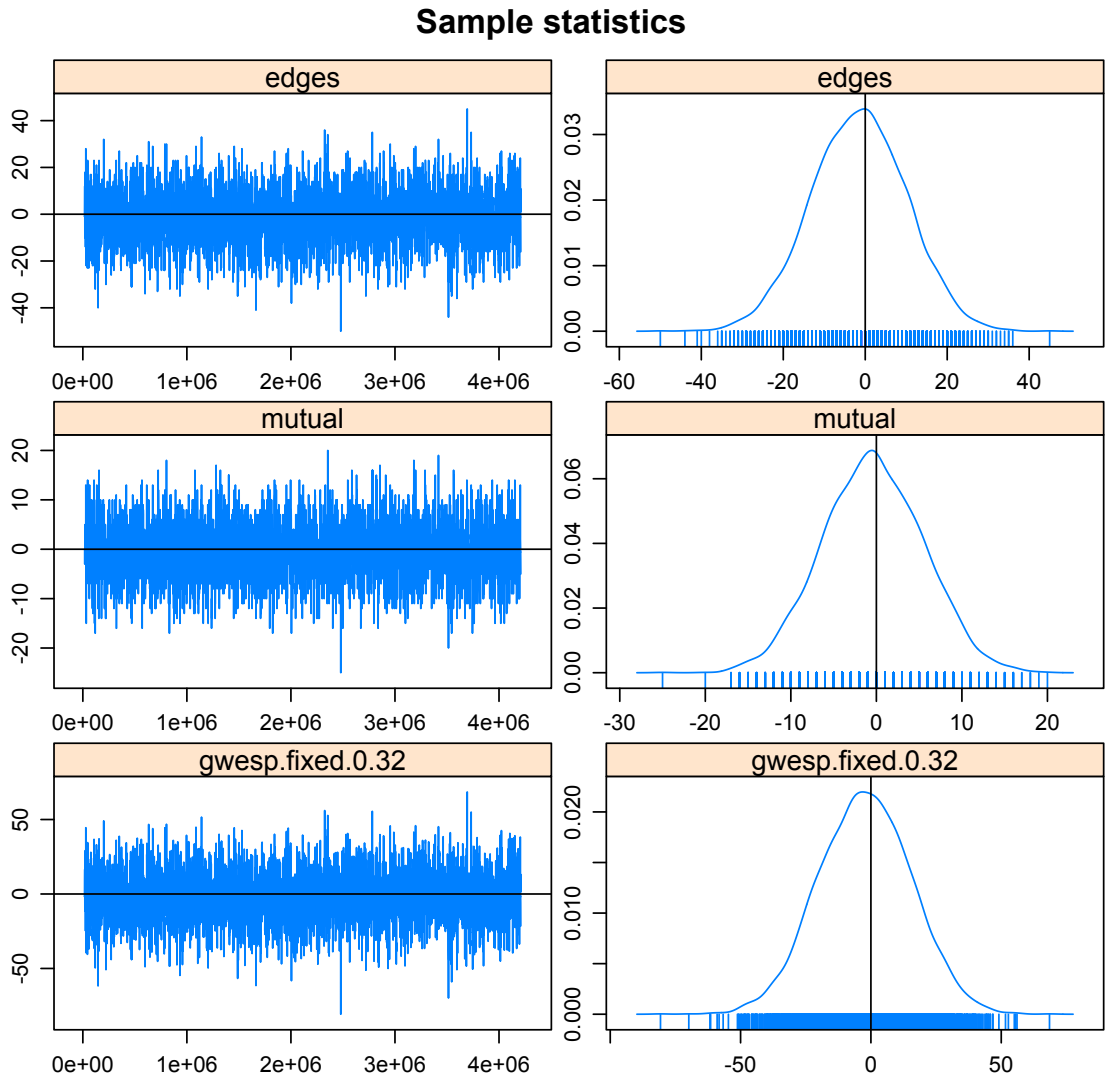
2.G Goodness-of-fit statistics for the applied ERGMs

Figure 2.5.1 Goodness-of-fit for the refined ERGMs



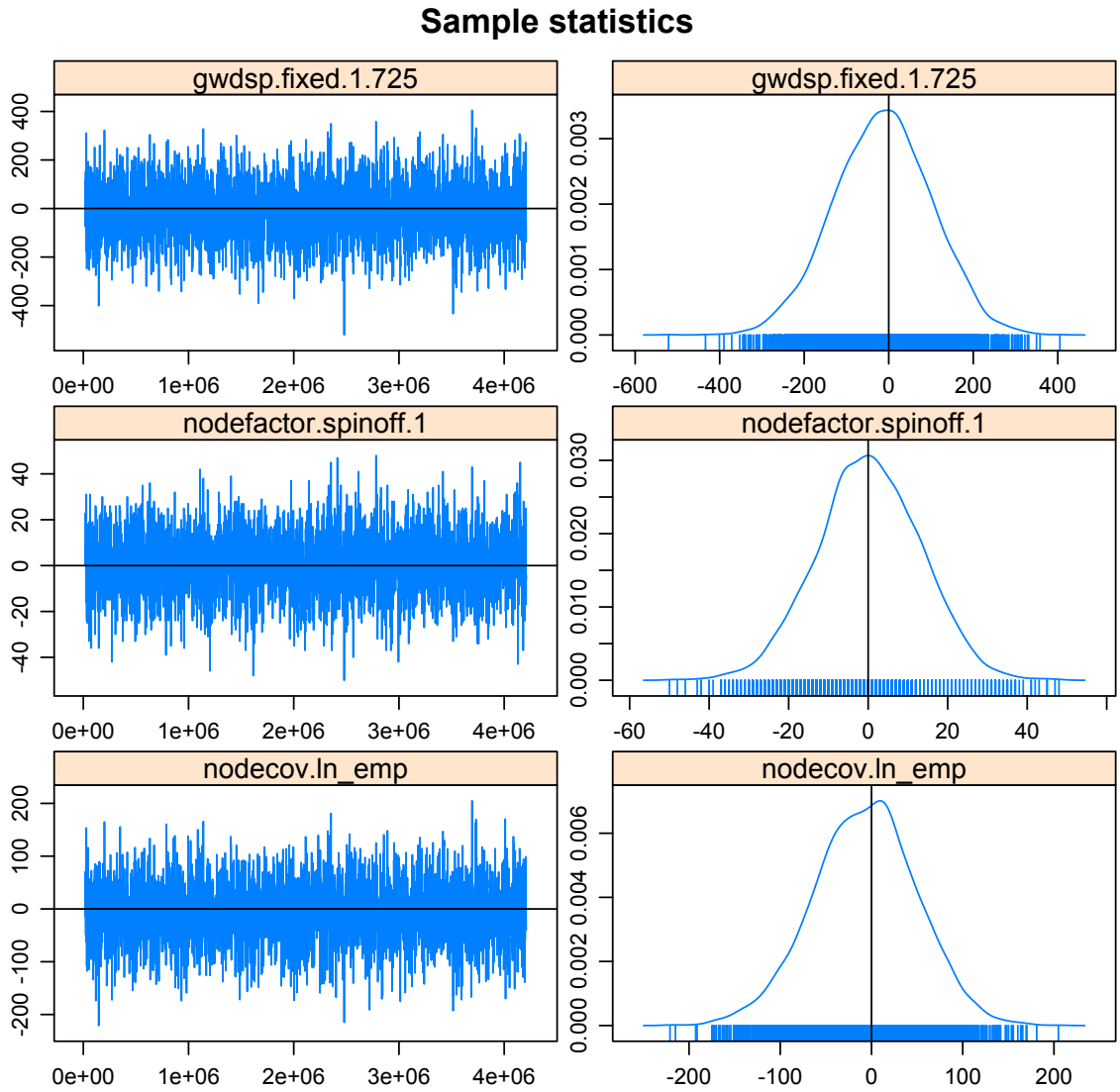
Source: Author's own data.

Figure 2.5.2 Goodness-of-fit for the refined ERGMs



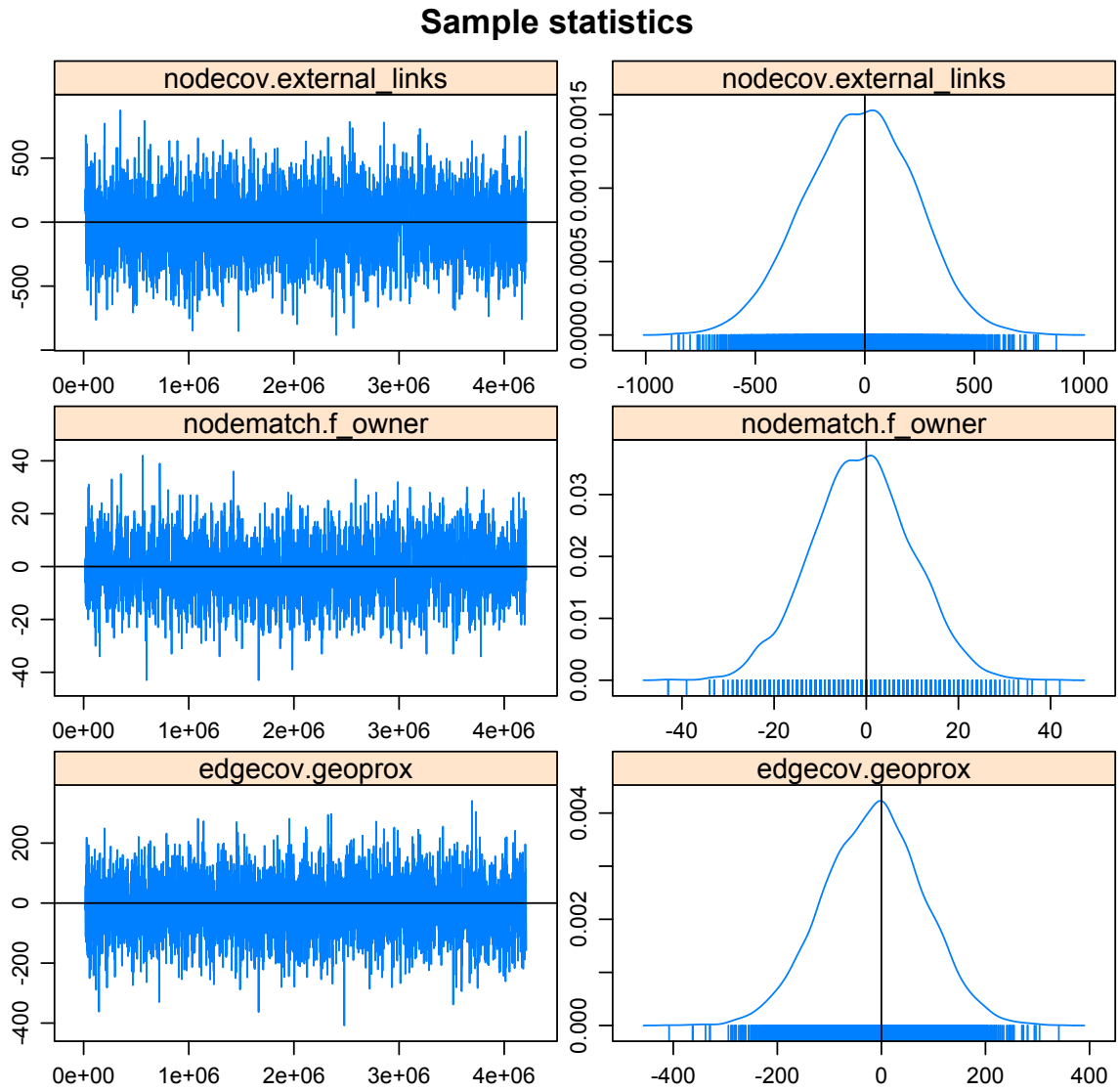
Source: Author's own data.

Figure 2.5.3 Goodness-of-fit for the refined ERGMs



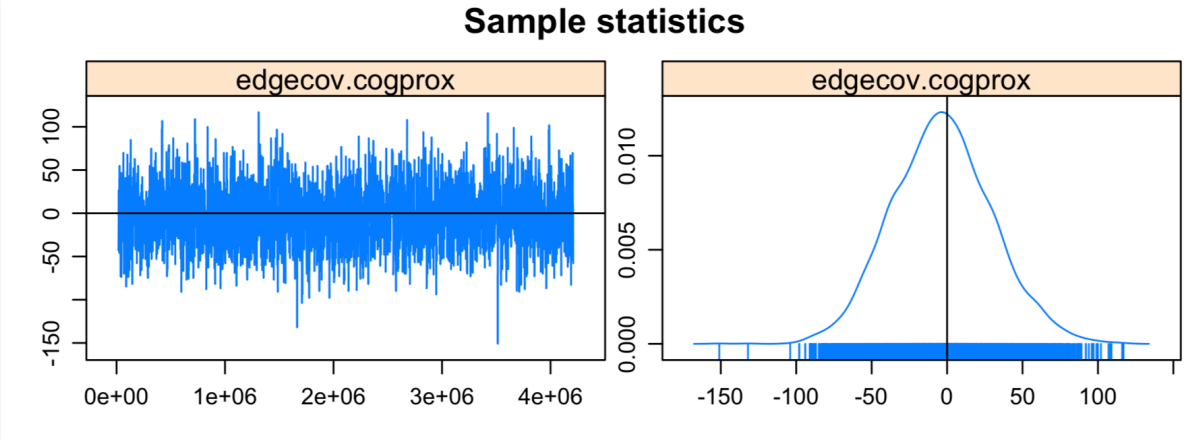
Source: Author's own data.

Figure 2.5.4 Goodness-of-fit for the refined ERGMs



Source: Author's own data.

Figure 2.5.5 Goodness-of-fit for the refined ERGMs



Source: Author's own data.

Chapter 3

Creation and persistence of ties in cluster knowledge networks

This chapter is based on a paper co-authored with Balázs Lengyel. The PhD candidate is the first author of the article, which has been published in 2018 as 'Creation and persistence of ties in cluster knowledge networks. Journal of Economic Geography, 18 (6), 1203-1226'.

3.1 Introduction

The idea that knowledge is not in the air available for everyone in industry specializations as opposed to what Marshall (1920) suggested has brought social networks into the forefront of cluster research (Breschi and Lissoni, 2009; Cantner and Graf, 2006; Cooke, 2002; Dahl and Pedersen, 2003; Fornahl and Brenner, 2003; Giuliani and Bell, 2005; Gordon and McCann, 2000; Ethridge et al., 2016; Sorenson, 2003). Despite distant ties might provide the region with new knowledge, most of the learning processes occur within certain spatial proximity (Bathelt et al., 2004; Glückler, 2007). Social ties are important for local knowledge flows because personal acquaintance reduces transaction costs between co-located actors and enhances the efficiency of mutual learning (Borgatti et al., 2009; Maskell and Malmberg, 1999). Knowledge networks that link "[…] *firms through the transfer of innovation-related knowledge, aimed at the solution of complex technical problems*" (Giuliani, 2010, p. 265) have been found very useful empirical tools in providing novel understanding of learning in clusters (Boschma and Ter Wal, 2007; Giuliani, 2007; 2010; Giuliani and Bell, 2005; Morrison and Rabellotti, 2009).

Scholars argue that the evolution of knowledge networks is closely related to the evolution of the cluster itself and therefore we can get new insights into cluster development by analyzing the dynamics of the underlying knowledge networks (Boschma and Fornahl, 2011; Glückler, 2007; Iammarino and McCann, 2006; Menzel and Fornahl, 2010; Martin and Sunley, 2011; Caniels and Romijn, 2003; Staber, 2011; Li et al., 2012; Ter Wal and Boschma, 2011). Path-dependent trajectories are claimed to characterize knowledge network change in clusters because tie selection, being an evolutionary process, is strongly influenced by the previous structure of the network, which is termed network retention (Glückler, 2007). On top of this process, technological or cognitive proximity in clusters is thought to further contribute to the establishment of ties and drive the network towards lock-in (Boschma and Frenken, 2010; Ter Wal and Boschma, 2011). Empirical evidence supports these theories by illustrating that endogenous network effects – such as triadic closure, reciprocity and status – influence tie selection and drives cohesive formulation of cluster knowledge networks (Giuliani, 2013) and that technological proximity further increases the probability of ties (Balland et al., 2016).

Notwithstanding the tendency towards cohesive formulation of social and collaboration networks (Powell et al., 2005), Glückler (2007) also emphasizes that variation of local networks is another major evolutionary process that characterizes path destructive development of regions. He claims that novelty not only arrives from extra-regional ties but can be generated by bridging and brokering loosely connected parts of the local network (Burt, 2004; Granovetter, 1973; Rosenkopf and Padula, 2008). Based on the arguments of Glückler (2007), we propose that both retention and variation in cluster knowledge networks should be

analyzed on the level of tie selection. Therefore, we need a theoretical framework, in which micro-motivations of tie selection helps the reproduction of the network and suggests simultaneous variation, which countervail against an existing trajectory. We argue that such a framework will help us understand how forces of retention and variation jointly drive the dynamics of cluster knowledge networks.

In this chapter, we enter the above discussion by claiming that tie creation and tie persistence in cluster knowledge networks have to be analyzed separately. The distinction is important because the micro-motivations of creating and maintaining ties might involve different costs and constraints (Jackson, 2008), different level of variety and in-depth learning (Rivera et al., 2010), as well as uncertainty (Dahlander and McFarland, 2013). In the next chapter, we argue that the firm commits itself easier to an existing tie with high opportunity costs if the knowledge of the source firm is highly applicable (Cohen and Levinthal, 1990). On the contrary, the firm is more likely to establish a new tie when the search costs and additional uncertainties of the new contact are relatively low (Dahlander and McFarland, 2013). Even though related costs, uncertainties and the value of knowledge access are hardly observable, micro motivations of firms can be inferred on by looking at the effect of network cohesion and proximity variables on the probability of ties (Giuliani 2013; Balland et al. 2016).

We decompose the hypotheses taken from the existing literature (Balland et al., 2016; Giuliani, 2013) into propositions to analyze the effect of triadic closure, geographical and cognitive proximities. This is straightforward because network cohesion and proximities can be argued to be univocal forces of retention and lock-in only in the case these support both tie creation and persistence. However, deviation from this pattern can provide us with new insights into the micro-motivations of network dynamics and also help us to include variation into the discussion. Further, the effects of triadic closure and cognitive proximity are not independent from each other and thus, we look at their joint effect, which allows us to make new conclusions on how the interplay between cohesion and cognitive proximity drives retention and variation in cluster knowledge networks.

Our empirical network data was collected by face-to-face interviews in the printing and paper product cluster in a Hungarian town in years 2012 and 2015. This network fits well to our aims because the cluster is in the mature phase and has a long history in the region; there is a variety of cognitive proximity across firms; and the majority of the local companies apply some kind of specialized technology to create unique paper products.

Applying stochastic actor-oriented models, we find that triadic closure and geographical proximity increase the probability of tie creation but does not influence tie persistence. These findings suggest that proximity in the network and in space lower the costs and uncertainties of the firm when it searches for new connections, but does not influence cohesive network formation and consequently is not a clear engine of retention. Further, cognitive proximity is positively correlated to the probability of tie persistence but firms create ties to cognitively proximate firms only if they do not share partners. This result implies that firms repeat contact and strengthen ties to those partners that have similar technological profile and thus can offer

more applicable knowledge. The last finding also anticipates that variation might indeed counter-act cohesive formation of cluster knowledge networks.

3.2 Literature and framework

3.2.1 Knowledge networks and cluster evolution

Social networks that span across company borders facilitate knowledge flows between firms and therefore have become a cornerstone for understanding why firms in clusters outperform firms outside clusters (Giuliani and Bell, 2005; Gordon and McCann, 2000; Sorenson, 2003). Geographical proximity is crucial for such binds between firms because it creates opportunities for face-to-face and frequent interactions, and by increasing the socializing potential facilitates trust-based social relationships (Storper and Venables, 2004). Such processes lead to the emergence of coherent local collectives, shared rules and norms, and consequently, to more effective local learning, while new impulses can be primarily accessed through extra-regional links (Amin, 2000; Asheim, 1996; Bathelt et al., 2004; Malmberg, 1997). Empirical findings support this view by showing that central firms in the knowledge network are more innovative than firms in the periphery (Boschma and Ter Wal, 2007), by illustrating that the extent of extra-regional links are associated with better performance (Fitjar and Rodriguez-Pose, 2011; Morrison and Rabellotti, 2009), and by showing that the density of individual ties between co-located firms fosters productivity growth in the region (Lengyel and Eriksson, 2017).

However, scholars also warn us that too coherent ecosystems and social environments make renewal difficult (Uzzi, 1997) and can govern regions into locked-in development paths (Grabher, 1993). This is at least partly because social tie formation in clusters is path-dependent and depends on the structure of the network itself (Glückler, 2007). Two of the well-documented phenomena of network evolution apply to cluster networks as well: central firms are likely to become more central (Barabási and Albert, 1999) and alters tend to requite ties or close triangles in the network (Granovetter 1985; Watts and Strogatz, 1998). Another source of path-dependency is driven by similarity-effect between co-located agents. Because similarity increases the likelihood of tie formation, which is often referred to as homophily in social sciences (McPherson et al., 2001), the high level of cognitive proximity between cluster firms breed cohesive tie formation and lock-in (Boschma and Frenken, 2010; Cantner and Graf, 2006).

To explain the changes in the knowledge network over time, Ter Wal and Boschma (2011) propose a macro perspective. They argue that a stable centre-periphery structure emerges over the growth stage of the cluster life cycle and the network becomes dense and cohesive only in the mature phase. An alternative micro perspective was suggested by

Glückler (2007) who claimed that partner selection prevail at the firm level and therefore the micro-foundations of tie creation are fundamental for understanding the evolutionary mechanisms of network change. He further argues that besides the retention mechanisms that cause path-dependence in local networks by the reinforcement of existing network structure, network variation appears as a set of mechanisms that enables the emergence of novelty and path-disruption.

Due to recent methodological developments, ideas connected to the micro-perspective of social network evolution in clusters became empirically testable (Snijders et al., 2010); however, only very few papers do such analyses. Giuliani (2013) pioneers this field and establishes a framework in which the macro outcomes of social network change are explained by its' micro-foundations. She points out that retention-driven endogenous network effects, such as cohesion and status, together with exogenous effects, such as firm level capability, establish a stable hierarchy in the cluster knowledge network in a way that firms with low absorptive capabilities hold back endogenous network effects from driving the network into absolute cohesion. Balland et al. (2016) contributes by comparing the endogenous network effects and exogenous proximity effects across business networks and technological advice networks and find that proximity effects only prevail in technological networks but network effects drive the dynamics of both technological and business networks.

In the next two sub-sections, we provide a new micro approach for cluster knowledge network evolution. Like previous studies, our framework contains selection and retention; however, we stretch the argument further to capture forces of variation as well (Glückler, 2007). In doing this, we separate tie persistence and tie creation and expose that endogenous network effects and proximity effects are not independent from each other. We posit propositions instead of hypotheses, which is a way to stress that, despite the theoretical argument regarding the micro-motivations of network evolution, the framework remains empirical (Uzzi, 1997) and the generality of the results regarding knowledge network evolution in clusters requires further investigations.

3.2.2 Tie creation and tie persistence

To find solution for a technical problem, the firm can either maintain ties by asking advice from existing contacts or can search for and create new ties. Both maintaining ties and searching for new partners demands direct costs – these might include time demand, need for financial resources, cognitive effort, or social constraint –, and the opportunity costs of allocating resources to the specific tie instead of other ties (Glückler, 2007; Hansen, 1999). Asking advice from existing contacts needs shared time and commitment and strengthening the connection thus is thought to involve large opportunity costs (Coleman, 1988; Uzzi, 1997); while asking a new partner demands some but arguably less effort (Burt, 2004; Granovetter, 1973).

There are further qualitative differences between creating a new tie or maintaining an existing one, for which one can apply the exploration versus exploitation dichotomy (Beckman et al. 2004; Levinthal and March, 1993; March 1991; Verspagen and Duysters 2004). On the one hand, exploring a new knowledge source offers opportunities for firms in clusters to find new variety of knowledge (Hansen, 1999; Reagans and McEvily, 2003) but involves uncertainties as well because no prior experience exists about the new partner (Dahlander and McFarland, 2013; Lavie and Rosenkopf, 2006). On the other hand, maintaining and thus strengthening the connection can ease the transfer of complex or tacit knowledge (Aral, 2016; Reagans and McEvily, 2003) and uncertainties are less profound when exploiting the link to an existing partner (Greve et al., 2010; Hanaki et al., 2007). Inter-organizational ties dissolve if the firm finds alternative ties that offer better and still affordable solutions and persist only if the tie represents a valuable connection (Seabright et al., 1992).

We argue on the base of this literature that the firm will select to maintain those existing ties with higher probability that provide access to relatively high value-cost ratio, because these ties might provide more relevant and more applicable knowledge than other existing ties. On the contrary, the firm is more likely to establish a tie when the search costs and additional uncertainties of the new contact are low compared to other possible new contacts. Because these theoretical micro-motivations of network dynamics are non-observable, one can derive them from the effect of observable factors, such as endogenous network effects, geographical and cognitive proximities.

Endogenous network effects – such as cohesion – decrease costs of new tie creation because a shared contact might help to establish the new connection and can also diminish uncertainties by providing information about potential partners (Granovetter, 1985). However, cohesion also increases the likelihood that new connections will give access to redundant knowledge (Hansen, 1999) and therefore too much cohesion harms variation in the network and, after a certain threshold, the performance of firms and the network itself (Aral and van Alstyne, 2011; Uzzi, 1997; Uzzi and Spiro, 2005). On the contrary, it is not clear how cohesion influences the costs of tie persistence. Strong and cohesive ties increase the willingness of the knowledge source to share complex knowledge and therefore decrease the relative costs of repeated communication (Reagans and McEvily, 2003) but cohesive ties also demand more time and commitment (Granovetter, 1973) and thus their maintenance can also extensively increase the opportunity costs of the tie (Glückler, 2007).

We opt for triadic closure as a measure of network cohesion and test how it influences tie creation and tie persistence. Previous results are mixed; Giuliani (2013) found that triadic closure had a positive effect on the probability of tie presence in cluster knowledge networks; while Shipilov et al. (2006) found that triadic closure only influences tie creation positively and has no significant effect on tie persistence. Staber (2011) also finds those ties are less durable that are brokered through a third party. However, according to the central tenet, cohesion is a main factor of network retention. Therefore, we propose positive correlation for both mechanisms and test these effects empirically.

Proposition 1A: Triadic closure is positively correlated to the probability of tie creation.
Proposition 1B: Triadic closure is positively correlated to the probability of tie persistence.

In case both propositions are supported, we can argue that cohesion leads to network retention by reducing costs and uncertainties of searching for new partners and by facilitating complex knowledge sharing. However, failure of either Proposition 1A or 1B would imply that cohesion is not absolute and firm motivations induce network variation as well.

Geographical proximity is thought to increase the opportunity to meet and formulate new relationships (Borgatti et al., 2009; Marmaros and Sacerdote, 2006; Rivera et al., 2010; Storper and Venables, 2004) and also to maintain contacts (Lambiotte et al., 2008; Lengyel et al., 2015) primarily through decreasing travel and transportation costs. However, geographical proximity also offers potentials to form weak ties (Wellman, 1996) and scholars argue that other types of proximities are more important to establish strong connections when geographical proximity is given (Boschma, 2005; McPherson et al., 2001). The physical closeness of actors decreases the costs of setting up a new relationship and also moderates the costs of repeating interactions. Therefore, we propose positive correlation for both tie creation and tie persistence.

Proposition 2A: Geographical proximity is positively correlated to the probability of tie creation.
Proposition 2B: Geographical proximity is positively correlated to the probability of tie persistence.

If both of these propositions are supported, one can argue that knowledge ties are concentrating in space because geographical proximity facilitates tie formation by decreasing costs of meeting new partners and of repetitive face-to-face interactions. However, the failure of either Proposition 2A or 2B would suggest that place-dependency does not dominate network evolution, and further micro-motivations might be more important for tie selection.

Cognitive proximity influences the dynamics of cluster knowledge networks (Balland et al., 2016; Boschma and Frenken, 2010) and evidence shows that similarity in knowledge increases the probability of interaction in groups (Carley, 1991; Galaskiewicz and Shatin, 1981). However, it is still not entirely clear how cognitive proximity influences tie creation and tie persistence separately. One might expect that cognitive proximity facilitates tie creation because it decreases the level of uncertainty related to new partners and thus the firm can expect accurate and useful advice from those partners that can understand the technical problem the firm faces (Cohen and Levinthal, 1990; Lane and Lubatkin, 1998; Nelson and Winter, 1982). However, the probability of finding redundant knowledge rises with cognitive proximity because there is an overlap in the knowledge bases of firms (Boschma, 2005). Furthermore, similarity of knowledge bases can lower the rising costs of repeated knowledge

transfer and thus it is easier to maintain a tie (Reagans and McEvily, 2003). Nevertheless, the role of cognitive proximity in tie creation and tie persistence needs further understanding and we propose a positive relation for both dynamics and will discuss the influence of cognitive proximity on knowledge network dynamics with the empirical results at hand.

Proposition 3A: Cognitive proximity is positively correlated to the probability of tie creation.
Proposition 3B: Cognitive proximity is positively correlated to the probability of tie persistence.

In case Proposition 3A gets empirical support, we can argue that cognitive proximity is important to establish relationships in clusters because new ties to cognitive proximate actors involve lower uncertainties. Evidence for Proposition 3B would underlie that ties last longer between firms with similar technological profiles because they understand each other, which eases knowledge transfer and might also increase the value of accessible knowledge and consequently, the high opportunity costs of strong ties are compensated. Support for both propositions would suggest that the dynamics of the network drives the cluster towards technological lock-in (Boschma and Frenken, 2010; Cantner and Graf, 2006; Broekel and Boschma, 2012). However, the failure of either proposition would imply that variation and path destruction are also taking place in cluster knowledge networks.

3.2.3 Interplay between network effects and cognitive proximity

Endogenous network effects and proximity effects are not independent from each other in network evolution because link formation induced by similarity usually establishes cohesive groups of similar individuals, which is commonly referred to as homophily in the sociology literature (McPherson et al., 2001). In turn, studies that focus on the origins of homophily claim that the high levels of homophily observed in social networks are to a large extent due to structural properties of the network, such as triadic closure and reciprocity, which further induce connections between similar individuals (Kossinets and Watts, 2009; Wimmer and Lewis, 2010). This issue tells us that it is difficult to disentangle cohesion effects and effects of cognitive proximity in knowledge network evolution. Admitting that this chapter cannot solve the problem, we aim to make a step towards understanding whether endogenous network effects and cognitive proximity strengthen or weaken each other in driving the dynamics of cluster knowledge networks.

It is difficult to overstate the importance of this effort for economic geography. Because proximity in too many dimensions of knowledge relations harm renewal capacities of regions (Grabher, 1993), "*[…] solution to such regional lock-in phenomena clearly lies in trying to re-organize the network relations such that interactions can take place between actors that are less proximate […]*" (Boschma and Frenken, 2010, p. 130-131). However, it is still unclear how network variation happens while network retention is clearly in action (Glückler, 2007). We

argue that the joint effect of endogenous network effects and cognitive proximity on network dynamics can provide us novel insights into the question. This problem has not been studied in economic geography before and there are hardly any empirical studies to base our expectations upon. An exception is Rosenkopf and Padula (2008) who find that similarity – in their case structural homophily that captures similarity in status instead of knowledge base – predicts tie formation between loosely connected parts of networks, but does not predict tie formation in cohesive sub-networks. Their results imply that network variation is only possible if network endogeneity and homophily weaken each others' effect.

One can look at the joint effect of dyadic network variables by using their interaction (Powell et al. 2005); in this chapter, we opt for the interaction between triadic closure and cognitive proximity. We borrow the argument of Rosenkopf and Padula (2008) to formulate our expectation and posit that ties are less likely to form and persist between cognitively proximate potential partners in the cluster if they also share contacts.

Proposition 4A*: The interaction of triadic closure and cognitive proximity is negatively correlated to the probability of tie creation.*
Proposition 4B*: The interaction of triadic closure and cognitive proximity is negatively correlated to the probability of tie persistence.*

In case Propositions 4A and 4B are verified, we could argue that sharing partners simplifies the creation and maintenance of connections to cognitively distant peers by reducing the uncertainty whether it is worth to establish the new knowledge access or not and by reducing the costs of repeated knowledge transfer. Alternatively, such findings could also suggest that the firm is more likely to reach out and keep relation to those partners with similar and easy to apply knowledge if they do not share partners because the likelihood of finding novelty is higher (Boschma, 2005; Granovetter, 1973; Hansen, 1999). In sum, verification of these propositions would provide new evidence that network endogeneity and network variation are simultaneously present in cluster evolution and are driven by the interplay between cohesion and cognitive proximity.

However, Huber (2012) does not find such clear negative relation between social proximity and cognitive proximity when looking at the importance of knowledge ties in the Cambridge IT cluster; whereas cognitive proximity and social proximity have not been found to co-evolve in the German R&D collaboration network either (Broekel, 2015). Therefore, we choose to keep the empirical nature of our expectation and discuss potential implications for cluster evolution with the research results at hand.

3.3 The study setting

3.3.1 Printing and paper product industry in Kecskemét

Printing and paper product industry has a long tradition in the region of Kecskemét[4]. The town is about 80 km south from Budapest, the capital of Hungary, and accounts for around 115.000 inhabitants with an economy rooted in agriculture as well as processing and manufacturing industries (heavy machinery and car manufacturing). The first printing-house called Petőfi Press was established in the 1840s and it still works under this name. Since the 1990s, after the planned economy collapsed in Hungary, numerous small and medium enterprises (SMEs) were born and created a strong local base for the industry. International companies have also located their facilities in the town (e.g. Axel-Springer). By now, the location quotient calculated from the number of employees shows significant relative concentration of both the manufacture of articles of paper and paperboard (LQ=4.602) and the printing and service activities related to printing (LQ=1.059).

The relatively high concentration and simultaneous presence of small and big firms resulted in intensive local competition, which requires flexible specialization of SMEs and the local industry as such. Almost all of the present companies apply some kind of specialized technology to create unique paper products (e.g. specifically printed, folded, unique paper products, packaging materials, stickers and labels). Firms typically deal with customized traditional goods or services, do not carry out R&D activities, the cluster is built around mature technological knowledge and smaller, customer-driven process-oriented innovations are typical in order to satisfy the customers' unique needs. In sum, the local industry can be characterized as an old social network-based cluster (Iammarino and McCann, 2006) and it provides appropriate conditions for analyzing the dynamics of the knowledge network. Firstly, as we discovered along the first round of interviews in 2012, there is a strong local network behind the clusters which is characterized by informal networking processes and based on the interactions of technicians to search for advice on technical problems that cannot be solved in house. For example, they may want advice on how to set a new type of printing machine or ask for experience with a special type of packaging carton. Secondly, the cluster is in a mature life-cycle stage as the number of firms is relatively stable and there are no external effects that might influence networking processes and we should control for.

[4] For a visual presentation of the location of Kecskemét in Hungary and the location of firms around the town see section 3.A in the Appendix.

3.3.2 Data collection and management

For the selection of the particular firms we used The Company Code Register (2011) by the Hungarian Central Statistical Office, which is a nation-wide firm level dataset with seat addresses, classification of economic activities and basic firm statistics. We chose all firms that had at least 2 employees, had the company seat in the urban agglomeration of Kecskemét and were classified under the industry code 17 (Manufacture of paper and paper products) or 18 (Printing and reproduction of printed media) in the Statistical Classification of Economic Activities of Eurostat (2008). Based on 2012 data, 38 firms suited the above conditions and we merged those firms that had identical addresses and similar names, which resulted in a final number of 35 firms.

We collected data by face-to-face structured interviews with skilled workers (mostly with co-founders, operational managers or foremen). The relational data was collected through the so called "roster recall" method (Wasserman and Faust, 1994); each firm was asked to report relations to any other cluster firms presented to them in a complete list (roster). The question formulated to collect knowledge network data was exactly the same as used in several studies before (Giuliani and Bell, 2005; Morrison and Rabellotti, 2009). This question is related to the transfer of innovation-related knowledge and only reveals the inter-firm linkages that are internal to the cluster and specifically address problem solving and technical assistance (Giuliani and Bell, 2005). This is meant to capture not only the bare transfer of information, but the transfer of contextualized complex knowledge instead. In our setting, revealed relationships are trust-based, informal connections that are vulnerable to the loss of confidence. We collected additional year-specific firm-level information about main activities, number of employees, type of ownership and external knowledge linkages of firms. We also used an open question to explore other important actors for knowledge sharing not mentioned in the roster.

We managed to get answers from 26 different companies in year 2012 and repeated the interviews in 2015 with the same firms. Compared to previous studies on cluster knowledge network evolution (Giuliani, 2013; Balland et al., 2016) we take a mid-time interval of three years to indicate significant changes in network relations. Burt (2000) suggests that non-repeated contacts vanish after three years. Although two companies were closed down during the years, other two were mentioned by the respondents in the open questions at the end of the roster. Thus, we collected 26 responses in year 2015 too and reached more than 70% of the local firms in the industry at both time points. The data gathering could be judged as a success as only one firm refused to answer our questions in 2012. Most of the non-responding actors were shut down or temporarily stopped their business activities and all of them were domestic small and medium-sized enterprises (SMEs).

The questions related to firms' knowledge transfers have been used to construct two directed adjacency matrices with n x n cells (where n stands for the number of respondents)

for the two time points, in which each cell reports on the existence of knowledge being transferred from firm i in the row to firm j in the column. The cell (i, j) contains the value of 1 if firm i has transferred knowledge to firm j and contains the value of 0 when no transfer of knowledge has been reported between firm i and j.

3.3.3 Descriptive analysis

The main characteristics of the examined firms did not change from 2012 to 2015[5]. Most of them are SMEs, there is only one firm with more than 100 employees and only a minority of them is foreign-owned (less than 20%). Two companies were closed down along the studied period, but two other companies joined to the sample by 2015. As we can clearly see in Table 3.1, the knowledge network became sparser over time. From the 223 knowledge ties apparent in 2012 only 110 linkages persisted. Interestingly, no firms became isolated by 2015. On average, actors asked for technical advice from 8 firms in 2012 and only from 6 firms in 2015. We used the Jaccard index to measure the stability of the network, which is higher than 0,3 and within appropriate limits for the analysis of network evolution (Ripley et al., 2017). The visual representation of the knowledge networks (Figure 3.1) suggests that the degree distribution is not proportional. In both cases the network is hierarchical and some actors have remarkably more connections than others. This is in line with previous studies that have shown the uneven and hierarchical nature of knowledge exchange in clusters (Giuliani, 2007).

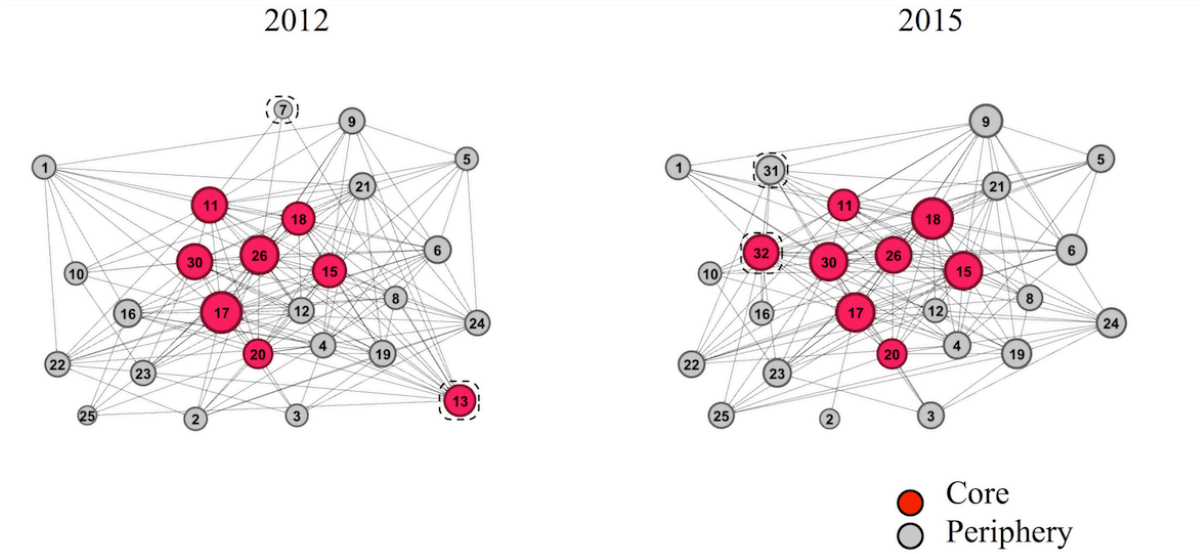
Table 3.1 Descriptive statistics of the knowledge network in 2012 and 2015

	2012	2015
Nodes	26	26
Ties	223	181
Density	0.295	0.239
Average degree	7.964	6.464
Ties created	-	71
Ties persisted	-	110
Ties dissolved	-	113
Isolates	0	0
Jaccard index	-	0.374

Source: Author's own data.

[5] Detailed descriptive statistics of the sample firms are provided in the Appendix (section 3.B).

Figure 3.1 The local knowledge network of the printing and paper product industry in Kecskemét in 2012 and 2015



Source: Author's own data.
Note: The size of the nodes is proportional to degree. Firms who left the 2012 sample or entered the 2015 sample are marked by dashed frame.

The high number of tie dissolution and the unstable nature of the core-periphery structure suggest that neither the network nor the cluster is in a growing stage (Ter Wal and Boschma, 2011)[6]. In line with that, the personal interviews in 2015 confirmed that the local competition had intensified. Some of the central firms in the 2012 knowledge network revealed that they do not share or dare to contact other firms for technical advice because they fear their market share, reputation, and know-how. These descriptive findings imply that the cluster under study is in the phase of its' life-cycle when increasing competition could cause secrecy in clusters as firms keep their technical solutions for themselves and tend to share less knowledge (Menzel and Fornhal, 2010) and not in the phase when competition stimulates firms to innovate as idealized by Porter (1990).

3.4 Methodology and variables

Similarly to previous papers on knowledge network evolution (Balland, 2012; Giuliani, 2013; Balland et al., 2013; Ter Wal, 2014; Balland et al., 2016), we apply stochastic actor-oriented models (SAOMs). These models can take account of three classes of effects that influence the evolution of networks (Ripley et al., 2017; Snijders et al., 2010). Firstly, endogenous or structural effects that come from the network structure itself (e.g. degree-related effects,

[6] As shown in detail in section 3.C in the Appendix, we find that both the composition and the density of linkages changed in the core of the cluster knowledge network.

triadic closure, reciprocity). Secondly, dyadic covariate effects e.g. similarity or proximity (commonly referred to as homophily or assortativity) between pair of actors. Thirdly, individual characteristics of actors are also taken into account because the ego-effect expresses the tendency of a given characteristic to influence the network position of the node. Further, SAOM estimations rely on three basic principles (Snijders et al., 2010). First, the evolution of the network structure is modeled as the realization of a Markov process, where the current state of the network determines its further change probabilistically. Second, the underlying time parameter *t* is continuous, which means that the observed change is the result of an unobserved series of micro steps and actors can only change one tie variable at each step. Third, the model is 'actor-oriented' as actors control and change their outgoing ties on the basis of their positions and their preferences.

In SAOMs, actors drive the change of the network because at stochastically determined moments they change their linkages with other actors by deciding to create, maintain or dissolve ties. Formally, a rate function is used to determine the opportunities of relational change, which is based on a Poisson process with rate λ_i for each actor *i*. As actor *i* has the opportunity to change a linkage, its choice is to change one of the tie variables x_{ij}, which will lead to a new state as $x, x \in C(x^0)$. Choice probabilities (direction of changes) are modeled by a multinomial logistic regression, specified by an objective function f_i (Snijders et al., 2010):

$$P\{X(t) \text{ change to } x | i \text{ has a change opportunity at time } t, X(t) - x^0\}$$
$$= p_i\left(x^0, x, v, w\right) = \frac{\exp\left(f_i\left(x^0, x, v, w\right)\right)}{\sum_{x' \in C\left(x^0\right)} \exp\left(f_i\left(x^0, x', v, w\right)\right)} \tag{3.1}$$

When actors have the opportunity to change their relations, they choose their partners by maximizing their objective function f_i (Broekel et al., 2014; Balland et al., 2013). This objective function describes preferences and constraints of actors. Choices of collaboration are determined by a linear combination of effects, depending on the current state (x^0), the potential new state *(x),* individual characteristics *(v),* and attributes at a dyadic level *(w)* such as proximities. Therefore, changes in network linkages are modeled by a utility function at node level, which is the driving force of network dynamics.

$$f^i\left(x^0, x, v, w\right) = \sum_k \beta_k s_{ki}\left(x^0, x, v, w\right) \tag{3.2}$$

The estimation of the different parameters β_k of the objective function is achieved by the mean of an iterative Markov chain Monte Carlo algorithm based on the method of moments, as proposed by Snijders (2001). This stochastic approximation algorithm estimates the β_k parameters that minimize the difference between observed and simulated networks. Along the iteration process, the provisional parameters of the probability model are progressively

adjusted in a way that the simulated networks fit the observed networks. The parameter is then held constant to its final value, in order to evaluate the goodness of fit of the model and the standard errors. For a deeper understanding of SAOMs see Snijders et al. (2010) and for an economic geography review see Broekel et al. (2014).

Table 3.2 demonstrates three different specifications of SAOMs (Ripley et al., 2017). Evaluation function compares the probability of presence to the absence of the tie at time *t+1* regardless of tie status at *t*. Creation function compares the probability of creating a previously not existing tie to not creating a tie; while the endowment function compares the probability of tie persistence to tie termination. These three specifications represent three different dependent variables of network evolution. Previous studies only looked at the evaluation models (Balland et al., 2016; Giuliani, 2013) and had to assume that the odds ratios in the creation and endowment models are identical (Ripley et al., 2017). However, these probability ratios typically differ, which is the case in our empirical sample as well. The differentiation between dependent variables in SAOMs is rarely applied (Cheadle et al., 2013) and empirical studies based on this distinction are completely missing from the economic geography literature.

Table 3.2 Tie changes considered by the evaluation, creation and endowment functions

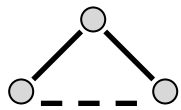
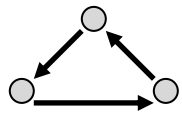
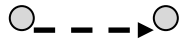
	Evaluation model		
	t	t+1	Number of ties
Creation	i j	i –> j	71
Persistence	i –> j	i –> j	110
Termination	i –> j	i j	113
No ties	i j	i j	462
Odds ratio			181/575
	Creation model		
	t	t+1	Number of ties
Creation	i j	i –> j	71
Persistence			
Termination			
No ties	i j	i j	462
Odds ratio			71/462
	Endowment model		
	t	t+1	Number of ties
Creation			
Persistence	i –> j	i –> j	110
Termination	i –> j	i j	113
No ties			
Odds ratio			110/113

Source: Author's own construction based on Ripley et al. (2017).

The effects of structural, dyadic, and individual variables are estimated in order to test the propositions; these variables are described in Table 3.3. To investigate how structural

effects or network cohesion shape the evolution of the knowledge network behind the examined cluster we investigate the role of triadic closure that is often used in SAOM papers and captures the notion when partner of partners become partners so that a triad is created (Giuliani, 2013; Balland et al., 2016). In order to control for other endogenous network effects, like other studies do, we include density (out-degree of actors), reciprocity and directed cycles (3-cycles).

Table 3.3 Operationalization of structural, dyadic and firm level variables

Structural variables			
	Description	Formula	Visualization
Triadic closure	Tendency toward triadic closure when two knowledge ties existed in the previous period	$T_i = \sum_{j,h} x_{ij} x_{ih} x_{jh}$	
Reciprocity	Tendency of mutual knowledge exchange	$R_i = \sum_j x_{ij} x_{ji}$	
Cyclicity	Tendency of knowledge exchange in cycles	$C_i = \sum_{j,h} x_{ij} x_{jh} x_{hi}$	
Density	Overall tendency of actors to ask advices	$D_i = \sum_j x_{ij}$	
Dyadic variables			
Geographical proximity	Physical distance of firms subtracted from the maximum distance in the sample		
Cognitive proximity	Number of digits two firms share in common in their NACE 4 codes		
Triadic closure X Cognitive proximity	Number of common third partners multiplied by the number of digits two firms share in common in their NACE 4 codes		
Firm level variables			
External knowledge ties	Number of knowledge linkages outside the region		
Age	Number of years since establishment		
Ownership	Equals *1* if foreign and *0* if domestic		
Employment	Total number of employees		

Source: Author's own construction based on Balland et al. (2016), Giuliani (2013), Snijders et al. (2010)
Note: The plain lines and arrows represent pre-existing ties, while the dashed arrows represent the expected ties that will be created if the corresponding structural effect is positive.

To capture the importance of dyadic effects on knowledge network tie formation, we focus on geographical proximity, cognitive proximity and the interaction of possible triads and cognitive proximity. Proximities are frequently used as dyadic effects in SAOM based knowledge network studies (Balland, 2012; Balland et al., 2013; Balland et al., 2016; Ter Wal, 2014). Geographical proximity is operationalized as the distance of the selected pair of firms subtracted from the maximum of the physical distance of firms. The variable takes higher

value as the distance between firms diminishes. We applied a valued measure for cognitive proximity corresponding to the number of digits the two firms have in common in their NACE 4 codes (Balland et al., 2016)7. This measure assumes that two firms have similar technological profiles and therefore are in cognitive proximity if they operate at the same sector category (Frenken et al., 2007). To control for the independence of network structural effects and actor similarity on tie creation and persistence, we also investigate the interaction variable of the number of common third partners and cognitive proximity on dyadic level.

The importance of external relationships has been highlighted in the cluster literature (Bathelt et al., 2004; Glückler, 2007; Morrison, 2008). To measure the effect of extra-regional connections as an individual characteristic, we used the number of external knowledge ties (mean it links to other regions in Hungary or abroad). Additionally, we used actor related control variables as type of ownership, age, and the number of employees.

Since our networks are directed, we can control for the effect of individual characteristics on incoming and outgoing ties (Ripley et al., 2017). Alter variables represent the effect of individual characteristics on the actor's popularity to other actors. A positive parameter will imply the tendency that the in-degrees of actors with higher values on this variable increase more rapidly. Ego variables represent the effect of individual characteristics on the actor's activity. A positive parameter will imply the tendency that actors with higher values on this variable increase their out-degree more rapidly. The differentiation is important in case of cluster knowledge networks as the motives behind knowledge sharing and knowledge exploration could be highly influenced by the characteristics and capabilities of firms.

3.5 Results

Table 3.4 presents the results of six SAOM specifications. Model (1) represents the general model while Model (2) contains the interaction of triadic closure and cognitive proximity as well. For both model settings we first estimate every effect by evaluation function, then we split our models by the applied creation and endowment functions on our four main variables (as indexed in Table 3.4). We opt to only change the underlying functions of triadic closure, geographical proximity, cognitive proximity and the interaction effect, while every other parameter is estimated by evaluation function only[8]. All parameter estimations in all models are based on 2000 simulation runs in 4 sub-phases. Parameter estimates can be interpreted as log-odds ratios, appropriate to how the log-odds of tie formation change with one unit

[7] More details and descriptive statistics of our cognitive proximity measure can be seen in the Appendix (section 3.D).

[8] The right way to split models to estimate creation and endowment in RSiena based SAOMs is still highly debated. Based on the instructions of Ripley et al., 2017, we opt to add the effects in question in either the creation or the endowment role into the same model. However, all along the many different model settings we tested, the sign, size and significance of our main explanatory variables were stable and our findings are robust.

change in the corresponding independent variable (Balland et al., 2016) because they are non-standardized coefficients from a logistic regression analysis (Steglich et al., 2010; Snijders et al., 2010). Since the null hypothesis is that the parameter is 0, statistical significance can be tested by t-statistics assuming normal distribution of the variable. The convergence of the approximation algorithms is sufficient for each model because all t-ratios are smaller than 0.1.

Table 3.4 Dynamics of the knowledge network

	Evaluation		Creation		Endowment	
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
Triadic closure	0.191***	0.218***	0.422***	0.462***	-0.103	0.003
	(0.032)	(0.036)	(0.059)	(0.072)	(0.108)	(0.131)
Geographical proximity	0.031	0.043	0.172*	0.259**	-0.103	-0.086
	(0.041)	(0.043)	(0.104)	(0.123)	(0.081)	(0.078)
Cognitive proximity	0.111**	0.276***	0.062	0.361***	0.194**	0.448***
	(0.050)	(0.077)	(0.092)	(0.137)	(0.083)	(0.146)
Triadic closure X Cognitive proximity		-0.049***		-0.141***		-0.055**
		(0.017)		(0.044)		(0.027)
External knowledge ties alter (evaluation)	-0.014	-0.015	-0.019	-0.027	-0.017	-0.015
	(0.017)	(0.016)	(0.017)	(0.018)	(0.018)	(0.018)
External knowledge ties ego (evaluation)	0.070***	0.081**	0.054***	0.068***	0.142***	0.128***
	(0.025)	(0.032)	(0.020)	(0.026)	(0.041)	(0.037)
Age alter (evaluation)	0.009	0.017	0.012	0.026*	0.021*	0.024**
	(0.011)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)
Age ego (evaluation)	-0.020*	-0.014	-0.015	-0.003	-0.037**	-0.028*
	(0.012)	(0.013)	(0.011)	(0.013)	(0.017)	(0.016)
Employment alter (evaluation)	0.001	0.001	0.000	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Employment ego (evaluation)	-0.000	-0.000	-0.000	0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Ownership similarity (evaluation)	0.048	0.111	0.176	0.255	0.098	0.126
	(0.186)	(0.207)	(0.208)	(0.216)	(0.198)	(0.200)
Cyclicity (evaluation)	-0.180***	-0.179***	-0.218***	-0.203***	0.183*	0.136
	(0.061)	(0.068)	(0.070)	(0.076)	(0.098)	(0.110)
Reciprocity (evaluation)	0.752***	0.701***	1.005***	0.989***	0.723***	0.641***
	(0.222)	(0.238)	(0.285)	(0.285)	(0.202)	(0.226)
Density (evaluation)	-1.600***	-1.778***	-1.959***	-2.265***	-1.265***	-1.347***
	(0.184)	(0.208)	(0.235)	(0.307)	(0.191)	(0.185)
Rate parameter (rate)	12.282	11.215	15.249	13.242	10.216	10.532
	(1.312)	(1.139)	(2.025)	(1.559)	(1.027)	(1.030)
Iteration steps	3898	4194	4141	4194	4191	4194
Convergence t-ratios	< 0.07	< 0.03	< 0.03	< 0.05	< 0.04	< 0.07

Source: Author's own data.

Note: Results of the stochastic approximation. The convergence of the models was good, as all t-ratios were smaller than <0.1. The coefficients are significant at the * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ level.

The coefficients of triadic closure are positive and significant in the evaluation models, which is in line with previous findings (Balland et al., 2016; Giuliani, 2013). We find that cohesion has a positive and significant effect in the creation models, but has no significant effect in the endowment models. These findings confirm Proposition 1A, but does not support 1B as triadic closure positively influences the probability of new tie creation, but do not influence the probability of tie persistence in the cluster knowledge network. These results suggest that the structure of the network promotes opportunities to establish connections and shared contacts reduce the costs and uncertainties of the search for new partners. However, our findings do

not support the idea that the maintenance of cohesive relationships is a general source of network retention in clusters.

Our second proposition concerns the role of geographical proximity as an influential factor of network dynamics. Unlike in a previous result (Balland et al., 2016), we find that the coefficient of geographical proximity is only significant and positive in creation models but does not influence the dependent variable in the evaluation and endowment models. Therefore, we confirm *Proposition 2A* and dismiss *2B*. This finding underlines the importance of micro-level geography and means that physical proximity provides opportunities for establishing knowledge ties, lowers costs and uncertainties of tie creation, but does not affect the assessment and maintenance of relationships. Consequently, place-dependency is not a general source of network retention. The results are also in line with the literature that questions the sufficiency of geographic proximity for knowledge transfer, learning and innovation and highlights the importance of other proximity dimensions (Boschma, 2005; Boschma and Frenken, 2010).

The third proposition addresses the role of cognitive proximity on tie creation and tie persistence in cluster knowledge networks. Unlike the previous two propositions, results in Model (1) and Model (2) are different. While the coefficients of cognitive proximity are positive and significant in both models for evaluation and endowment, the effect of cognitive proximity on tie creation turns positive and significant only in Model (2). Therefore, we can not accept *Proposition 3A* but can confirm *3B*. These results suggest that firms are more likely to maintain strong ties to partners with similar technological profiles. One can think of various possible implication of this result. Cognitive proximity might help the persistence of ties by reducing the costs of knowledge transfer and therefore enabling the partners to repeat the interaction. In turn, the strong relations that emerge by persisting cognitively proximate ties might foster the transfer of complex knowledge between firms in the cluster.

Finally, our fourth proposition posit that endogenous network effects and cognitive proximity are not independent and therefore we use a dyadic level variable to see how the interaction of the number of common partners and the extent of cognitive proximity affects tie creation and tie persistence. As we proposed, the interaction variable has a negative effect on both creation and persistence of ties. This result confirms both *Proposition 4A* and *4B*. Results in Model (2) suggest that the creation and persistence of a tie between two firms is less likely if they share many common partners and are cognitively proximate at the same time. In this case cognitive proximity in itself also supports tie creation, as firms might expect valuable knowledge from firms with similar technological profile, but they can not get any information about the partner via indirect relations. Cognitive proximity and therefore the value of expected advice seems to be a major force behind tie persistence, however, firms maintain costly strong ties to actors only if they cannot get access to the knowledge indirectly. These results lead to the conclusion that cohesive network effects and the effect of cognitive proximity are not independent and by the analysis of their interplay we can get a much better picture about the evolutionary process of knowledge network formation in clusters. It seems

that previously identified forces of retention counteract each other and rather help actors to vary their relations in order to find new varieties of knowledge.

Additionally, we included structural and firm level control variables in both models. The rate parameter indicates the estimated number of opportunities for change per actor, which refers to the stability of the network over time. The positive and relatively high value suggests that there were significant changes in the formation of new ties. Meanwhile, the negative and highly significant coefficients of density indicate that firms tend not to form and maintain knowledge linkages with just any other firm in the cluster (Snijders et al., 2010; Ripley et al., 2017). Similar coefficients were found for density previously (Balland et al., 2016; Giuliani, 2013). The negative and significant effect of cyclicity in most of our models indicates that actors create their relationships with their partner's partner in a certain hierarchy, but knowledge does not circulate among them. Instead, a dominant actor is more likely to provide it to the other two partners in the triad. However, cyclicity does not a significant affect when we test for the persistence of knowledge ties.

Further, the significance of the number of external knowledge ties as an ego effect control variable suggest that firms that build and maintain more linkages to actors outside the region establish and maintain their local ties more likely. As only the ego effect of external ties proved to be significant, it seems that firms with more external ties mostly establish out-going local linkages, therefore, seek for advice and absorb knowledge from cluster firms but share their own experiences with others to a lesser extent. Findings suggest that external stars, firms that have strong extra-cluster knowledge relations, but weak, absorption oriented intra-cluster linkages (Giuliani and Bell, 2005; Morrison et al., 2013) have significant influence on local tie formation. The role of age is still questionable as it has significant coefficients in model versions with endowment effects but has lower influence on tie formation in any other model versions. The significance of the age alter effect on tie durability suggests that older firms are more likely to give advice and share experience to their regular partners, but ask for technical help less frequently. The size and ownership of firms do not influence their knowledge tie formation.

A variety of robustness checks were carried out in order to confirm the stability of the results[9]. First, we have run both Model (1) and Model (2) stepwise with different combinations of variables. Since the model settings to decompose the evaluation function into creation and endowment functions is still debated, we also tried many other model specifications. Besides the presented models, we estimated every parameter by creation and endowment effects too and also tried the incorporation of both creation and endowment effects into the same model. Results remained the same in every case. We have also tried to include in-degree or network status as a control variable but it had no significant effect on tie formation and led to large *t* values of convergence. Every model has been run with only ego and only alter variables of individual characteristics as well. Along the large variety of different simulation runs, the size,

[9] The correlation tables of all presented SAOMs can be seen in the Appendix (section 3.E).

sign and significance of the estimates of the main explanatory variables were stable. The inclusion of both ego and alter versions of firm level characteristics further improved both our model convergences and interpretation. Second, in order to ensure our results on the different effects of proximities, we also applied Mann-Whitney tests for the distribution of proximity values in case of tie creation and tie persistence.

Table 3.5 Distribution of proximity values in case of tie creation and tie persistence

	Created ties		No ties
Number of ties	71		462
Average geographic proximity	8.676		7.807
Mann-Whitney test (*p*-value)		0.001	
Average cognitive proximity	1.929		1.894
Mann-Whitney test (*p*-value)		0.776	
	Persisted ties		Dissolved ties
Number of ties	110		113
Average geographic proximity	8.336		8.407
Mann-Whitney test (*p*-value)		0.881	
Average cognitive proximity	2.209		1.584
Mann-Whitney test (*p*-value)		0.005	

Source: Author's own data.

In Table 3.5, we compare the distribution of geographical proximity and cognitive proximity between created ties versus lacking ties (as in the creation model), and between persisted ties versus terminated ties (as in the endowment model). The *p*-values suggest that in case of tie creation the value of geographic proximity is significantly higher for created ties than for lacking ties, while the value of cognitive proximity is higher for persisted ties than for dissolved ties. The distribution tests further strengthen the robustness of our SAOM based results.

3.6 Conclusions and discussion

According to the first results of this chapter, triadic closure and geographical proximity increase the probability of tie creation, but do not influence tie persistence. These findings mean that firms select those new partners with higher likelihood that they share third partners with or that are in physical proximity. This suggest that being close in the network and in space creates opportunities for face-to-face meetings and speeds up information flow, and thus lower costs and uncertainties of searching new knowledge ties. However, our results do not support the idea that these ties also persist on the long run and promote retention in the network. Cohesive and geographically proximate ties are equally likely to be terminated than non-cohesive and physically distant relations. A straightforward interpretation of the latter finding is that firms choose to maintain knowledge ties driven by the content of accessible

knowledge and once the tie has been established, network structure and spatial location does not play a primer role.

Indeed, we find that cognitive proximity favours the persistence of ties but a positive and significant effect for tie creation has been found only when we introduced the interaction between triadic closure and cognitive proximities to the model. The first result suggest that a firm is more likely to repeat communication and maintain a knowledge tie with cognitively proximate partners than with cognitively distant peers. Our interpretation is that the value of advice or the applicability of transferred knowledge increases with cognitive proximity and therefore these ties are more valuable for firms. An alternative explanation is that cognitive proximity decreases the costs of knowledge transfer and therefore, firms can repeat interaction to have access to complex knowledge even if the opportunity costs of strong relations are increasing.

The negative and significant co-efficient of the interaction between triadic closure and cognitive proximity has far-reaching implications for the evolution of cluster knowledge networks. This finding suggests that the two sources of path-dependency, namely network retention driven by endogenous network effects and lock-in driven by cognitive proximity, do not strengthen each other. On the contrary, these forces seem to counter-act each other. A straightforward explanation of why firms ignore those ties that are cohesive in terms of network structure and also in terms of technological profile is that they are looking for new varieties of knowledge in the cluster. Consequently, network retention and network variation are simultaneously present in local knowledge networks.

Notwithstanding the new insights we provide, further research is needed to focus on the interference between retention and variation forces in knowledge networks. Based on our results, we propose that the creation and persistence of ties have to be analysed separately, because the micro level motivations of creating and maintaining ties are different. Further, we posit that the joint effect of endogenous network formation and proximities have to be investigated to get a clearer picture on how ties form in clusters. Such research shall aim not only to understand the patterns of relational change, the selection and retention mechanisms of network evolution, but also to take steps towards the recognition of forces that vary relational structures in clusters in a way that establishes new diversities in clusters. These together will allow us to fine-tune our understanding on how social networks and industry clusters co-evolve.

We have to emphasize the explanatory nature of our study and highlight some of the limitations and related future research opportunities. First of all, our results are based on a relatively small network with only a few nodes. Because stochastic models and especially the decomposition of creation and persistence of network ties in SAOMs requires large datasets, the generalization of our results should be careful. Based on the literature, other types of proximities, knowledge base or absorptive capacity of firms and the interplay of these with other structural variables are also need to be investigated (Giuliani, 2013; Balland et al., 2016). It must be stressed that the complex mixture of the analyzed factors might lead to different

dynamics across regions and industries because specializations differ in terms of thresholds of costs and benefits of cooperation (Gordon and McCann, 2000) and because the level of market uncertainties – e.g. strengthening competition or external shocks – might strongly influence network dynamics (Beckman et al., 2004). Further, our exercise is based on a mature cluster of printing and paper product creation with increasing level of competition. Therefore, the conclusions might be limited to traditional manufacturing clusters, and network dynamics in other stages of cluster lifecycle could be different (Ter Wal and Boschma, 2011; Suire and Vicente, 2014). Cohesive forces might have more influence on network change in an earlier life-cycle stage; competition or the fear from technological lock-in could change the willingness of cooperation in a later, mature or declining phase. According to the general thought, cognitive proximity has a dominant role in cluster lock-in (Boschma, 2005; Broekel and Boschma, 2012), which could intensify competition in clusters as well. This is an important point that future research shall address because repeated knowledge sharing increases the similarity of knowledge bases between co-located firms, which might lead competition and consequently thinning cooperation. Therefore, we shall also understand better the differences between tie creation and tie persistence in growing and in shrinking knowledge networks. The task is urgent because our models regarding tie persistence are not conclusive at all. A potential question can be, how does secrecy and free-riding influence knowledge network evolution? Further insights might be get from agent-based simulation models, in which agents punish those partners that are not sharing their knowledge by deleting the ties to them (Rand et al., 2011).

Additional limitation is – similarly to many studies on this topic – that the implications are based on the inter-firm alliance literature; however, advice networks might change more rapidly and the decision behind tie creation and persistence might be less strategic or even less conscious. Moreover, we are unable to control for the pre-existing friendships or other social ties among entrepreneurs, which might result more robust estimates. Moreover, our cognitive proximity measure simplifies the differences in knowledge bases of firms and therefore comparison to Giuliani (2013) is difficult. Further, ties are assumed to be identical in terms of transmitted content. Thus, the volume, depth and diversity of information content of the communications should be looked at (Aral and van Alstyne, 2011). This would allow us to investigate how the value of advice influences the persistence of ties, which we could not do in this chapter.

Another key issue for future research is the availability of longitudinal knowledge network data. With longer and more detailed relational datasets on cluster knowledge networks we might get answers to several, still open questions. First, we might get a better picture about how network dynamics change along the cluster life-cycle as we can investigate how the importance of structural and proximity effects change over time. Second, longitudinal data with more than two time points is needed to investigate tie re-creation, which might be driven by different forces than tie creation. Third, by using relational data on individual level

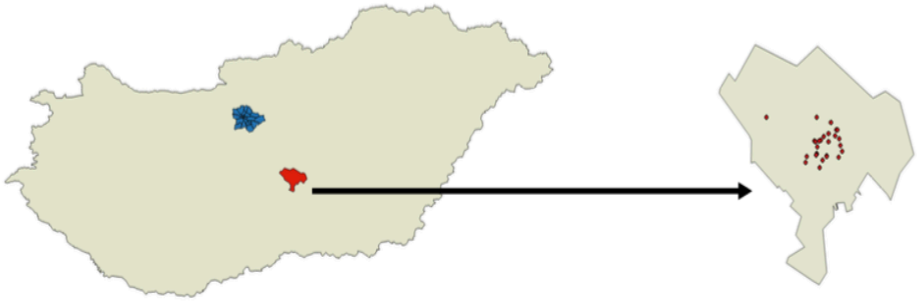
rather than firm level we might get much more accurate understanding on the motives of tie creation and persistence.

3.7 Appendix

3.A The location of Kecskemét, Hungary

The town of Kecskemét is about 80 km south from Budapest, the capital of Hungary, and accounts for around 115.000 inhabitants with an economy rooted in agriculture as well as processing and manufacturing industries. Most of the printing presses are located at the outskirts of the town but some of the firms dealing with pre-printing processes operate in the center of the town.

Figure 3.2 The location of Kecskemét and the firms in printing industry around the town



Source: Author's own data.
Note: (left) Kecskemét on the map of Hungary. The town of Kecskemét and Budapest, the capital of Hungary are highlighted. (right) The location of firms in the urban agglomeration of Kecskemét.

3.B Descriptive statistics of the sample in 2012 and 2015

The main characteristics of the examined firms did not change from 2012 to 2015. Two companies were closed down along the studied period, but two other companies joined to the sample by 2015. However, most of the firms were founded along the 1990s when self-owned firm foundation became possible in Hungary. The technological profile of the cluster is diverse; but printing dominates and less firms deal with paper product creation and pre-printing processes. The examined firms are mainly SMEs. There is only one firm with more than 100 employees and only a minority of them is foreign-owned (less than 20%). Interregional relations in terms of both export and net revenue ratio and extra-regional knowledge exchanges decreased over time. About 50% of the firms carried on export in both 2012 and 2015, but the number of external knowledge ties decreased from 7 to 4 on average.

Table 3.6 Descriptive statistics of the sample in 2012 and 2015

Characteristics	Number of firms		
	2012 (N=26)	Entry/exit	2015 (N=26)
Year of establishment			
Up to 1990	2		2
1990s	14		14
2000s	8		9
2010s	2		1
Entry		2	
Exit		2	
Main activities			
Paper product creation	7		6
Printing	12		11
Pre-printing processes	4		6
Other related activities	3		3
Size (number of employees)			
Small (1-10)	18		18
Medium (11-100)	7		7
Large (101-)	1		1
Average number of employees per firm	27		26
Ownership			
Domestic	21		21
Foreign	5		5
Exporters	13		11
Average number of knowledge linkages outside the region	7		4

Source: Author's own data.

3.C The change of core and periphery in 2012 and 2015

Similarly to previous studies (Giuliani and Bell, 2005), the core/periphery model of Borgatti and Everett (1999) identifies a cohesive group of central firms with high number of connections to each other and a group of peripheral firms loosely connected to the core and to each other at both points in time (Table 3.7). The fall of density affected every part of the network relatively equally. Both the core and the periphery became less connected by 2015 and the number of knowledge exchanges between the two parts also decreased. Furthermore, the composition of the core transformed as 25% of the core firms changed to periphery and 12.5% of them closed down.

Tabel 3.7 Core and periphery in 2012 and 2015, density and dynamics

	The density of linkages (knowledge transfer from row to column)		
	Core	Periphery	Final fit
2012			0.837
Core (n_c=8)	0.839	0.486	
Periphery (n_p=18)	0.389	0.163	
2015			0.841
Core (n_c=7)	0.738	0.391	
Periphery (n_p=19)	0.361	0.146	
	Stability of the core-periphery structure (dynamics in rows)		
Persistence	62.5%	88.8%	
Change	25%	5.6%	
Exit	12.5%	5.6%	

Source: UCINET 6 applied to author's own data.

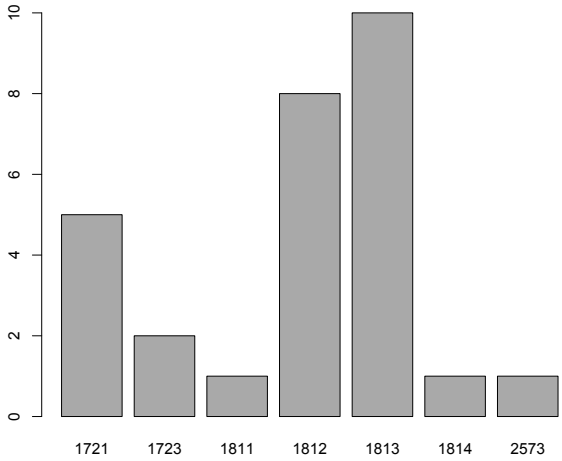
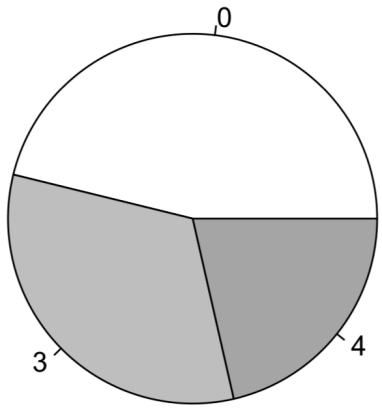
Note: The density of a network is the total number of ties divided by the total number of possible ties. The percentages are calculated on the population of firms present in 2012 (26 firms), therefore, it includes incumbents of 2015 but not new entrants.

3.D Descriptive statistics of cognitive proximity variable

To measure cognitive proximity of firms we used sectoral proximity which is based on the number of digits the two firms have in common in their NACE 4 codes (Balland et al., 2016). This assumes that two firms have similar technological profile and therefore are in cognitive proximity if they operate at the same sector category, which is in line with the related variety literature (Frenken et al., 2007).

In our case, the printing and paper product cluster consists of 28 firms and only 22.2% of all their possible connections have cognitive proximity values of 4, 33.6% have 3 and 44.2% have 0 as cognitive proximity measure. Because there are no 1 or 2 values of cognitive proximity (even though it is possible to have them) we also tried our basic model with a range of 0-2 for cognitive proximity as a robustness check. Results showed no difference.

Table 3.8 Descriptive statistics of the cognitive proximity indicator

Distribution of NACE codes among firms	Distribution of cognitive proximity values for all possible ties
	

Source: Authors' own data.

3.E Correlation tables for SAOMs

Network statistics can be highly correlated just because of their definition (Ripley et al., 2017). In stochastic actor-oriented models estimated parameters are allowed to be correlated, but their correlation could be important for the interpretation. Therefore, we append the correlation tables for all our model versions.

Tabel 3.9 Correlation of variables in Model (1) with evaluation function

	Density	Reciprocity	Triadic closure	Cyclicity	Cognitive proximity	Geographical proximity	Age alter	Age ego	Ownership similarity	External knowledge ties alter	External knowledge ties ego	Employment alter	Employment ego
Density													
Reciprocity	-0.190												
Triadic closure	-0.490	0.187											
Cyclicity	0.177	-0.503	-0.718										
Cognitive proximity	-0.070	-0.064	0.267	-0.087									
Geographical proximity	-0.004	0.028	0.005	-0.024	0.016								
Age alter	-0.101	0.085	-0.128	0.039	-0.189	-0.117							
Age ego	0.017	-0.080	-0.151	0.064	-0.209	-0.061	0.051						
Ownership similarity	-0.606	-0.050	-0.023	-0.032	-0.197	-0.036	0.173	0.137					
External knowledge ties alter	-0.081	-0.123	0.083	-0.092	-0.067	-0.235	0.055	0.000	0.232				
External knowledge ties ego	-0.056	0.048	0.089	-0.105	0.075	-0.265	0.101	-0.261	0.129	0.148			
Employment alter	-0.232	0.018	0.048	-0.078	0.122	-0.066	0.056	0.063	0.312	0.095	-0.013		
Employment ego	-0.066	-0.110	-0.118	0.110	0.041	-0.077	0.003	0.078	0.221	0.050	-0.003	0.147	

Source: Authors' own data.

Table 3.10 Correlation of variables in Model (2) with evaluation function

	Density	Reciprocity	Triadic closure	Cyclicity	Cognitive proximity	Geographical proximity	Common partners X cognitive proximity	Age alter	Age ego	Ownership similarity	External knowledge ties alter	External knowledge ties ego	Employment alter	Employment ego
Density														
Reciprocity	-0.188													
Triadic closure	-0.532	0.137												
Cyclicity	0.210	-0.521	-0.669											
Cognitive proximity	-0.270	-0.194	0.338	0.022										
Geographical proximity	-0.080	-0.048	0.021	0.015	0.130									
Common partners X cognitive proximity	0.310	0.171	-0.248	-0.068	-0.748	-0.096								
Age alter	-0.165	-0.018	-0.026	0.059	0.116	-0.066	-0.285							
Age ego	-0.001	0.026	-0.182	-0.046	-0.175	-0.140	0.079	0.029						
Ownership similarity	-0.620	0.095	0.055	-0.173	-0.050	0.022	-0.137	0.193	0.203					
External knowledge ties alter	-0.010	-0.067	0.010	-0.059	-0.131	-0.199	0.091	0.086	0.121	0.170				
External knowledge ties ego	-0.146	-0.061	0.284	-0.089	0.338	-0.177	-0.437	0.118	-0.299	0.113	0.041			
Employment alter	-0.272	0.068	0.108	-0.129	0.123	-0.053	-0.085	0.030	0.052	0.313	0.126	0.008		
Employment ego	-0.118	0.042	0.054	-0.072	0.108	-0.123	-0.107	0.045	0.054	0.224	0.078	0.165	0.132	

Source: Authors' own data.

Table 3.11 Correlation of variables in Model (1) with creation function

	Density	Reciprocity	Triadic closure	Cyclicity	Cognitive proximity	Geographical proximity	Age alter	Age ego	Ownership similarity	External knowledge ties alter	External knowledge ties ego	Employment alter	Employment ego
Density													
Reciprocity	-0.372												
Triadic closure	-0.648	0.550											
Cyclicity	0.331	-0.736	-0.773										
Cognitive proximity	-0.078	0.141	0.281	-0.144									
Geographical proximity	0.015	-0.077	-0.092	0.030	-0.123								
Age alter	-0.170	0.041	0.034	-0.053	-0.225	-0.064							
Age ego	0.009	-0.056	-0.070	-0.049	-0.235	-0.014	0.082						
Ownership similarity	-0.618	-0.026	0.126	-0.126	-0.208	-0.052	0.177	0.200					
External knowledge ties alter	0.012	-0.190	-0.097	0.046	-0.114	-0.132	0.095	0.114	0.248				
External knowledge ties ego	-0.059	0.094	0.098	-0.071	0.036	-0.306	0.070	-0.150	0.042	0.068			
Employment alter	-0.260	0.014	0.118	-0.112	0.132	-0.109	0.063	0.033	0.406	0.199	0.098		
Employment ego	-0.230	0.065	0.089	-0.100	0.059	-0.069	0.108	0.101	0.330	0.081	0.097	0.219	

Source: Authors' own data.

Table 3.12 Correlation of variables in Model (2) with creation function

	Density	Reciprocity	Triadic closure	Cyclicity	Cognitive proximity	Geographical proximity	Common partners X cognitive proximity	Age alter	Age ego	Ownership similarity	External knowledge ties alter	External knowledge ties ego	Employment alter	Employment ego
Density														
Reciprocity	-0.375													
Triadic closure	-0.744	0.526												
Cyclicity	0.339	-0.732	-0.735											
Cognitive proximity	-0.331	-0.008	0.335	-0.001										
Geographical proximity	-0.224	-0.008	0.154	-0.099	0.082									
Common partners X cognitive proximity	0.473	0.050	-0.374	0.014	-0.672	0.002								
Age alter	-0.145	0.006	0.020	0.046	0.091	-0.196	-0.196							
Age ego	-0.204	0.040	0.182	-0.215	-0.048	-0.127	-0.127	0.168						
Ownership similarity	-0.635	0.156	0.320	-0.291	-0.011	-0.128	-0.128	0.049	0.240					
External knowledge ties alter	0.138	-0.064	-0.066	-0.095	-0.175	0.143	0.143	0.007	-0.055	0.117				
External knowledge ties ego	-0.142	0.049	0.155	-0.054	0.235	-0.338	-0.338	0.022	-0.181	0.111	0.112			
Employment alter	-0.206	0.012	0.132	-0.074	0.186	-0.084	-0.034	0.011	-0.047	0.294	0.111	0.003		
Employment ego	-0.224	0.080	0.116	-0.076	0.111	-0.101	-0.075	0.025	-0.002	0.327	-0.016	0.168	0.226	

Source: Authors' own data.

Table 3.13 Correlation of variables in Model (1) with endowment function

	Density	Reciprocity	Triadic closure	Cyclicity	Cognitive proximity	Geographical proximity	Age alter	Age ego	Ownership similarity	External knowledge ties alter	External knowledge ties ego	Employment alter	Employment ego
Density													
Reciprocity	-0.277												
Triadic closure	0.030	-0.450											
Cyclicity	-0.268	0.355	-0.864										
Cognitive proximity	-0.012	-0.041	-0.027	0.066									
Geographical proximity	0.021	-0.047	-0.059	0.061	0.016								
Age alter	-0.067	0.127	-0.123	0.072	-0.221	-0.096							
Age ego	0.078	-0.274	0.397	-0.417	-0.022	-0.041	-0.126						
Ownership similarity	-0.534	-0.030	0.017	-0.105	-0.112	-0.056	0.118	0.011					
External knowledge ties alter	-0.029	-0.189	0.292	-0.347	-0.117	-0.261	0.073	0.084	0.285				
External knowledge ties ego	-0.082	0.268	-0.513	0.404	0.045	-0.230	0.165	-0.440	0.257	0.008			
Employment alter	-0.198	0.034	-0.090	0.052	0.111	-0.060	0.060	0.030	0.291	0.127	0.113		
Employment ego	-0.197	-0.019	0.044	-0.036	0.164	-0.096	0.013	0.003	0.264	0.039	0.068	0.061	

Source: Authors' own data.

Table 3.14 Correlation of variables in Model (2) with endowment function

	Density	Reciprocity	Triadic closure	Cyclicity	Cognitive proximity	Geographical proximity	Common partners X cognitive proximity	Age alter	Age ego	Ownership similarity	External knowledge ties alter	External knowledge ties ego	Employment alter	Employment ego
Density														
Reciprocity	-0.221													
Triadic closure	0.133	-0.355												
Cyclicity	-0.371	0.193	-0.865											
Cognitive proximity	-0.189	-0.108	0.082	0.012										
Geographical proximity	-0.013	-0.013	-0.186	0.166	-0.061									
Common partners X cognitive proximity	0.217	0.131	-0.162	-0.035	-0.796	-0.011								
Age alter	-0.117	0.055	-0.045	0.006	0.006	0.008	-0.167							
Age ego	0.104	-0.214	0.396	-0.440	-0.032	-0.091	-0.085	-0.009						
Ownership similarity	-0.529	0.015	-0.054	-0.017	0.059	-0.073	-0.128	0.165	0.053					
External knowledge ties alter	0.046	-0.127	0.190	-0.264	-0.110	-0.188	0.004	0.136	0.152	0.206				
External knowledge ties ego	-0.089	0.205	-0.364	0.319	0.203	-0.221	-0.089	0.006	-0.332	0.258	-0.123			
Employment alter	-0.136	-0.030	0.145	-0.153	0.123	-0.091	-0.056	0.031	0.092	0.282	0.128	0.049		
Employment ego	-0.168	0.042	-0.006	0.019	0.120	-0.102	-0.038	0.016	-0.010	0.220	0.009	0.094	0.075	

Source: Authors' own data.

Chapter 4

Explaining dynamics of relatedness: the role of co-location, complexity and collaboration

This chapter has been produced in collaboration with Tom Broekel and Ron Boschma. The PhD candidate is the first author of the research paper.

4.1 Introduction

The relatedness concept has drawn lots of attention in Management Studies, Economic Geography and Innovation Studies (Teece et al., 1994; Boschma, 2017; Hidalgo et al., 2018). Relatedness refers to the fact that two activities (technologies, industries, products, jobs) are not identical but share commonalities. Such commonalities may originate from two activities belonging to the same overarching technological or economic field, or they share complementarities or similarities (Nooteboom, 2000; Boschma, 2005). There is ample evidence that relatedness of technologies conditions whether a technology will be invented or adopted by a firm, region or country. Moreover, relatedness has been shown to influence performance and diversification of firms (Fornahl et al., 2011; Neffke et al., 2012) and regions (Frenken et al., 2007; Neffke et al., 2011). Recently, relatedness has also found its way into policy as an evaluation and benchmarking dimension (Balland et al., 2018; Fitjar et al., 2019).

However, so far, we know little about how (technological) relatedness comes into existence, and how it develops over time (Menzel, 2015). In fact, in most of the empirical studies relatedness is either treated as time-invariant, or as an exogenously given component. While we might expect that relatedness is rather stable in the short run, this is less likely to be the case in the long-run (Balland et al., 2015; Boschma, 2017). Relatedness has been used as an independent variable to explain all kinds of empirical phenomena, like the economic performance of firms and regions, but few studies have analyzed relatedness as dependent variable, and how relatedness changes and evolves over time. This motivates the chapter in studying the emergence and evolution of technological relatedness.

We seek to identify some factors contributing to the dynamics of relatedness between technologies. We focus on three factors that might shape technological relatedness: co-location, complexity and collaboration. Making use of patent data, our empirical study shows that co-location and complexity of technology pairs determine the emergence of technological relatedness and its development in Europe during the period 1980-2010. Moreover, we provide evidence that the importance of these factors is conditional on technologies being combined by collective efforts of inventors.

This chapter is organized as follows. Section 4.2 provides a short introduction on the concepts of relatedness and its possible interplays with co-location, complexity, and collaboration. Section 4.3 introduces the empirical data, key variables and modelling approach. Section 4.4 presents and discusses findings in detail. The chapter concludes with a discussion that outlines limitations, implications and possibilities for future research.

4.2 Dynamics of relatedness

Two activities (such as technologies, industries, occupations or research fields) are considered related when they are based on similar knowledge, skills or other inputs (Hidalgo et al., 2018). There are various dimensions of relatedness (like relatedness across skills, technologies, or products), depending on the type of similarities that one considers (Boschma, 2017). Many studies focus on technological relatedness, defining the degree to which two technologies are proximate to each other in a technological or cognitive way (Breschi et al., 2003; Kogler et al., 2013; Boschma et al., 2015; Rigby, 2015).

Relatedness has become a key input to outline possible re-combination and diversification opportunities in particular. There is consensus that the probabilities of firms, regions and countries to enter new and specific activities is a function of the number of the related activities they are specialized in (Neffke et al., 2011; Boschma et al., 2015; Hidalgo et al., 2018). Similarly, regions and countries with access to related variety tend to outperform economically those that are either highly specialized or overly diversified (Frenken et al., 2007; Hidalgo et al., 2007; Fornahl et al., 2011).

However, the relatedness literature puts less emphasis on relatedness as an evolving property. We know little about how relatedness comes into existence, and how it develops over time (Balland et al., 2015; Menzel, 2015). In most studies on regional diversification, relatedness is often being treated as time-invariant or exogenously given: relatedness is included in the empirical model as an independent variable that is held constant over the observed time period (e.g. Breschi et al., 2003; Frenken et al., 2007; Neffke et al., 2011). While it makes sense to assume that relatedness is stable in the short run (technologies tend to be related to other technologies for some fixed period), this is less likely to be the case in the long-run (Cowan et al., 2007; Balland et al., 2015; Boschma, 2017) when technological paradigm shifts (Dosi, 1982) reshuffle the technology space (Rigby, 2015), as happened with the rise of electronics and biotech (Krafft et al., 2011, 2014). In some studies, relatedness is modeled as time-variant, with empirical variations between time periods (e.g. Boschma et al., 2015; Rigby, 2015). However, these studies do often not discuss how relatedness emerges or reaches a particular level, and why it changes over time. In other words, relatedness is often used as an independent rather than dependent variable, with few studies focusing on how relatedness evolves over time.

Relatedness (and changes thereof) might not be independent of other factors. Co-location may be crucial here. Geographical proximity enhances knowledge spillovers between economic agents (Jaffe et al., 1992; Audretsch and Feldman, 1996). It thereby enhances localized learning processes and lays the foundation for technological (re-)combinations. The geographical co-location of actors with expertise in different technologies makes it more likely

that they engage in knowledge exchange and thereby increase the likelihood of these technologies to be (re-)combined, and thus increase their levels of relatedness. This builds on Jacobs (1969) who argued that diversity of ideas, knowledge or technologies in cities enables and triggers cross-fertilization and re-combinations. The co-location of technologies can bring together previously unrelated technologies and enhance the discovery of new complementarities between technologies which make them related as a result (Desrochers and Leppälä, 2011). Moreover, co-location reinforces already existing technological relatedness and lead to its further intensification and possibly to a situation of lock-in (Boschma, 2017). The process in which co-location leads to or alters technological relatedness has not yet been studied in detail. Therefore, we test the following hypothesis 1:

Hypothesis 1: co-location enhances the emergence and intensification of relatedness between technologies

Complexity is also expected to matter for technological relatedness because it influences combinatorial processes. The complexity of a technology is often understood as a function of the number of its (sub)components and their interdependence (Simon, 1962; Fleming and Sorenson, 2001). Combining two technologies requires a basic understanding and mastering of both. Gaining such an understanding is more difficult, the more complex technologies are (Hidalgo and Hausmann, 2009; Balland and Rigby, 2017). Therefore, the complexity of two to-be-combined technologies increases the required re-combinatorial efforts (Fleming and Sorenson, 2001; Sorenson and Fleming, 2004). This will even hold when the combination process is conducted by a team. Even in this case, all members of the team must have a basic understanding of the involved technologies, which, as argued above, is more difficult in case of two complex technologies. Moreover, due to the greater knowledge diversity inherent to more complex technologies, it is likely that parts of these technologies are at greater cognitive distances. Complex technologies are likely to be made of relative different and unrelated (sub)components (Broekel, 2019). This will make communication and learning within teams harder, and consequently collaborative (re-) combination. So, combining two complex technologies is more difficult than combining two simple ones.

Technologies also have a value that reflects on their supply and demand. Technologies that are simple to copy, and which can be moved easily over space, represent less economic value (Maskell and Malmberg, 1999). For this reason, technological complexity is argued to correlate positively to the socio-economic value of technologies (Fleming and Sorenson, 2001). Put differently, combinations of complex technologies are associated with higher economic rewards (Dalmazzo, 2002; Hidalgo and Hausmann, 2009). So, while these are more difficult to create and to develop, they offer larger economic incentives. Therefore, combinations of complex technologies might be even more common than the combination of less complex technologies (Broekel, 2019). So complexity acts as (initial) barrier to technological combinations and thereby to relatedness, but once this barrier is overcome, the

higher social-economic benefits associated to the combination of complex technologies may actually support the development of relatedness. By driving technologies to be more frequently combined, complexity may reduce the distance between the two and foster the emergence of new similarities and complementarities, and thus new relatedness.

In sum, complexity represents a dimension of technologies that matters for relatedness by shaping the likelihood and frequency of their (re-)combination. While economic incentives may represent a powerful opposing force, we expect that more complex technologies are less likely to be combined. Consequently, we expect complexity to hamper technologies becoming more related. However, over time, complexity may facilitate the development of relatedness. We formulate the second hypothesis as follows:

Hypothesis 2: more complex technologies are less likely to be related, but complexity may support the development of relatedness

Human collaboration creates a pool of ideas, skills and knowledge that enables the combination of technologies (Hidalgo, 2015). Many inventions and technological (re-)combinations are done in a collaborative fashion (Wuchty et al., 2007; Gilsing et al., 2008). Yet, not all invention processes are collaborative in nature. In many instances, individual inventors collect and combine knowledge from different sources, which may or may not be in geographical vicinity. Whether inventive activities are done individually or collaboratively may matter for the impact of co-location and complexity on relatedness.

Geographical proximity can be expected to be particularly important in cases in which technologies are combined in a collaborative manner. While individuals may also have a bias towards local knowledge sources (Broekel and Binder, 2007), geographical proximity greatly helps in realizing face-to-face interactions and by lowering their transaction costs. Co-location is of much greater relevance in collaborative settings in which actors with heterogeneous knowledge have to identify each other first, learn how to interact, establish efficient knowledge sharing and learning, and eventually jointly engage in interactive invention processes. Consequently, co-location can be expected to be of greater relevance for relatedness in case technological (re-)combinations are done in collaboration.

Whether technological relatedness emerges and intensifies as a result of individual or collaborative invention processes, may also matter for the effect of complexity on relatedness. It is well accepted that complex technologies require more collaborative efforts (Hidalgo and Hausmann, 2009; Balland and Rigby, 2017; Broekel, 2019). In fact, one may argue that collaboration is a strategy to deal with high complexity. Accordingly, we might expect higher complexity in cases of collaborative invention processes, and hence a larger negative influence of complexity on the existence and intensification of relatedness in cases of collaboration. However, by overcoming the challenges of complexity, collaboration may actually weaken the observable effect of complexity on relatedness. Put differently, given the strong link between complexity and collaboration (Broekel, 2019), it might be difficult to isolate the effect of

complexity when accounting for collaborative efforts. Our third hypothesis emphasizes this two-sided view on collaboration and complexity:

Hypothesis 3: collaboration amplifies the importance of co-location for technological relatedness, while it weakens the effect of complexity on technological relatedness

In sum, we expect co-location and complexity to impact on the presence and strength of relatedness between technologies, but both effects will be mediated by collaboration.

4.3 Empirical approach

4.3.1 Dataset and dependent variable

We rely on patent data as our primary database. We use the OECD REGPAT database (version 2018) covering patents registered by the European Patent Office (EPO). It contains detailed information on patents' application year, technology classes, inventors and inventor locations since 1976. The pros and cons of patent data have been extensively discussed, so we refrain from this and refer to the relevant literature (Griliches, 1990; Desrochers, 1998).

We focus on the period 1976-2010 that offers reliable data. Figure 4.1A illustrates the growth of the number of patent applications during this period. To track the evolution of technological relatedness, we constructed seven non-overlapping 5-year time periods from 1976 to 2010, with the first covering the years of 1976-1980, the second 1981-1985, and so on. By pooling the data of multiple periods, we increase the stability of our measures, as patent numbers tend to fluctuate strongly between years. This is particularly the case when they are aggregated at the regional or technological level (Buerger et al., 2012).

To empirically represent technological relatedness, we need to define technologies first. We follow existing studies and consider technologies being represented by four-digit CPC classes (Balland et al., 2018; Broekel, 2019). We exclude technologies starting with the letter *Y*[10] that indicates cross-technological patents, i.e. patents that cannot be clearly assigned to other classes. This leaves us with 646 distinct technologies. Figure 4.1B illustrates that almost all of these technologies are present in the dataset since 1980. Accordingly, four-digit CPC classes appear relatively consistent during the examined period.

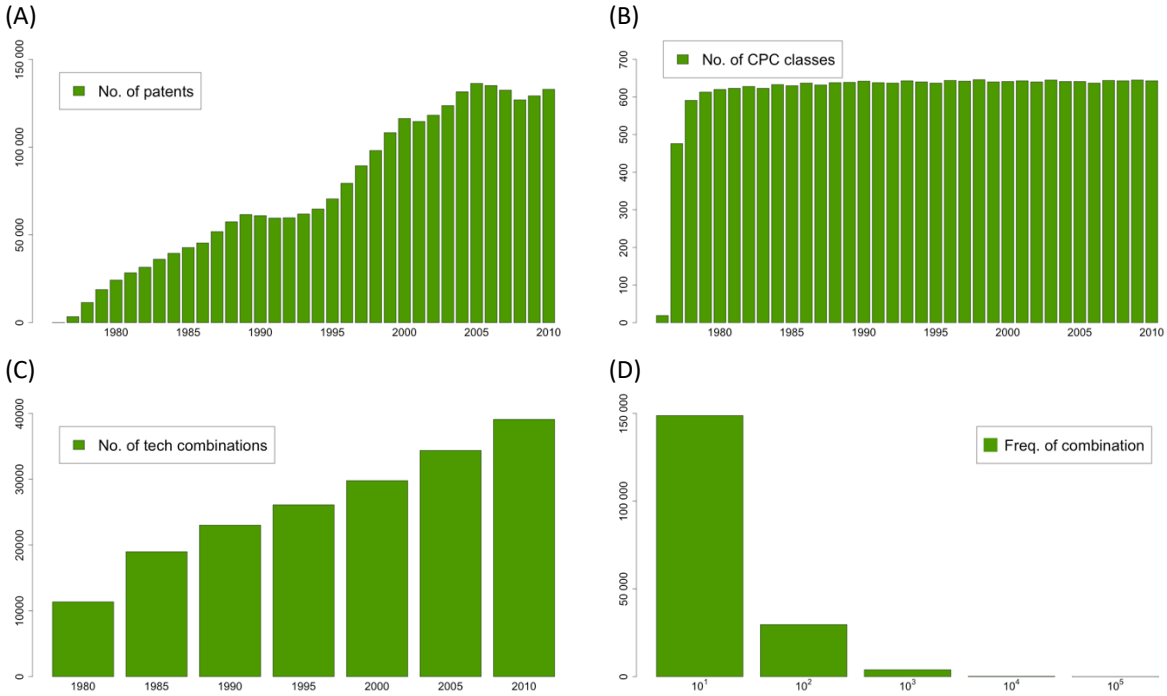
There are many ways how technological proximity can be estimated. Most studies use the information on technologies jointly appearing on patents, so-called co-occurrences, as an indication of combinatorial innovation processes and technological distance (Breschi et al., 2003). We follow this literature and concentrate on the joint appearance of technologies (four-digit CPC) classes on patents.

[10] The 'Y' CPC class is only a technical category with significant overlap to other classes.

The 646 technologies translate into 208,335 potential technological combinations (excluding self-combinations) which serve as units of observation. For each of these pairs, we count the number of occurrences on patent documents within each of the seven time periods. More precisely, we count how frequently different four-digit CPC classes are combined on patents during the given period, which implies we do not allow for multiple combinations of the same four-digit classes for simplification.

Figure 4.1C visualizes the number of distinct technological combinations for each five-year period. Their number is increasing linearly over time. This growth is both in absolute and relative terms. The share of realized combinations of technologies (appearing on at least one patent) on the potential number of combinations grows from around 5% to 18%. This is explained by the strongly growing numbers of patents and the more or less constant number of technologies (4-digit CPC classes). Figure 4.1D highlights there are few technology combinations that appear in high frequencies (>100 times), while the vast majority is realized in rather small numbers (<10).

Figure 4.1 Patents, technology classes and technology combinations



Source: Authors' own construction based on OECD REGPAT 2018

Notes: (A) Number of patents in each year, 1976-2010

(B) Number of different 4-digit CPC technology classes that appear on patents in each year, 1976-2010

(C) Number of observed distinct technological combinations. We measure technologies as 4-digit CPC technology classes and by combination we count the co-occurrence of 2 technologies on patents in the selected 5-year long period.

(D) Distribution of combination frequencies in the full sample over all the periods. For this chart, we excluded combinations with 0 frequency. Note that the x axis is on a log10 scale.

In general, each patent represents a chance of two technologies co-occurring in the data. Therefore, the likelihood of two technologies jointly appearing on patents is primarily determined by the size of the two technologies, i.e. by the number of patents that are assigned to these technologies (jointly and independently). The literature suggests a range of approaches to control for this purely statistical effect. For instance, some studies employ the Cosine index (Breschi et al., 2003; Ejermo, 2003) or alternative measures of co-occurrence standardization (Kogler et al., 2013; Mewes, 2019). However, specifying the correct standardization / normalization is far from straightforward (van Eck and Waltman, 2009).

In contrast to many other studies, we are not so much interested in which technologies are statistically significantly related and which are not. In contrast, we seek to approximate technological distance between two technologies in a continuous manner. We therefore follow the idea of Neffke (2009) and Neffke et al. (2011) and focus on revealed relatedness. That is, the absolute numbers of co-occurrences of technologies serve as dependent variable in a regression. To account for the fact that large technologies have more possibilities of co-occurring on patents, we include the absolute numbers of patents that are assigned to either of the two technologies, alongside other variables as explanatory factors. Put differently, these numbers of co-occurrences that are predicted by these two variables can be seen as the baseline number of co-occurrences. These are the numbers that can be expected given the size of two technologies. Additional variables considered in the regression model are then used to explain the observed deviations from this baseline number, i.e. they give insights into why some technological combinations are observed more or less frequently than what can be expected by just considering the numbers of patents of two technologies. In this context, we are interested in the roles geographical distance and complexity play.

4.3.2 Co-location of technologies

As discussed above, we expect technologies to become more related when they are frequently co-located. Empirically, this corresponds to the overlap of their spatial distributions. We firstly assign patents to the NUTS2 regions of their inventor addresses. Patents with multiple inventors in multiple regions are assigned to multiple regions (no fractional counting). Next, we obtain the spatial distribution of a technology's patents, by counting the number of patents assigned to this technology per region. On this basis, we assess the overlap in two technologies' distributions with the co-agglomeration measure proposed by Ellison et al. (2010). For two technologies *i* and *j*, the measure looks as follows:

$$COAGGLOM_{i,j} = \frac{\sum_{r=1}^{R}(s_{ir}-x_r)(s_{jr}-x_r)}{1-\sum_{r=1}^{R}x_r^2}. \tag{4.1}$$

whereby P_{ir} is the number of patents in technology *i* in region *r* and $s_{ir} = \frac{P_{ir}}{\sum_{r'=1}^{R} P_{ir'}}$ is the share of technology *i* in region *r* (and for *j* accordingly). x_r is the mean of these shares in region *r* across all technologies. The index has been used in multiple studies investigating industrial co-agglomeration (Ellison et al., 2010; Diodato et al., 2018). Its main advantage is that it is largely invariant to the distribution of inventors, applicants and of the considered spatial units (Diodato et al., 2018).

We calculate the co-location of the 208,335 technology pairs in all seven periods. To reduce the noise in the data, we set the patent counts to zero for all regions with less than ten patents in a particular technology. Moreover, we exclusively consider the appearing 316 EU and EFTA NUTS-2 regions and patents associated to these11. To make the score comparable across technologies, we consider its z-score with 0 mean in the estimations.

4.3.3 Technological complexity

Empirically approximating technological complexity is a challenge with the literature offering few approaches that are applicable to the data at hand. Balland and Rigby (2017) transferred the idea of the Economic Complexity index of Hidalgo and Hausmann (2009) to patent data and calculated an index of technological complexity. The approach assesses complexity of technologies by the method of reflection. In essence, the method utilizes information on the extent technologies are ubiquitous (few regions are specialized in these) and how diversified regions are (in how many technologies they are specialized). By means of an iterative procedure, the index sets these two aspects into relation and measures the complexity of technologies by their tendency of co-concentration with other ubiquitous technologies.

Earlier, Fleming and Sorenson (2001) employed an *N/K model* to derive a technological complexity index on the basis of the frequencies with which patent subclasses co-occur on patents. The underlying argument is that more complex combinations are harder to realize and thereby less likely to be observed. The difficulty of (re-)combining two classes is empirically assessed by evaluating the currently observed co-occurrence of patent subclasses on patents in relation to the cumulative frequency of their co-occurrence in previous years.

While both approaches have their merits, Broekel (2019) argues that some of the core assumptions underlying these measures are problematic. Moreover, he shows that both indices fail in resembling a number of stylized facts commonly associated with technological complexity. However, more importantly, in our analyses both indices bear the danger of being very closely related to key variables in the model. The complexity index of Balland and Rigby (2017) is based on spatial co-location of technologies, which is already captured by our measure of co-agglomeration. Fleming and Sorenson's (2001) approach utilizes the frequency

[11] Detailed descriptive statistics of the coagglomeration variable can be found in the Appendix (section 4.A and 4.B).

of patent class co-occurrences on patents as central ingredient, which is also the basis of our technological relatedness measure. To avoid the potential of (artificial) empirical overlap, we turn to an alternative measure of complexity that does not raise these concerns.

We adopt the structural complexity index developed by Broekel (2019). It measures technological complexity by modeling technologies (4-digit CPC classes) as combinatorial networks of (knowledge) components. While in this, co-occurrence information of patent classes on patents is used as well, in this case, these networks are dichotomized eliminating the frequency information, which is central in our measure of technological distance. Broekel (2019) argues further that the combinatorial networks of complex technologies are characterized by a greater diversity of (sub-)network topologies. Network science shows that a greater diversity implies a larger information content of these networks. Broekel (2019) argues that this demands larger R&D efforts for the invention of such technologies and also represents a greater obstacle to the learning or copying of this technology. Both aspects make complex technologies more exclusive and potentially economically more valuable.

To capture this diversity, Broekel (2019) proposes the *network diversity score* of Emmert-Streib and Dehmer (2012), as empirical approximation of *structural diversity* of technologies i.e. as empirical measure of technological proximity. Following Broekel (2019), the network diversity score is estimated as follows. For each technology *T* (four-digit CPC class), its combinatorial network is constructed in period *t*, by extracting all patents with at least one patent subclass (ten-digit CPC subclass) belonging to technology *T*. Next, the co-occurrence matrix of all patent subclasses appearing on these patents is created and dichotomized with all positive entries being one, and all others remaining zero. This matrix represents the combinatorial network of technology *T* in which complexity is assessed by means of the *network diversity score* (NDS). In the estimation of the NDS, multiple subsamples are drawn from the networks' main component. For each subsample *i*, the share of modules (α_{module}), the variability of module sizes (v_{module}), the variability of the Laplacian matrix ($V_{\text{Laplacian}}$), and the relation of graphlets of size three and four ($r_{\text{graphlets}}$) are calculated and the individual *NDS* score is estimated as:

$$iNDS_i = \frac{\alpha_{\text{module}} * r_{\text{graphlets}}}{V_{\text{Laplacian}} * V_{\text{module}}} \tag{4.2}$$

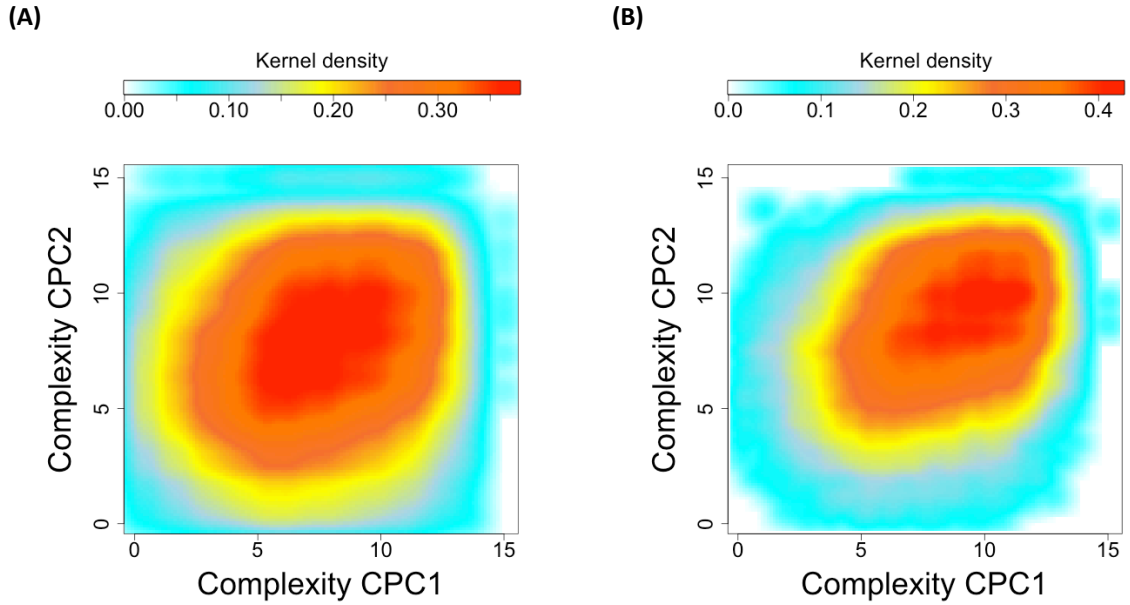
The *iNDS* is subsequently averaged over sample networks giving the NDS for this technology's combinatorial network. It is log-transformed and multiplied by -1 to obtain the final complexity measure of *structural diversity*, which signals higher complexity with larger values.[12]

[12] For a more detailed introduction to the measure see Broekel (2019).

In practice, we use the values of *structural complexity* given in Broekel (2019).[13] These are annual complexity values for each of the 646 four-digit CPC classes. As structural complexity scores are relative stable over short time (Broekel, 2019), we used the complexity values of technologies given at the end of each period.

As discussed in Section 4.2, we expect complexity to influence the likelihood of two technologies being combined, i.e. co-occurring on patents. Crucially, this influence might be negative as more complex technologies are more difficult to be combined. However, complexity may also have a positive impact by signaling the potential of greater rewards associated to the combination of two complex technologies. In any case, it is less the individual complexity values of technologies that matter but rather their joint configuration. We model this with the sum of structural complexity values of two technologies (*COMPLEX_SUM*).

Figure 4.2 Complexity of possible and realized technology combinations



Source: Authors' own construction based on OECD REGPAT 2018

Note: For the construction of these charts we do not split the dataset to periods.

(A) The density of all the possible technological combinations on a complexity-complexity plot (1980-2010)

(B) The density of all observed technological combinations on a complexity-complexity plot (1980-2010)

Figure 4.2A gives a first impression on this by visualizing how all technology pairs' complexity values potentially align. Figure 4.2B then shows how these values are actually distributed across the observed (co-occurring) technology pairs. The figures suggest that most of the realized combinations are pairs with mid or higher value of complexity, which supports our theoretical argumentation of complexity positively correlating to the (economic and technological) potential of combinations of technologies. To control for the possibility that the sum of two technologies' complexity is relatively high as a result of a combination of a highly

[13] The data can be downloaded via: https://doi.org/10.1371/journal.pone.0216856.s007

complex and a rather simple technology, we also control for the absolute difference in two technologies' complexity values (*COMPLEX_ABSDIFF*).

4.3.4 Collaboration and relatedness

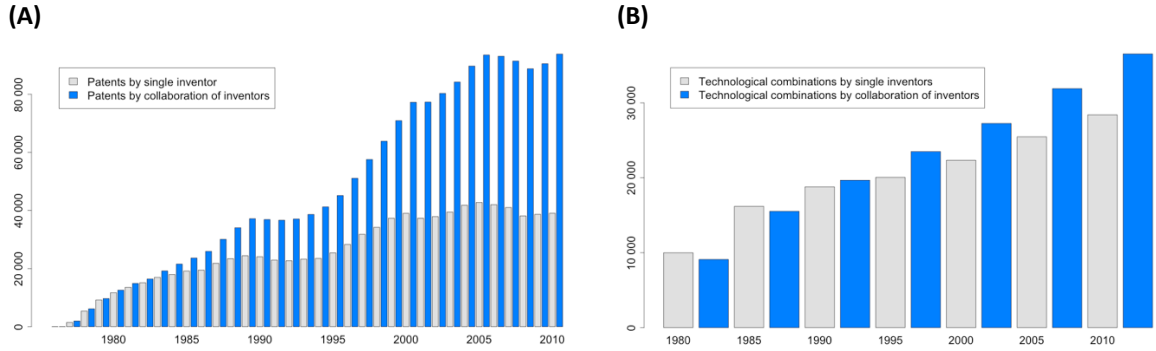
As discussed in Section 4.2, we expect collaboration to mediate the influence of co-location and complexity on technological relatedness. To isolate its influence, we decided to split the data and run the analyses separately on different subsamples. We divided the whole dataset into patents with multiple inventors (COLLAB, 1,692,750 patents in our entire sample) and patents with a single inventor (SINGLE, 910,228 patents in our entire sample). For both of these data sets, we estimated the relatedness measure anew. That is, the SINGLE measure counts the number of patents by single inventors that combine two distinct technologies during the given period, while the COLLAB measure counts the co-occurrence of technologies on patents created in collaboration during each of the 5-year periods. The relatedness measure based on the multi-inventor patents will be called COLLAB in the following and represents the degree of relatedness that is achieved between technologies exclusively through collaborative invention activities. In contrast, SINGLE approximates the extent to which individual inventors are able to combine technologies.

Since the beginning of the 1980s, more patents are produced in collaboration than by single inventors (Figure 4.3A). Since the 2000s, this relation has further increased in favor of patents produced by teams. This is also reflected in the contribution of collaborative inventions to combination of technologies. Figure 4.3B shows that, after the first two periods, collaboration accounts for larger portions of technological combinations than individual inventions, implying that the majority of invention processes is collaborative more recently.

In addition to the relatedness measure based on all patents, we conduct our empirical analyses using these two relatedness variables (REL_COLLAB, REL_SINGLE) as dependent, and all the other variables as explanatory. By comparing the results, we explore in what context co-location and complexity are of greater importance for the development of technological relatedness.

Of course, in practice, both types of invention processes (collaborative and individual) will take place simultaneously and most likely also interdependent. For instance, an inventor may make an individual invention combining two technologies, which is based on his or her previous collaborative work. Accordingly, it is impossible to truly isolate both dimensions and the results should therefore be interpreted with caution. Nevertheless, we believe it is a relatively easy way to explore in what settings (individual and collaborative invention processes) co-location and complexity exert a greater impact.

Figure 4.3 Patterns of collaboration on patents



Source: Authors' own construction based on OECD REGPAT 2018

Note: Technology pairs are the edges in the technology space with non-zero weight.

A) Number of patents produced by single inventors and collaboration teams along the period of 1980-2010.

(B) Number of observed technology pairs by single inventors and collaboration teams along the 1980-2010 period.

4.3.5 Empirical model

As pointed out, we follow the idea of Neffke (2009) and Neffke et al. (2011) by concentrating on revealed relatedness. This implies that our dependent variables account for the three different relatedness measures (REL_FULL, REL_SINGLE, REL_COLLAB). They represent the dyadic relation between two technologies implying that our empirical model seeks to explain a relation/link. More precisely, they are the absolute counts of co-occurrences of two technologies *i* and *j* on patents. To capture the fact that large technologies have more opportunities to potentially link, these absolute counts are related to the absolute numbers of patents that are assigned to either of the two technologies. As we also consider additional relational variables, i.e. variables that describe the relation between technologies *i* and *j*, e.g. their degree of geographical co-location, the revealed relatedness framework actually corresponds to a gravity regression approach commonly used in the trade and spatial spillover literature (Burger et al., 2009; Scherngell and Barber, 2009; Gómez-Herrera, 2012). In these models, the strength of links – in our case the strength of relatedness between two technologies – are estimated in a regression framework that contains both dyadic (relation) covariates and variables related to both technologies involved in the dyad.

Our dependent variables are counts suggesting the use of Poisson regressions. However, only a minority of technological combinations are realized in every period and many technologies are not related at all (Figure 4.1C). Due to induced over-dispersion of the variable, we apply Zero-Inflated Negative Binomial (ZINB) regressions. ZINBs consider the existence of two latent groups within the population: a group that has strictly zero combinations and a group having counts other than zero. Correspondingly, the estimation process takes place in two parts (Greene, 1994; Burger et al., 2009). The first part of ZINBs is

the zero-inflation part that contains a logit (or probit) regression of the probability that there is no technological combination at all. The second part contains a Poisson regression of the probability of each count for the group that has combination count other than zero.

We have no reasons to believe that a different set of factors is responsible for the (first) realization of technologies combinations and the intensification thereof. Accordingly, we include the same explanatory variables into both parts of the regression. Our model specification for the zero-regression is the following:

$$\Pr(y_{ijt} = 0) = \beta_0 + \beta_1 \log(y_{ijt-1}) + \beta_2 COAGGLOM_{ijt} + \beta_3 COMPLEX\_SUM_{ijt} + \beta_4 COMPLEX\_ABSDIFF_{ijt} + \beta_5 \log(N_{it}) + \beta_6 \log(N_{jt}) + \eta_t \quad (4.3),$$

with $\Pr(y_{ijt} = 0)$ being zero in case of no co-occurrences between technology *i* and *j* in period *t,* and one otherwise. y_{ijt-1} is the strength of technological relatedness between technology *i* and *j* in the previous period. *COAGGLOM, COMPLEX_SUM* and *COMPLEX_ABSDIFF* denote the characteristics of the technology pair *i* and *j* in period *t*. N_{it} and N_{jt} are the number of patents in technology *i* and *j* at time *t* as in gravity model specifications, and η_t is a period fixed effect. For the count part, we use the corresponding specification:

$$\Pr(y_{ijt}) = \gamma_0 + \gamma_1 \log(y_{ijt-1}) + \gamma_2 COAGGLOM_{ijt} + \gamma_3 COMPLEX\_SUM_{ijt} + \beta_4 COMPLEX\_ABSDIFF_{ijt} + \gamma_5 \log(N_{it}) + \gamma_6 \log(N_{jt}) + \delta_t \quad (4.4),$$

with δ_t being the period fixed effect.

4.4 Results

The results of the regression models are presented in Table 4.1. As outlined above, each model has two parts. Firstly, the zero-part (logit) shows the contribution of the explanatory variables to the existence of technological relatedness (at least one co-occurrence of two technologies on patents). Secondly, the count part presents the explanation for the degree of technological relatedness conditional on the existence of at least one co-occurrence of two technologies on a patent in the relevant period.

The table present three models. Model 1 (REL_FULL) is our main model that represents how relatedness is generally influenced by the co-location and complexity of technologies. Model 2 (REL_SINGLE) provides the results for technological relatedness as observed on the basis of individual invention processes. Model 3 (REL_COLLAB) shows the results for technological relatedness, as resulting from collaborative innovation activities.

Table 4.1 Results of the multivariate gravity models, zero-inflation negative binomial regressions (ZINB) on the strength of technological relatedness

	Technological relatedness		
	FULL (1)	SINGLE (2)	COLLAB (3)
Count model			
COAGGLOM (z-score)	0.102***	0.121***	0.137***
	(0.003)	(0.004)	(0.004)
COMPLEX_SUM	0.040***	0.036***	0.041***
	(0.001)	(0.001)	(0.001)
COMPLEX_ABSDIFF	-0.032***	-0.043***	-0.033***
	(0.002)	(0.002)	(0.002)
REL log lag1	1.015***	1.072***	1.003***
	(0.002)	(0.003)	(0.003)
NO. PATS CPC1 log	0.072***	0.079***	0.083***
	(0.002)	(0.003)	(0.003)
NO. PATS CPC2 log	0.074***	0.078***	0.088***
	(0.002)	(0.003)	(0.002)
Constant	-1.108***	-1.248***	-1.298***
	(0.132)	(0.027)	(0.024)
Period FE	Yes	Yes	Yes
Zero-inflation model			
COAGGLOM (z-score)	0.056***	0.059***	0.066***
	(0.004)	(0.004)	(0.004)
COMPLEX_SUM	-0.051***	-0.030***	-0.056***
	(0.002)	(0.002)	(0.002)
COMPLEX_ABSDIFF	-0.060***	-0.052***	-0.070***
	(0.003)	(0.003)	(0.003)
REL log lag1	3.261***	3.425***	3.541***
	(0.020)	(0.025)	(0.026)
NO. PATS CPC1 log	0.532***	0.577***	0.512***
	(0.004)	(0.006)	(0.005)
NO. PATS CPC2 log	0.471***	0.477***	0.460***
	(0.004)	(0.005)	(0.004)
Constant	-6.835***	-6.925***	-6.110***
	(0.038)	(0.044)	(0.036)
Period FE	Yes	Yes	Yes
Log-likelihood	-66,200	-43,200	-53,790
AIC			
Observations	1,092,636	1,092,636	1,092,636

Note: The regressions are based on all the combinations that have available data for every variable in every period. For an easier interpretation, we changed the +/-signs of coefficients in the zero-inflation part. Standard errors in parentheses, *p<0.1; **p<0.05; ***p<0.01

Based on the logit part of the FULL (1) model specification, we find that the zero- and count model are very similar in terms of coefficients' sign. There is only one instance the two differ. It suggests that the existence of relatedness (zero-model) and its intensification (count model) are driven by more or less the same factors in the same way. Notably, the coefficients do vary in magnitude indicating that some factors vary in importance between these two

situations. Given the large number of observations, all variables are significant. We will therefore focus the interpretation on the coefficients' signs and magnitude.

Our (few) control variables show the expected signs. Previous relatedness (REL lag1) is the strongest positive predictor of subsequent relatedness. That is, technologies that are already related will continue to be so (zero-model) and tend to intensify this relation (count model). This suggests indeed that path-dependent developments are dominant and structural breaks are not the norm. The two mass variables, the number of patents of the considered technologies (NO. PATS) are also significantly positive and their coefficients are relatively large. The finding reflects that technologies characterized by many patents have more opportunities to co-occur on patents and hence are likely to be related (zero-model) and having comparatively larger relatedness values than smaller technologies.

The first key variable is *COAGGLOM* approximating the extent to which concentration of technology 1 tends to appear in the same regions as technology 2. The negative and significant coefficient for *COAGGLOM* suggests that the more two technologies are co-located, the more likely they are related and the higher their degree of relatedness is. This clearly confirms our first hypothesis, co-location and thereby geographical proximity between technologies facilitates relatedness. The estimated coefficient is positively significant in all models. It implies that co-location of technologies increases the chances that they are related at all (null-model) and that co-location tends to increase their relatedness (count-model). This finding has to be seen in the light of many studies on related diversification which show a strong link between co-location and relatedness (Neffke et al., 2011; Kogler et al., 2013). The present study confirm that this process is bi-directional: co-location drives the development of relatedness and vice versa. Knowledge spillovers stimulated by co-location indeed influence (re-)combinatorial invention processes which result in technologies converging in terms of relatedness.

The results on complexity are less straightforward. Notably, we observe a significantly negative coefficient for COMPLEX_ABSDIFF in all models. It suggests that technologies at the opposite spectrum of complexity are less likely to be related (zero-model) and less likely to increase their level of relatedness (count model). Accordingly, technologies at similar levels of complexity are more likely to be related. The coefficient of COMPLEX_SUM is the only one changing signs between the zero-model (negative) and the count-model (positive). The result of the zero-model suggests that at the upper end of the complexity spectrum, technologies are less likely to be complementary to each other. This is not so much the case at the lower end. It implies that less complex technologies are more likely complementary with each other and hence may be combined and become related. This confirms the first part of Hypothesis 2: more complex technologies are less likely related. The result of the count-model shows that technologies at the upper end of the complexity spectrum to be more likely to intensify their relatedness than those at the lower end. The finding confirms the second part of Hypothesis 2: complexity does support the intensification of relatedness, conditional on two technologies having been combined before. Most likely, it is the larger economic rewards associated to the

combination of complex technologies that, in contrast to less complex ones, leads to more efforts being made in such endeavors.

In Hypothesis 3, we argued that the effects of co-location and complexity may differ between invention processes in a collaborative or individual fashion. Against our expectations, we do not find the results to differ dramatically between model 2 (SINGLE) and model 3 (COLLAB). All variables maintain their levels of significance and their coefficients keep their signs. However, we do observe some differences in the magnitude of the coefficients. With one exception, we find the coefficients in case of individual inventions (SINGLE) to be consistently smaller. As a result, *COAGGLOM* and COMPLEX_SUM are of smaller relevance for establishing relatedness (zero-model) and for its intensification (count model) in case of individual inventions. As expected, the importance of co-location for relatedness is stronger in case of collaborative invention processes. The coefficient of co-location is about 8% and 20% larger (zero and count models) in case of collaborative invention (COLLAB) than in those where single inventors conduct the invention (SINGLE).

The results for complexity are less clear. When technologies have been combined before (count models), collaborative R&D is more likely to combine complex technologies than individual inventors. The coefficient of (COMPLEX_SUM) is about 14% larger in the COLLAB than in the SINGLE setting. In this case, the greater complexity and the potential rewards associated to it seem to increase the willingness to form teams. However, this difference between collaborative and non-collaborative patenting is even more pronounced when combining simple and complex technologies (COMPLEX_ABSDIFF). This combination is generally less likely (negative coefficient), although the reluctance towards these combinations is smaller in case of collaborative action. The coefficient of COMPLEX_ABSDIFF in the COLLAB model is about 23% smaller than its counterpart in the SINGLE model. We believe that research teams are better able handling the complexity (of the involved complex technology) and hence, are comparatively less reluctant to approach such situations (large values of COMPLEX_ABSDIFF).

Our findings for the case of two technologies having previously combined (count model) are somewhat different. In this case, complexity always acts as barrier. This barrier is particularly large when combinatorial activities have to be done collaboratively. Again, we believe economic incentives to play a role here. Combining two technologies that have never been combined before involves definitely more risk than developing a technology combination which has been shown to work in the past. Rewards are much harder to prognosticate in these scenarios. Accordingly, our results may reflect that actors refuse to invest more resources (setting up a research team) when the outcome is highly uncertain. Consequently, COMPLEX_SUM obtains the larger negative coefficient in the models of COLLAB than in those of SINGLE with the coefficient in COLLAB being almost 87% higher. Even when just one complex technology is involved (large values of COMPLEX_ABSDIFF), this difference is still very pronounced with the corresponding coefficient being ca. 35% larger.

4.5 Conclusions

This chapter seeks to explain technological relatedness and identify factors contributing to its development. Doing so, we shed light on an issue that received little attention so far in studies on relatedness in economic geography and innovation studies: relatedness is usually treated as independent variable shaping the co-location of technologies (e.g. Kogler et al. 2013; Rigby 2015), rather than a dependent variable that requires explanation itself.

We argue and show empirically that co-location of technologies also enhances relatedness between technologies: the more two technologies are co-located, the more likely they are related and the higher their relatedness. Our study also shows that highly complex technologies are less likely to be related. However, when their complementarity has been confirmed by a successful combination and hence, a minimum level of relatedness has been achieved, complex technologies are more likely to be combined. We argue that economic incentives could be responsible for this finding, as, despite higher risks, combinations of complex technologies promise larger economic rewards. Lastly, we explored the mediating role of collaboration. In line with our expectations, we found collaboration to strengthen the impact of co-location on relatedness, while its mediating role on the impact of complexity is very pronounced in cases of technologies being combined for the first time. In this case, the role of complexity as barrier to relatedness particularly hampers collaborative efforts.

Our exercise has implications for future research. Firstly, our study suggests possible endogeneity issues when studying the interplay between relatedness, co-location and complexity of technologies. For instance, the relation between co-location and relatedness can be considered bi-directional: our study shows that co-location drives the emergence and development of relatedness, while many studies suggest this relationship also works in the other direction (e.g. Hidalgo et al., 2007; Neffke et al., 2011; Rigby, 2015). Changes in relatedness are also likely to influence and shape technological complexity. While interesting (Balland et al., 2015; Broekel, 2015; Balland et al., 2018), we leave this discussion to future studies, as this is beyond the scope of the chapter.

Secondly, technological relatedness can also be seen as a network of technologies, where each technology is a node and every tie represent the strength of relatedness (Hidalgo et al., 2007). So far there is no study that comprehensively examines the (network) evolutionary process behind the technology space (Hidalgo, 2009). Our study also takes the first steps to understand how unrelated technologies become related by time (Desrochers and Leppälä, 2011; Castaldi et al., 2015), or in other words, how unconnected or less connected technologies become tightly connected. Moreover, it requires further investigation whether regions that pioneer the recombination of unrelated technologies and therefore contribute to the evolution of relatedness also thrive economically in the long run (Pinheiro et al., 2018). This key challenge for research needs to be taken up systematically in future studies.

Thirdly, we found that differences in the level of complexity between technologies act as barrier to technological relatedness. There is little discussion on this issue in the literature. It might be that simple and complex technologies represent different underlying knowledge infrastructures and human capital, which restrict their complementarity and consequently their integration. In this sense, complexity also captures fundamental boundaries between technologies. More research is clearly needed in this direction.

Finally, while the chapter provides insights on determinants of relatedness, we only explored the role of few factors. The consideration of further aspects like the role of institutions would surely add to our understanding of the underlying mechanisms. Moreover, our relatedness measure captures technological relatedness through patents, but should be tested with other data and relatedness measures (Boschma, 2017). And the generality of the results could be further improved by tests in other geographical settings than Europe.

4.6 Appendix

4.A Descriptive statistics of the dataset

Table 4.2 Descriptive statistics of the entire dataset

Variable	Number of observations	Mean	Min	Max	Standard Deviation
No. tech pairs	1,458,345	1.905	0	13,373	39.389
Co-occurrence of technologies (SINGLE)	1,458,345	0.547	0	2860	8.934
Co-occurrence of technologies (COLLABORATION)	1,458,345	1.358	0	10,513	31.510
Coagglomeration	1,458,345	-0.0002	-0.199	1.057	0.048
Colocation	1,458,345	0.019	-0.135	1	0.079
Complexity SUM	1,394,145	14.275	0	28.158	4.659

Source: Authors' own construction based on OECD REGPAT 2018
Note: The entire dataset contains all the possible CPC technology combinations for all the periods. For a few technology classes the complexity measure is not available.

4.B Descriptive statistics of the co-agglomeration variable and further robustness checks

As a robustness check we also measure the co-concentration of technology pairs by an index inspired by Porter (2003) and Diodato et al. (2018). It quantifies technologies' co-location as the correlation between their relative technological advantage (RCA) vectors:

$$COLOCATION = corr(RTA_{ir}, RTA_{jr}) \tag{4.5}$$

where RTA_{ir} is the relative overrepresentation of technology *i* in region *r*. The equation of RTA is the following:

$$RTA_{ir} = \frac{\frac{P_{ir}}{\sum_i P_{ir}}}{\frac{\sum_r P_{ir}}{\sum_i \sum_r P_{ir}}} \tag{4.6}$$

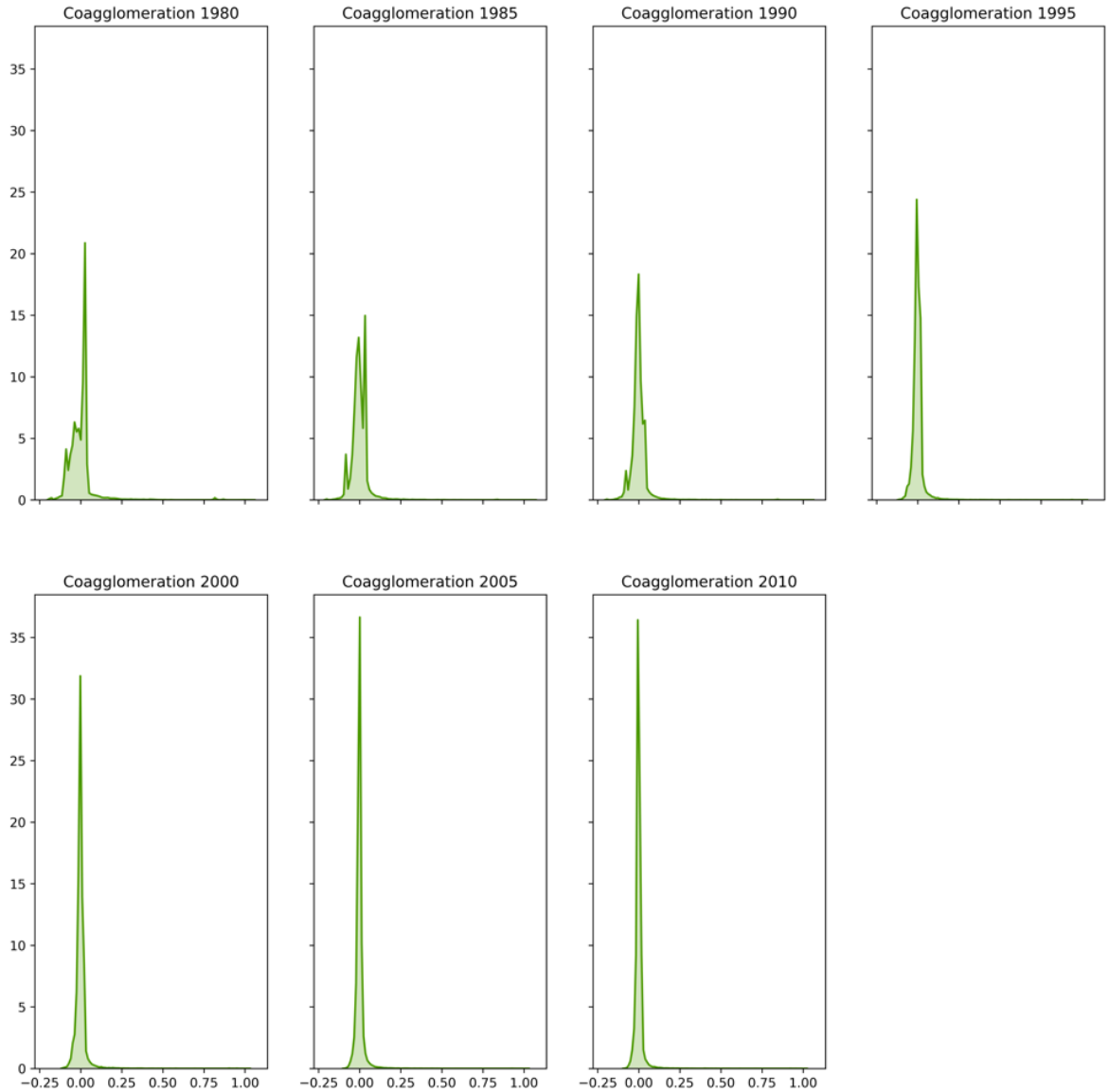
P_{ir} is the number of patents in technology *i* in region *r*.

Table 4.3 The Top 5 most co-agglomerated technologies in every period and their colocation value

CPC1	CPC2	Year	Coagglom	Colocation	CPC1	CPC2	Year	Coagglom	Colocation
G04B	G04C	1980	1.038	1.000	A62B	B31F	1985	1.058	1.000
G04B	G04G	1980	1.038	1.000	A23D	C11C	1985	1.051	1.000
G04C	G04G	1980	1.038	1.000	B25D	D21D	1985	1.050	1.000
D01G	D03D	1980	1.036	1.000	G04F	G04G	1985	1.044	1.000
D05B	F24C	1980	1.031	1.000	B65C	C25F	1985	1.002	1.000
B25C	D02H	1990	1.051	1.000	D05C	E03D	1995	1.021	1.000
A22B	F02G	1990	1.031	1.000	A23B	A23F	1995	1.020	1.000
B25F	F02N	1990	0.997	1.000	B42C	H01C	1995	1.012	1.000
B61G	C12C	1990	0.984	1.000	B44C	C06B	1995	0.994	1.000
B61G	F03D	1990	0.984	1.000	C08C	F16P	1995	0.984	1.000
D03J	C10K	2000	1.022	1.000	C10J	C10K	2005	1.022	1.000
G04D	G04R	2000	1.018	1.000	G04D	G04R	2005	1.011	1.000
B60Y	F01B	2000	1.016	1.000	C09G	F04F	2005	0.979	1.000
B42C	F23H	2000	1.011	1.000	A41B	B09B	2005	0.975	1.000
B42C	G01Q	2000	1.011	1.000	A62D	B03B	2005	0.955	1.000
B63C	F41J	2010	1.017	1.000					
A43D	C12G	2010	1.000	1.000					
D23G	D43L	2010	0.990	1.000					
B27K	F28G	2010	0.982	1.000					
D02J	D06L	2010	0.956	1.000					

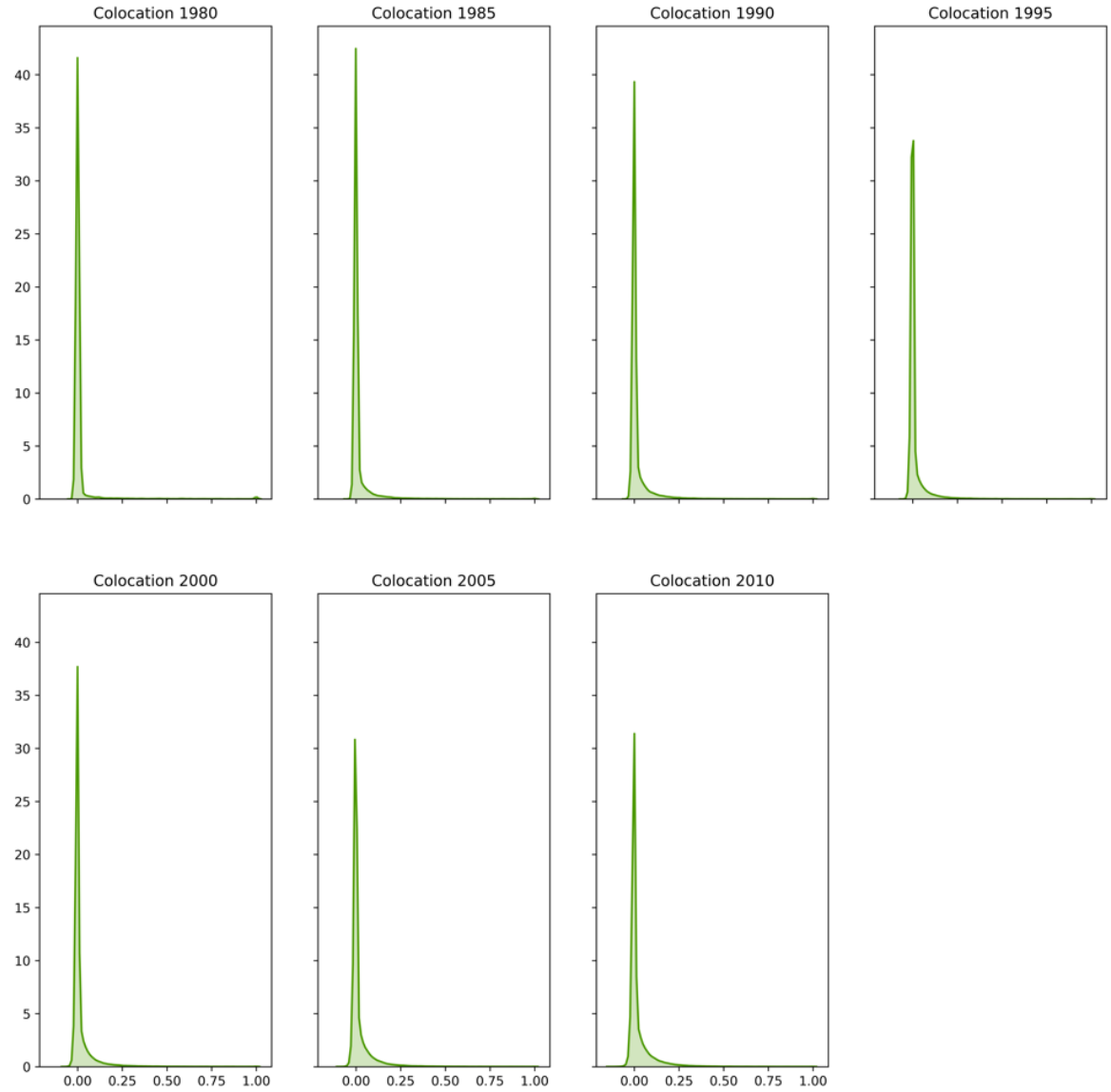
Source: Authors' own construction.

Figure 4.4 Kernel density estimations of the coagglomeration variable for every period



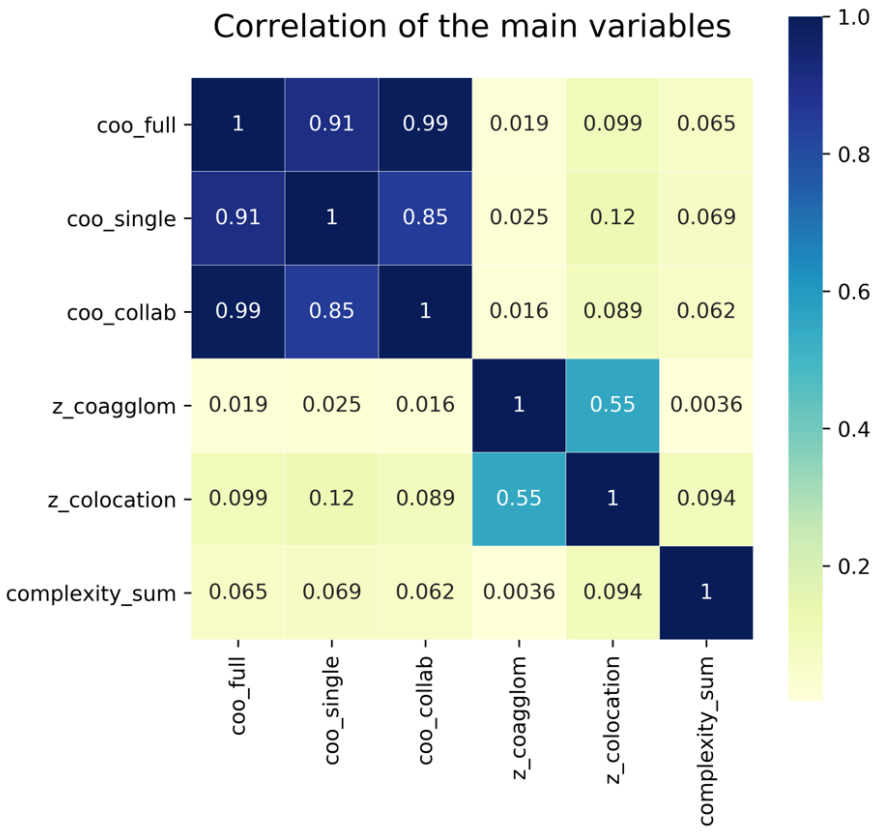
Source: Authors' own construction.

Figure 4.5 Kernel density estimations of the colocation variable for every period



Source: Authors' own construction.

Figure 4.6 Correlation matrix of the most important variables



Source: Authors' own construction.

Chapter 5

Brokering the core with periphery – collaboration networks and individual success in the Budapest film industry

This chapter is adapted from "Brokering the core with periphery – collaboration networks and individual success in the Hungarian film industry", a paper produced in collaboration with Gergő Tóth and Balázs Lengyel. It is currently in a revise-and-resubmit stage for the journal PLOS One. The PhD candidate is the first author of the research paper.

5.1 Introduction

Creative work takes place in teams, where social relations greatly impact the breadth and depth of how collaborators can combine ideas and skills (Caves, 2000; De Dreu and Weingart, 2003). Therefore, a vast literature of network analytic research has looked at the structure of social relations in which individuals and teams are embedded to understand how their position in the social network influences their success (Barabási, 2018b; Cattani and Ferriani, 2008; Fleming et al., 2007; Powell et al., 1996; Uzzi and Spiro, 2005; de Vaan et al., 2015; Vedres, 2017).

A highly reflected concept to explain individual success in cultural fields and creative industries is the core/periphery structure of social relationships (Cattani and Ferriani, 2008; Cattani et al., 2014). A core/periphery structure is characterized by a densely connected cohesive subgroup of core actors and a set of peripheral actors that are only loosely connected to the core (Borgatti and Everett, 1999). The selective core usually represents influential, central actors of a community and their position is often associated with privileges, prestige and control (Clauset et al., 2015; Morgan et al., 2018). The periphery on the other hand consists of a wider variety of actors who are likely to develop alternative views with non-canonical ideas to voice their relevance to the field (Cattani et al., 2014) and have arguably less collaboration commitments but more free capacities (Csermely et al., 2013). Studies on creative industries with a core/periphery approach demonstrated that an intermediate position between the cohesive core and the less connected periphery is favourable for creative achievements due to a threshold between recognition and novelty (Cattani and Ferriani, 2008). On the one hand, being close to the core might help the recognition of creative works. One the other hand, the pressure of dense schedules and to conform to established norms of the field is less on peripheral actors who can thus can come up with fresh ideas.

The importance of intermediate positions in social networks is reflected in the structural hole literature as well and are often associated with brokerage. By bridging otherwise disconnected parts of the network, brokers mediate and control the flow of ideas, information and knowledge (Burt, 2004; 2005; Granovetter, 1973; Fleming et al., 2007; Zaheer and Soda, 2009) and control the use of resources in the network. Consequently, broker position is thought to provide network advantages that enhance individual outcomes. More recently, edge-betweenness have been claimed to better identify brokers in networks than bridging structural holes because it reflects on the individual's control over flows in the entire network (Everett and Valente, 2016). Despite the core/periphery structure has attracted recent attention in various domains (e.g. Bastos et al., 2018; Crespo et al., 2014; Csermely et al., 2013;

Kojaku and Masuda, 2018), the question how brokerage in core/periphery networks influence individual success is surprisingly under-researched.

In this chapter, we propose that creators are more likely to achieve individual success if they belong to the core of the collaboration network that characterizes the industry and gain additional advantages if they are in broker position at the same time. We further argue that these individuals are even more likely to achieve success if they broker the core with the periphery of the network. This case individual brokers enjoy the advantages of being in the core and at the same time, have access to new ideas and free capacities from the periphery.

The context of our empirical exercise is provided by the Budapest film industry, which is a suitable example of a project-based industry heavily relying on localized networks of a broad range of actors (Bielby and Bielby, 1999; Caves, 2000; De Vaan et al., 2013). It also improved its international reputation recently by two American Academy Awards in subsequent years and have collected various other prestigious honours in Europe including Golden Palm and Golden Bear. In order to explain how the structure of social relations influence success in terms of award winning, we constructed time-varying collaboration networks of movie creators from 1990 to 2009. This unique dataset allows us to test how core/periphery position and brokerage influence the likelihood of award winning on the Hungarian Film Week, the major festival for new movies in the country. Results suggest that being in the core and being a broker of the network both increase the individuals' chance to win award. Besides the implementation of betweenness based brokerage measure by Everett and Valente (2016), we construct a new way to capture the role of brokers in core/periphery networks. Our new Gatekeeping index identifies those creators who act as bridges between core and peripheral nodes in their ego network. Combining brokerage with Gatekeeping, we find evidence that those core brokers who act as gatekeepers and connect the core with the periphery enjoy additional likelihood of award winning.

The chapter is organized as follows. In section 5.2, following a brief literature review, we present our hypothesis on how core/periphery structure and brokerage influence success in creative industries. Next, we describe the context and the collected data and provide descriptive statistics of the longitudinal relational dataset. The applied statistical model and variables are introduced in Section 5.4. The final section includes a discussion of main findings and implications; details the limitations of the study and outlines some possible future research issues.

5.2 Brokers in core/periphery collaboration networks

Individual outcomes have been associated with the ability to access and mobilize resources through social relations (Coleman, 1988; Lin, 2002; Putnam, 2001). Therefore, structural patterns of social networks are fundamental to understand individual success. Two of the most reflected theories in this regard are the core/periphery structure of social networks

(Borgatti and Everett, 1999) and the theory of social network brokerage (Burt, 1992; Everett and Valente, 2016). In this section, we argue that an approach combining these two central tenets will provide us with new details on how social capital induces individual success.

Since the seminal work of Borgatti and Everett (1999), the tendency that collaboration networks form a core/periphery structure has become a common understanding in social network analysis. In this central tenet, networks self-organize into a densely connected core of central nodes, while the periphery refers to a sparsely connected, usually non-central set of nodes that are loosely linked to the core (Csermely et al., 2013). Core/periphery structure usually emerges in networks of creative production that depends on intensive project-based collaboration (Faulkner, 1983; Jones, 1996; Giuffré, 1999; Cattani et al., 2014). Consequently, core has been separated from periphery in several real-world social networks such as academic networks (Clauset et al., 2015), online communities of musicians (Dahlander and Frederiksen, 2012), industry clusters of wineries (Giuliani and Bell, 2005), or the feature film industry of Hollywood (Cattani and Ferriani, 2008; Cattani et al., 2014).

It is often argued that individuals located in the core of creative production enjoy the social capital concentrating in a relatively small number of established players with the necessary material resources, political influence and social relations to enforce their central role in the creation of cultural and creative products (Anheier et al., 1995; Fraiberger et al., 2018). Core network position correlates with prestige by signalling legitimacy, experience and credibility (Peterson and Anand, 2004; Clauset et al., 2015; Morgan et al., 2018) and can facilitate the completion and legitimation of artistic or creative work. Creators in the core have better opportunities for innovation and outstanding performance because of superior access to information (Borgatti and Everett, 1999). The close proximity of the core to the entire network implies that ideas originating in the high-prestige core spreads faster and easier throughout the field (Abramson and Rosenkopf, 1997), whereas ideas originating from the lower prestige peripheral creators must filter through many more intermediaries (Clauset et al., 2015; Morgan et al., 2018).

Information access and control have central importance in the theory of network brokerage as well (Burt, 1992). Brokers use weak ties to access diverse information that can be combined in novel ways (Aral, 2016; Burt, 2004; Granovetter, 1973); and they bridge otherwise disconnected parts of social networks, which enables brokers to mobilize social capital. Moreover, brokers can consider new ideas and opportunities earlier and easier as they have connections to a wider set of groups (Fleming and Waguespack, 2007). More recently, Everett and Valente (2016) argued that brokerage means control over bridging ties in the network and they emphasized that a node's brokerage is a function of the bridging scores of the edges it is incident to. They suggested betweenness-based measures to capture brokerage by considering the entire network structure, not just a node's local environment. Nevertheless, the widely accepted common understanding is that brokering over ideas, information and resources induce individual success (Burt, 2004; Granovetter, 1973; Fleming et al., 2007). However, the very detailed elaboration of the above theories over the last

decades left an open question: how do brokerage and core/periphery position jointly influence nodal outcomes and individual success?

In the first step of establishing such a framework, we propose that being in the core of the network and being a broker of the network provides different benefits and thus are complementary sources of success. While core positions yield recognition of the community in general and therefore indicate the ability of individuals to mobilize social capital associated with elite groups (Anheier et al., 1995; Cattani et al., 2014), brokerage provides diversity and enables one to organize social capital embedded in more than one and relatively distinct groups (Burt, 2004).

Therefore, it is a plausible expectation that core and broker positions help individual outcomes through separate mechanisms in the network. Consequently, one can expect positive association between both coreness and brokerage and the measures of individual success in a common empirical model. Further, we can also posit that individuals gain extra probability in case they are in the core and are brokers at the same time, since they have advantage to combine distinct ideas into creative outputs and will be likely to be recognized for doing this. To articulate this point, we formulate a hypothesis for empirical research:

Hypothesis 4: The probability of individual success is increased by both core membership and brokerage and these network effects reinforce each other.

Although it is not in our explicit focus, Hypothesis 4 implicitly refers to an important feature of creative project-based and focused work, in which a right mix of cohesion and diversity is needed (Uzzi and Spiro, 2005). The core, which is cohesive by definition, arguably provides effective exploitation of ideas due to the increased willingness of cooperating and sharing through strong connections (Aral and van Alstyne, 2011; Obstfeld, 2005; Reagens and McEvily, 2003). Brokers in the core therefore pair cohesiveness with diversity and thus their good ideas that are better due to their broker status (Burt, 2004) can be better exploited due to their core status.

Following this logic, a further question naturally emerges: brokerage between which groups yield success? Is it brokerage in the core community, which is theoretically possible because network cores are usually far from being fully connected graphs (Borgatti and Everett, 1999; Csermely et al., 2013), or is it brokerage between the core and periphery of the network?

To motivate the relevance of this question, we refer to a recent literature that argues for the importance of intermediate core/periphery position in the network. Cattani and Ferriani (2008) demonstrated that individuals located between the core and the periphery of the social network are in most favourable position to achieve creative results because connections to the core members provides advantages of prestige and recognition; while connections to the wider periphery give access to radically new and non-confirmative ideas as well. According to their argument, central actors can become too entrenched in the

prevailing conventions of a community and thus tend to ignore the potential contributions of new ideas and knowledge from outside (Schilling, 2005; Dahlander and Frederiksen, 2012; Cattani and Ferriani, 2008). In other words, a stable core could lead to a lock-in situation and result the circulation of common, rather than novel ideas, information and knowledge. In contrast, creators from the periphery might lack the necessary endorsement and visibility to the recognition of their work, but they are more likely to bring fresh, novel ideas to the system. Consequently, those core brokers who channel in ideas from the periphery to the core are arguably in a better position than those brokers that bridge core with other core groups.

Besides channelling information and ideas, brokers can take advantage of mobilizing resources in the network, such as free capacities for work. This is important because creators share their time among partners and benefit more from a collaboration if their partner focuses on their shared project and not on other collaborations (Jackson and Wolinsky, 1996). In this logic, core creators do not have much free capacities since they are involved in dense collaboration networks and the amount of network ties they can manage is limited (König et al., 2010; Stiller and Dunbar, 2007). Therefore, peripheral creators might have more working capacities available since they are involved in less projects and favour collaboration with central creators (Barabási et al., 2002; Hojman and Szeidl, 2008). Consequently, broker creators who bridge the core and the periphery control more capacities than those creators who broker core groups with each other.

In sum, those core brokers who bridge the core and the periphery have access to core advantages, such as prestige and credibility, and at the same time, have the ability to link new creators with new ideas from the periphery and use their capacities. Our line of argument leads us to formulate the following hypothesis:

Hypothesis 5: The probability of individual success is higher for the members of the core who also broker the core and the periphery.

5.3 Context and data

5.3.1 The Budapest film industry

The Hungarian film production is in many ways similar to other continental European film scenes. The production of feature films depends heavily on public founding, which is a significant difference to the Hollywood film industry. As most of the Hungarian feature films are produced in Hungarian language, it results a small and relatively closed market. However, there are several film studios, film schools and celebrated award winner filmmakers with worldwide reputation. Moreover, Budapest has become a favoured place for international film production with a great supply of film studios around the city. This created a learning

opportunity for many Hungarian movie creators who get experienced by working for foreign movie projects. Economic activities in relation to film production also concentrate heavily around Budapest. Specifically, 70% of all the companies (around 2000 firms) operating under the industry code 5911. Motion picture, video and television programme production activities (NACE Rev 2) are located in the urban agglomeration of Budapest[14]. In spatial concentrations of project-based creative industries like the film production around Budapest, the exploitation of locally concentrated knowledge mainly happens through participation in social networks (Scott, 2000; Grabher, 2002; Sydow and Staber, 2002; Ibert, 2004; Perretti and Negro, 2007).

Outstanding Hungarian movies used to be celebrated during the Hungarian Film Week in Budapest, which was the most prominent film festival in the country up till the close past. It was established in 1965 and with few years of hiatus during the 1970s it was a central event in the national film production for decades. To the 1990s the festivals program became similar to many other international film festivals and besides the main prize, directors, cinematographers, editors and writers were awarded in separate categories. Due to a drastic change in the financial supporting system of Hungarian movie production, the industry totally stopped in years 2012 and 2013. Even though Hungarian movies won numerous international awards in the 2013-2018 years including the Best Foreign Language Film at the 88th Academy Awards, Best Live Action Short Film at the 89th Academy Awards and Golden Bear in the main competition section of the 67th Berlin International Film Festival, the 43rd Hungarian Film Week in 2012 was the last festival organized to celebrate Hungarian movies.

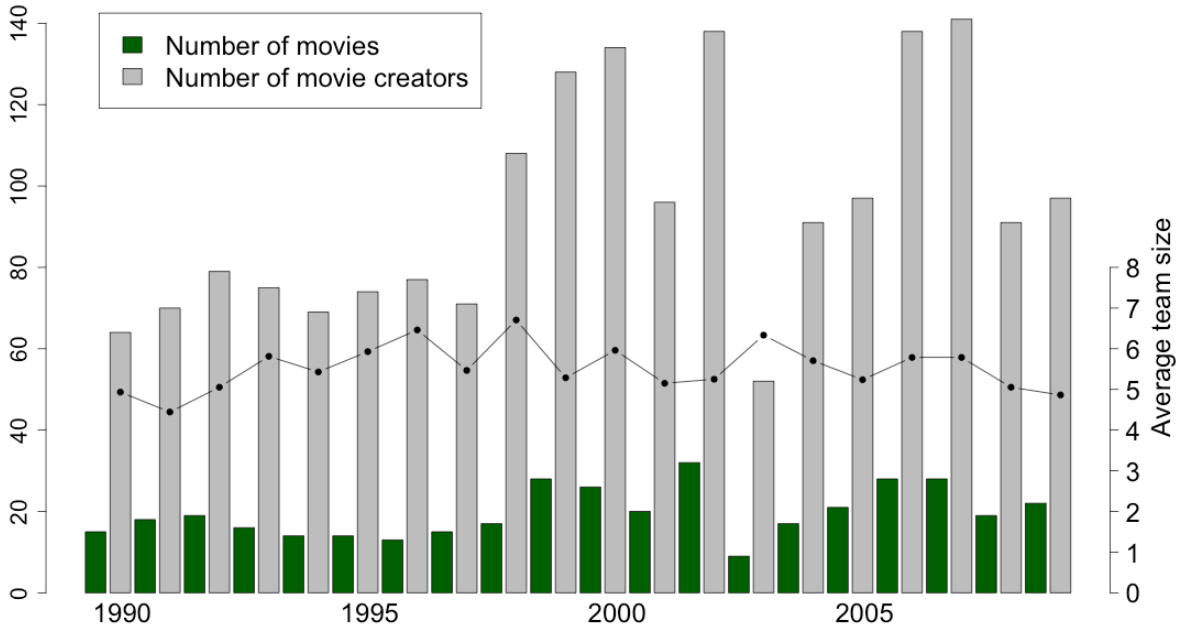
5.3.2 Data and network construction

Our main data source is the online database of the Hungarian Film Archive[15]. It contains detailed information on the title, production year, name and role of movie creators, production companies and cast members for every feature film submitted to the national archive between 1912 and 2011. We use this dataset to reconstruct the collaboration networks of movie creators in retrospection. Additionally, we use the Hungarian film yearbooks to collect data on awards[16]. Even though data is available on movies in a nearly 100-year time span, we focus our analysis on the 1990-2009 period because detailed and continuous data on Hungarian Film Week awards are only available for this period[17]. Despite the political and economic changes in the early 90s, the selected interval was a relatively productive period of the Hungarian film industry.

[14] To look at the distribution of firms in Hungary we used the OPTEN Company Database, hhtp://opten.hu
[15] The online database of the Hungarian Film Archive is available on http://mandarchiv.hu/tart/jatekfilm
[16] The Hungarian film yearbooks are available online on http://mandarchiv.hu/cikk/4643/Filmevkonyv
These yearbooks provide detailed information on Hungarian movies produced in the given year, festival attendance, awards, number of attendance and financial revenues in between 1979-2010.
[17] Distribution of produced movies in 1912-2011 can be found in the Appendix (Table 5.4).

Since movie creation is project based, film industry's collaboration networks are constantly created and re-created as individuals collaborate on a specific project, disband when the project ends, and then combine for a new project, often with new partners (Jones, 1996). We analyse the unipartite projection of the bipartite affiliation network of movie creators and movies. We construct networks of individuals in which a link between any two movie creators indicates collaboration on a movie. In a similar fashion to Cattani and Ferriani (2008), we considered movie creators as the following members of the production crew: cinematographer, director, editor, producer and writer. Figure 5.1 shows that the number of movies and the number of movie creators are very much varying along the 1990-2009 period, however, the average team size of movies is relative stable.

Figure 5.1 Collaboration patterns in movie production



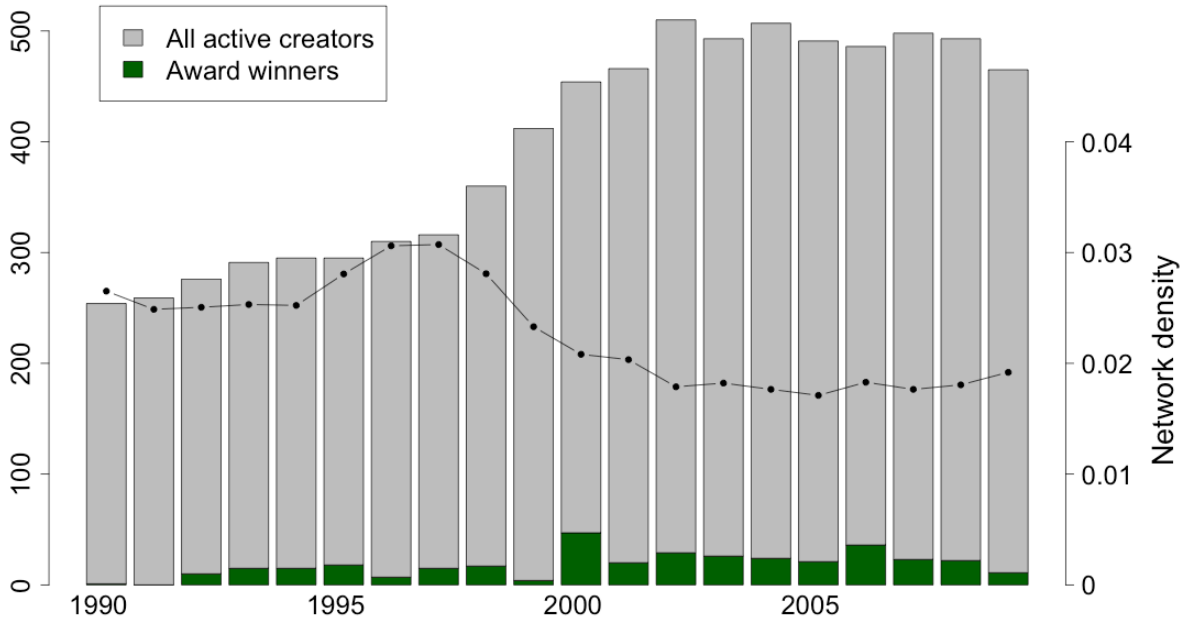
Source: Author's own construction based on the Hungarian National Film Archive.

We assume that the relationships formed through the production of a movie last for 7 years, similarly to other network studies related to project-based creative production (e.g. Uzzi and Spiro, 2005; Cattani and Ferriani, 2008)[18]. Thus, the adjacency matrix of the collaboration networks for a given year records ties formed in that year and in any of the previous six years. Therefore, each network observation covers the co-production linkages of a 7-year time interval, covering the period of 1984-2009. As an example, the network of 2000 is based on all the movies created during the 1994-2000 period, while the network of 2001 represents collaborations on movies in the 1995-2001 period. Figure 5.2 presents the number

[18] As a robustness check we also run the same analysis with 5-year intervals and we have similar results (see Appendix Table 5.16).

of active movie creators per year based on 7-year moving windows[19]. As the number of active individual movie creators nearly doubled from early 1990s to early 2000s, the collaboration network became even more sparse.

Figure 5.2 Active creators, award winners and network density



Source: Author's own construction based on the Hungarian National Film Archive.
Note: Number of active movie creators are based on 7-year moving-windows and represent our final sample. Award winners are also part of the active creator group.

5.4 Variables and methodology

5.4.1 Dependent variable – success in film industry

We use data on award winning in the Hungarian Film Week to operationalize individual success; which is a fairly established practice to measure individual creative performance in the tradition of creativity research (Simonton, 2004; 2009). During the Hungarian Film Week, movie creators were awarded for the quality of their performance by a selected jury of film critics, producers and industry peers. Besides the main prize for the best feature film, individual achievements were also awarded in separate professional categories (e.g. best cinematographer, editor, director).

There are several reasons for looking at awards won at the Hungarian Film Week only and analysing only awards as indicator of success. First, even though a lot of Hungarian movies

[19] The exact number of movies and movie creators per year we establish our networks on can be found in the Appendix (Table 5.5).

competed at international film festivals, there is a remarkable variance in the prestige of these events, which makes it difficult to compare success achieved at major festivals and at small ones. Second, other success indicators from the Hungarian Film Week data, such as being nominated for an award, are not available for every year. Third, other possible success indicators in other data, such as the attendance, number of screenings or Internet Movie Database (IMDb) ratings are not available for every movie.

The dependent variable is award winning, which equals 1 in a given year if the movie creator won an award in an individual category (e.g. best cinematography or best editing) and zero otherwise. In case the movie won the best movie prize, every creator is marked by a dummy in our dataset. Since we are focusing on individual success rather than team success, we tested our models in two alternative settings: by removing best picture award winners and leaving them in the sample. Because the coefficient signs were identical and effect sizes were significantly not different across the two settings, we decide to report results taken from the full network. Our final sample contains 361 award winner movie creators out of 8401 individuals along the period of 1990-2009. The share of award winner individuals in all the active creators per year are also presented in Figure 5.2.

5.4.2 Independent variables – coreness and brokerage

The main explanatory variables in our empirical framework are core/periphery position and brokerage in the movie creator network. We measure all individual movie creators' positions on a core/periphery continuum using Borgatti and Everett's (1999) algorithm[20]. Following the approach of Cattani and Ferriani (2008) and Dahlander and Frederiksen (2012), we estimate each nodes degree of coreness by the continuous measure of core/periphery. As a result, our *Coreness* indicator refers to the degree of closeness to a densely connected network core for each movie creator. The indicator is high for core creators, low for peripheral creators and have a medium value for creators in an intermediate position. The computation of the index is made by the built-in continuous core/periphery procedure in UCINET VI (Borgatti et al., 2002). We apply the described procedure to all network matrices of years 1990-2009.

As one unit change in *Coreness* represents a shift from the absolute periphery (where the value is 0) to the absolute core (value is 1), one unit change of the variable would be difficult to interpret in a regression framework. Therefore, we standardized *Coreness* into *z*-score with zero mean by the formula $z(x) = (x - \bar{x})/sd(x)$, where $\bar{x}$ is the mean of x, and $sd(x)$ is the standard deviation of x.

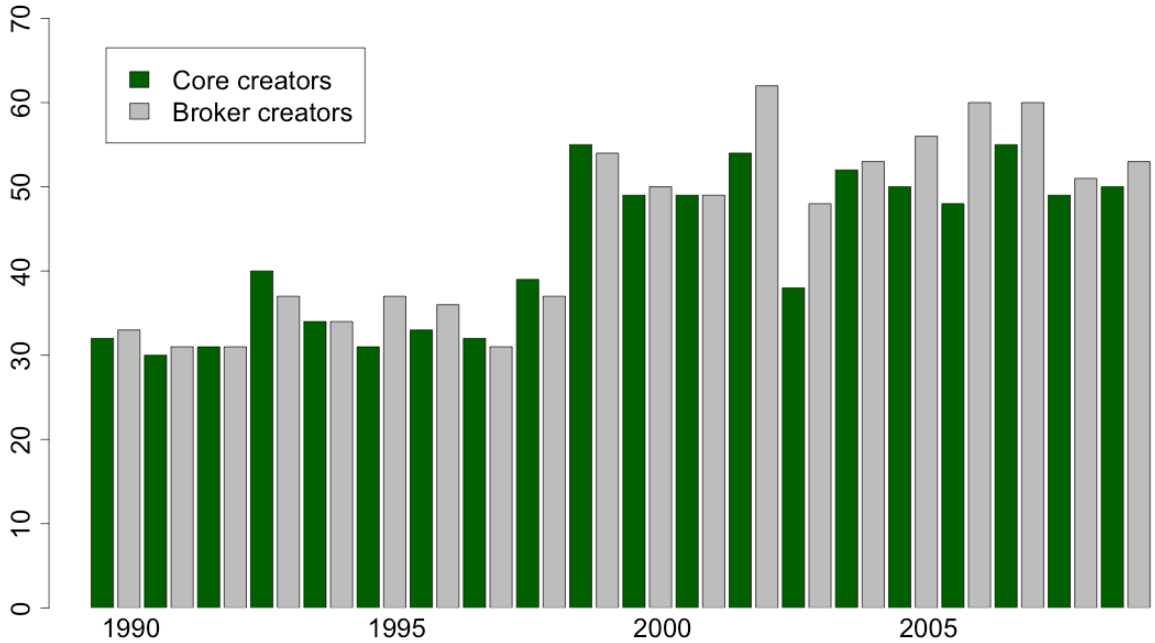
Brokerage is the other key variable in our empirical setting. We use the recently suggested betweenness measure by Everett and Valente (2016). We decide to apply this new

[20] As a robustness check, we tested the k-core indicator, an alternative measure for network coreness. Because results were similar, we decided to report on the core/periphery indicator that is more commonly used in the literature. K-core based results can be found in the Appendix (Table 5.9).

measure instead of network constraint (Burt, 1992) because we aim to identify brokerage by considering the full network structure and will zoom into ego-networks only when we qualify brokers' bridging positions between core and periphery. However, as a robustness check we tested Burt's constraint measure as a brokerage indicator and got similar results[21]. Further discussion on other possible measures can be found in the Conclusions and discussion section.

The measure is computed in two steps. First, we calculate the edge betweenness centrality measure for every tie in the network. Second, for each node assign a brokerage score which is the average of the edge centralities which are incident to the focal node. The indicator takes high value if the focal actor has several ties that are part of many shortest paths. Consequently, a positive correlation between Brokerage and the dependent variable would imply that brokers are more likely to succeed. For easier comparison we also standardized *Brokerage* into *z*-score with zero mean.

Figure 5.3 Change in the number of core and broker creators



Source: Author's own construction based on the Hungarian National Film Archive.
Note: Number of movie creators are based on 7-year moving-windows. The number of core and broker creators are based on the *Core* and *Broker* dummy variables.

For a more detailed understanding on how the joint effect of core/periphery position and brokerage influence actors' individual success, we created dummy variables from the *Coreness* and *Brokerage* indices. A similar practice was applied in Cattani and Ferriani (2008) to indicate whether movie creators were in the core of the network. The dummy variable is called *Core* that takes the value 1 in case the continuous *Coreness* value of the individual is in

[21] As a robustness check, we tested constraint as an alternative measure for network brokerage. We reported the main result in the Appendix (Table 5.10).

the top 10 percentile (above 0.90) of the measure's scale and zero otherwise. The *Broker* dummy takes the value of 1 in case *Brokerage* measured by the described betweenness indicator is in the top 10 percentile (above 0.90) of the measure's scale. Figure 5.3 presents the change in the number of *Core* and *Broker* creators along the examined period. Based on the applied dummy variables, the number of Core and Broker movie creators are nearly identical in every point in time[22].

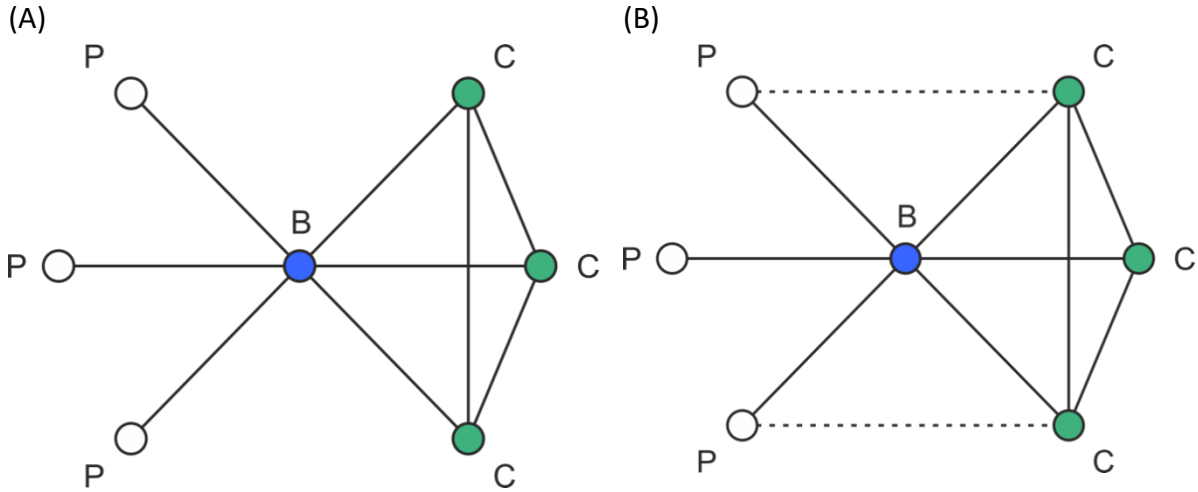
To understand how brokerage induces success in core/periphery networks, we also aim to identify those brokers who bridge core with periphery in the network. To do this, we introduce a new measure we call *Gatekeeping*, which refers to how much a node act as a bridge between the core and the periphery in its ego network. The measure partly builds on the 'gatekeepers of knowledge' concept by Allen (1977). It suggests that gatekeepers are a small group of individuals in the core of an information network, with access to external sources of information that allow them to control over the diffusion of new knowledge in the network. Equation (5.1) summarises the construction of our *Gatekeeping* measure.

$$\mathrm{Gatekeeping_i} = 1 - \frac{\mathrm{L_{cp}+1}}{\mathrm{||v_c|| \times ||v_p||+1}}, \tag{5.1}$$

where L_{cp} refers to the remaining number of links between core and peripheral actors in the ego network of *i*, without the focal actor. In the denominator, $\|v_c\|$ refers to the number of core individuals in the ego network of creator *i* and $\|v_p\|$ accordingly for peripheral nodes. As a result, the indicator is the inverse of the observable ties between core and peripheral actors compared to all the possible ties between the two modules in the ego. Figure 5.4 shows two hypothetical cases. In case of Figure 5.4A gatekeeping indicator has a relatively high value as there are both core and peripheral nodes in the ego network, but the focal actor is the only connection between core and peripheral nodes. From this example one can expect that the focal creator will enjoy the benefits of bridging structural holes across core and periphery. In case of Figure 5.4B gatekeeping indicator has a relatively low value as there are more ties in the ego network between core and periphery in which the focal actor is not involved. One cannot expect that the ego will benefit from connecting the core with periphery because core neighbours also have direct access to additional advantages residing at peripheral collaborators.

[22] We also tested the Core and Broker dummies with cut off points at 0.75 percentiles and our main results did not change significantly. The related model outputs can be found in the Appendix (Table 5.12).

Figure 5.4 Illustration of high (A) and low (B) *Gatekeeping* indicators



Source: Author's own construction.

The *Gatekeeper* indicator identifies those movie creators who connect the core and periphery of the network. This is a dummy variable that takes value 1 in case the value of Gatekeeping indicator is in the top 10 percentile (above 0.90) of the measure's scale and 0 otherwise. Combined with core and broker dummies, this variable allows us to test how the network position connecting the core and the periphery influence success of core and broker individuals. Moreover, by the combination of brokerage on the global network level and gatekeeping on the ego level between core and peripheral nodes we might get a better understanding on how brokerage works in core/periphery networks.

5.4.3 Control variables

Since creative projects are thought to be most successful when the underlying collaboration network has the optimal level of small-world property (Uzzi and Spiro, 2005), we control for those two variables that constitute small world networks. First, *Clustering* is the fraction of closed triangles in our observed network and closed triangles in a random graph with similar properties. Second, *Average Path Length* is the fraction of average path length in the observed network and the average path length of a random graph with similar properties.

Because of the applied 7-year moving-windows to create more stable and connected creator networks, we also control for the number of *Films Per Window* that the given network structure is based on. Moreover, we used the variable *Creators Per Window* to control for the number of active movie creators in the 7-year period or the number of nodes in the network in the given year.

Professionals new to the industry might receive disproportional attention from award voters, as they might prefer new talents over old veterans (Cattani and Ferriani, 2008). To

account for this effect, for each movie creator in every year a *Newcomer* dummy variable was created that takes 1 if a professional is a new participant of movie production and 0 otherwise. Finally, we applied year fixed effects and individual role fixed effect as well, which refers to the main role of the creator in the period being classified to cinematographer, director, editor, producer or writer categories.

5.4.4 Model construction

Let us assume that creative performance is a latent continuous unobserved variable, which is related to individual success measured by awards. Because of the dependent variable y_i can take only two values, we apply a binary logistic specification. The dependent variable takes two values, 0 when the event of award winning has not occurred and 1 if creative performance yielded an award.

Therefore, we estimate a pooled, logistic regression model with year and role fixed effects and we cluster standard errors at the creator level. The empirical model is defined in the following:

$$Pr(y_{it} = 1) = \alpha + \beta_1 Coreness_{it} + \beta_2 Brokerage_{it} + \beta_3 Gatekeeping_{it} + \beta_4 N_t + \beta_5 Z_{it} + \varphi_t + \omega_i + u_{it} \quad (5.2)$$

, where *Coreness* (and its squared term to test the non-linear effect), *Brokerage*, and *Gatekeeping* denote the network characteristics of creator *i* at year t. N_t stands for the network structure variables controlling for *Average Path Length* and *Clustering* of the creator network in the given year. Z_{it} is for the collection of control variables. φ_t is a year fixed effect and ω_i is a categorical fixed effect for the main role of creators' in movie production.

In the second part of the analyses, we run logistic regressions on the interactions of the dichotomized *Core*, *Broker*, and *Gatekeeper* variables to disentangle these effects on individual success. In our last model, we introduce a three-way interaction to capture the creators' role as a broker in both global and local core/periphery setting:

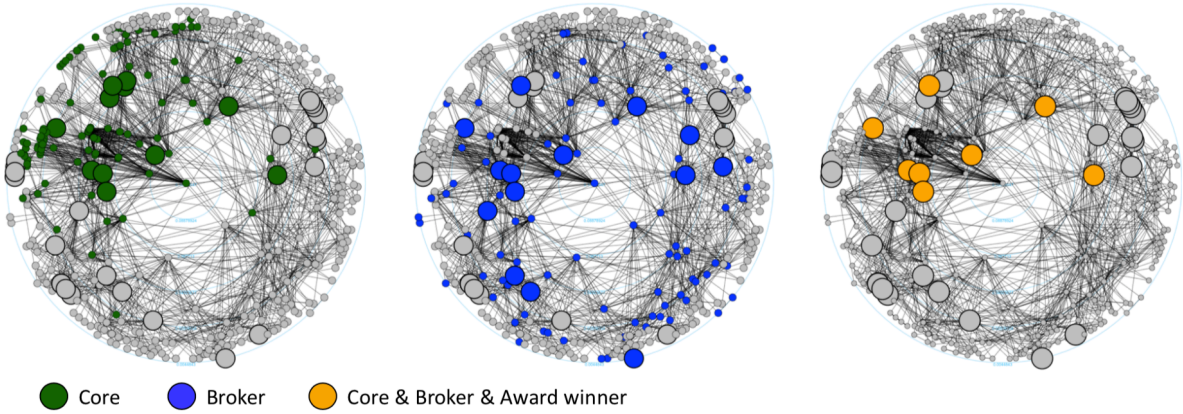
$$Pr(y_{it} = 1) = \alpha + \beta_1 Core_{it} + \beta_2 Broker_{it} + \beta_3 Gatekeeper_{it} + \beta_4 (Core_{it} \times Broker_{it}) + \beta_5 (Core_{it} \times Gatekeeper_{it}) + \beta_6 (Broker_{it} \times Gatekeeper_{it}) + \beta_7 (Core_{it} \times Broker_{it} \times Gatekeeper_{it}) + \beta_8 N_t + \beta_9 Z_{it} + \varphi_t + \omega_i + u_{it} \quad (5.3)$$

, where besides the introduced dichotomized variables we use the same model setting.

5.5 Results

Figure 5.5 illustrates and summarises the basic idea behind this study through the example of the 2006 creator network. Intuitively, we can see that non-core and non-broker creators also won awards, but individual creators are more likely to win, if they are both in the core and are brokers at the same time.

Figure 5.5 Representation of core, broker and award winner creators in the network of 2006



Source: Author's own construction based on the Hungarian National Film Archive.
Note: Bigger nodes represent award winners in all three graphs. Nodes with higher degree centrality are closer to the center of the circular layout. The network is based on a 7-year moving window. To make the visualization more intuitive, we opt to use the 0.75 cutoff on Coreness and Brokerage to identify Core and Broker creators.

Table 5.1 presents the absolute numbers and the share of award winners in the different creator groups underpinning our intuitions detailed above. Around 4% (361 from 8401) of creators won an award. Both award winning creators in core (13%) or broker (10%) positions are over represented compared to non-core and non-broker award winners. Moreover, from the 213 creators who occupy the special position as brokers who are also part of the core, 41 (19%) won an award.

Table 5.1 Number of core, broker and award winner creators in the final sample

	Award winner (n=361)	No winners (n=8040)	Full sample (n=8401)	Share of award winners (4%)
Non-core & Non-broker	156	5820	5976	3%
Core	107	717	824	13%
Broker	85	751	836	10%
Core X Broker	41	172	213	19%

Source: Author's own construction based on the Hungarian National Film Archive.
Note: Number of movie creators are based on 7-year moving-windows. The number of core and broker creators are based on the *Core* and *Broker* dummy variables. Because of the logit specification, data for 1991 was removed from the regression.

Table 5.2 presents the coefficients from four different logit regression models that test our first hypothesis. Control variables are included in all models and we introduce explanatory variables in a stepwise manner. In model (1) we introduced both linear and quadratic forms of Coreness into the estimation. The coefficient of *Coreness* (z-score) is significant and positive, while the quadratic term negatively correlates with the dependent variable, suggesting that creators are more likely to receive an award for their contribution to a movie in a given year when they are closer to the core of the network, however, it has diminishing returns.

We introduce *Brokerage* measured by the betweenness indicator in model (2) without controlling for *Coreness*. The positive and significant coefficient suggests that *Brokerage* significantly induces creators' likelihood of award winning. Most importantly, neither the significance level nor the direction of correlation changes for *Coreness* and *Brokerage* variables in model (3). Already in line with *Hypothesis* 4, which will be further looked at in Table 5.3, results reported in model (3) suggest that creators are more likely to receive an award when they are closer to the core and their broker status further increases the probability of award winning. Consequently, these findings support the idea that being part of the network core and brokering the network provide complementary benefits for creative workers.

In model (4) we test the importance of our *Gatekeeping* variable on award winning, independently from *Coreness* and *Brokerage*, while in model (5) we include *Gatekeeping* to our extended model. The coefficients in the final model version show similar results and *Gatekeeping* has positive and significant effect too. It indicates that both brokerage on the global network level and gatekeeping in the ego network influences individual success in core/periphery networks.

Table 5.2 Network position and individual success – results of logit regressions

	Dependent variable – award winning				
	(1)	(2)	(3)	(4)	(5)
Coreness	0.860***		0.884***		0.569***
(z-score)	(0.089)		(0.080)		(0.090)
Coreness^2	-0.085***		-0.091***		-0.033**
(z-score)	(0.021)		(0.017)		(0.014)
Brokerage		0.466***	0.446***		0.124*
(z-score)		(0.051)	(0.041)		(0.073)
Gatekeeping				0.885***	0.670***
(z-score)				(0.056)	(0.083)
Clustering ratio	-3.106	-2.301	-1.866	-4.251	-2.255
	(24.034)	(24.070)	(24.113)	(24.733)	(24.964)
Path length ratio	22.062	20.915	21.558	25.669	24.561
	(23.690)	(23.483)	(23.722)	(24.249)	(24.558)
Creators Per Window	0.014	0.009	0.013	0.018	0.020
	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
Films Per Window	0.042	0.063	0.050	0.031	0.026
	(0.101)	(0.102)	(0.102)	(0.105)	(0.106)
Newcomer (dummy)	2.080***	2.141***	2.432***	2.749***	2.853***
	(0.137)	(0.134)	(0.141)	(0.153)	(0.146)
Constant	-25.178*	-26.254*	-27.129*	-26.681*	-28.364*
	(14.851)	(14.963)	(14.911)	(15.397)	(15.481)
YEAR FE	Yes	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes	Yes
BIC	2707.371	2755.513	2618.668	2593.761	2532.006
Log likelihood	-1237.396	-1265.939	-1188.572	-1185.064	-1140.766
N	7672	7672	7672	7672	7672

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01

Standard Errors are clustered at the creator level. The descriptive statistics (Table 5.6) and correlation matrix (Figure 5.7) of our main variables can be found in the Appendix. VIF values are also provided in Appendix Table 5.8.

Bayesian information criterion (BIC) is used to compare our models. Looking on the BIC statistics of the stepwise specifications, the model selection criteria improves by the inclusion of our main explanatory variables. The variance inflation factor (VIF) test indicates no serious problems of multicollinearity[23].

We argue by Hypothesis 4 that core and broker advantages are complementary and therefore effects of these network positions reinforce each other. In order to check the validity of this idea we have to look at the interaction term between the two main variables, which is more reliable in case of dichotomized variables. We introduce our dummy variables

[23] VIF values for our final continuous model can be found in the Appendix (Table 5.8).

stepwise in model (6) and model (7). Model (9) in Table 5.3 shows with these new dummy variables that in addition to being in the core, being a broker is also a crucial feature of individual success since both coefficients are positive and significant. Further, model (9) provides additional evidence that *Core* and *Broker* variables strengthen each other because their interaction term is also significant and positive. This result suggests that core creators who are brokers as well have higher probability of receiving an award than creators merely part of the core. This supports Hypothesis 4 that brokerage and core membership have complementary advantages and both effects help individual success by reinforcing each other.

Further, we posit that intermediation between core and periphery is done most effectively by core brokers who can benefit the most from being in the core and having access to new ideas and controlling free capacities in the periphery. We introduce *Gatekeeper* dummy in model (8) and include a three-way interaction of *Core*, *Broker* and *Gatekeeper* in a stepwise manner from model (10) to model (12). We find that *Core* and *Broker* creators have extra likelihood of award winning when they are also *Gatekeepers*, meaning that connections are rare between their core and peripheral neighbours in the ego network. The strong, positive and significant coefficient of the three-way interaction effect indicates that being a core and a broker creator at the same time especially increases the likelihood of award winning if the creator acts as a bridge between the core and the periphery. This verifies Hypothesis 5 and suggests that being in the core of a community and also bridging the core to the periphery significantly helps individual success.

Table 5.3 Relationship between Core and Broker position and award winning

	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Core	1.638***			1.509***	1.298***		1.145***
	(0.124)			(0.137)	(0.149)		(0.143)
Broker		1.641***		1.397***		1.046***	0.744**
		(0.182)		(0.295)		(0.317)	(0.349)
Gatekeeper			2.051***		1.841***	1.935***	0.594
			(0.152)		(0.362)	(0.202)	(1.036)
Core X Broker				2.699***			2.378***
				(0.198)			(0.426)
Core X Gatekeeper					2.556***		2.184***
					(0.166)		(0.213)
Broker X Gatekeeper						2.264 ***	1.926***
						(0.187)	(0.340)
Core X Broker X Gatekeeper							2.509***
							(0.182)
Constant	-48.028***	-46.205***	-49.736***	-51.424***	-52.232***	-50.543***	-13.721***
	(17.035)	(16.758)	(17.006)	(17.003)	(17.127)	(16.958)	(1.437)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Role FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
BIC	2710.52	2755.46	2683.06	2635.49	2624.50	2686.85	2697.33
Log lik.	-1228.37	-1265.91	-1229.71	-1196.98	-1191.49	-1222.66	-1276.38
Observations	7672	7672	7672	7672	7672	7672	7672

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01 Further control variables that are not reported in the table include Clustering, Average Path Length, Creator Movie Window, Films Per Window, Newcomer. Standard Errors are clustered at the creator level.

5.6 Conclusions and discussion

In this chapter, we aim for a novel understanding on how brokerage in core/periphery networks influence individual success. Our empirical exercise is based on a unique dataset of Hungarian feature films, from which we created time-varying collaboration networks of individual movie creators. Our results confirm that creators have the highest chance to win an award if they are in the core and broker the network at the same time. This core and broker positions in the network support individual success by providing complementary benefits for creators. Moreover, by applying a novel way to consider core/periphery structure in identifying network brokerage, we find that core creators who connects the core and the periphery are in the most favourable position to win awards. Major contribution of the chapter is therefore the new evidence that brokerage indeed matters in dichotomized core/periphery networks so that successful broker individuals diversify their relations between core and peripheral creators.

Complementarity of coreness and brokerage arguably arises from the disparate advantages they provide in mobilizing social capital. While network core yields quicker and wider recognition of creative outputs through prestige effects (Fraiberger et al., 2018), brokerage provides better inputs for creativity through diverse social connections. In the latter regard, periphery has been claimed important for creative production in previous literature because ideas resided there are fresher than conventions in the core (Schilling, 2005; Dahlander and Frederiksen, 2012; Cattani and Ferriani, 2008). Core brokers, therefore, can take advantage of controlling over new ideas from the periphery and their implementation in prestigious social circles. Finally, it might be easier for a core broker to mobilize capacities of her peripheral collaborators because core connections are usually time poor; whereas peripheral creators might devote much of their time to the project with a prestigious core artist.

To account for brokerage between core and periphery in the network, we applied two novel methodological approaches. First, we implemented the betweenness centrality measure suggested by Everett and Valente (2016) to capture brokerage better than ego-network variables. Second, we provide a simple measure we call gatekeeping to identify how much actors connect the core and the periphery in their ego network. Certainly, this measurement can be further developed and its' potential to capture the gatekeepers of knowledge in network needs further investigation (Allen, 1977; Morrison, 2008; Giuliani, 2011, Vicente et al., 2011). Moreover, we think that further methodological research is needed to generalize the mechanisms of brokerage in core/periphery structures. For example, one may bring our approach closer to the original measurements of brokerage and decompose network constraint in weighted networks. In the recent work, we assumed that

the whole team of collaborators contribute to the production of the movie equally and considered an unweighted graph only.

As our research focuses on the Hungarian film production that is primarily concentrated around the urban agglomeration of Budapest, it also connects well to previous studies on how networks matter for the exploitation of locally accumulated industrial knowledge (e.g. Giuliani and Bell, 2005; Boschma and Ter Wal, 2007; Morrison, 2008; Eriksson and Lengyel, 2019). In economic geography and related fields project-based creative industries in agglomerations have been studied frequently through collaboration networks (see Johns, 2006; Balland et al., 2013; Capone and Lazzeretti, 2018). We contribute to this discussion with evidence that core/periphery structure and brokerage matters in agglomerations also for individual performance. However, as we were unable to collect the necessary locational data, we cannot observe the influence of geographical concentration on individual success precisely.

Another issue one has to face when generalizing the argument comes from the usual limitation due to contextual focus and social network simplification. The findings of this study are based on a specific era of the Hungarian film production that is certainly not in the core of the global movie production. Even though the observed community covers a relatively isolated industry, international success of Hungarian artists or co-working relationships in foreign movie productions provided access to external social capital for the Hungarian film makers, which is invisible in our data. A more puzzling question is whether the same results hold if one looked at the global movie production, in which there are probably more than one core group of artists with their own peripheries (Everett and Borgatti, 1999). Brokerage between core groups might offer more fresh ideas in such networks if there are substantial qualitative differences residing in these subnetworks.

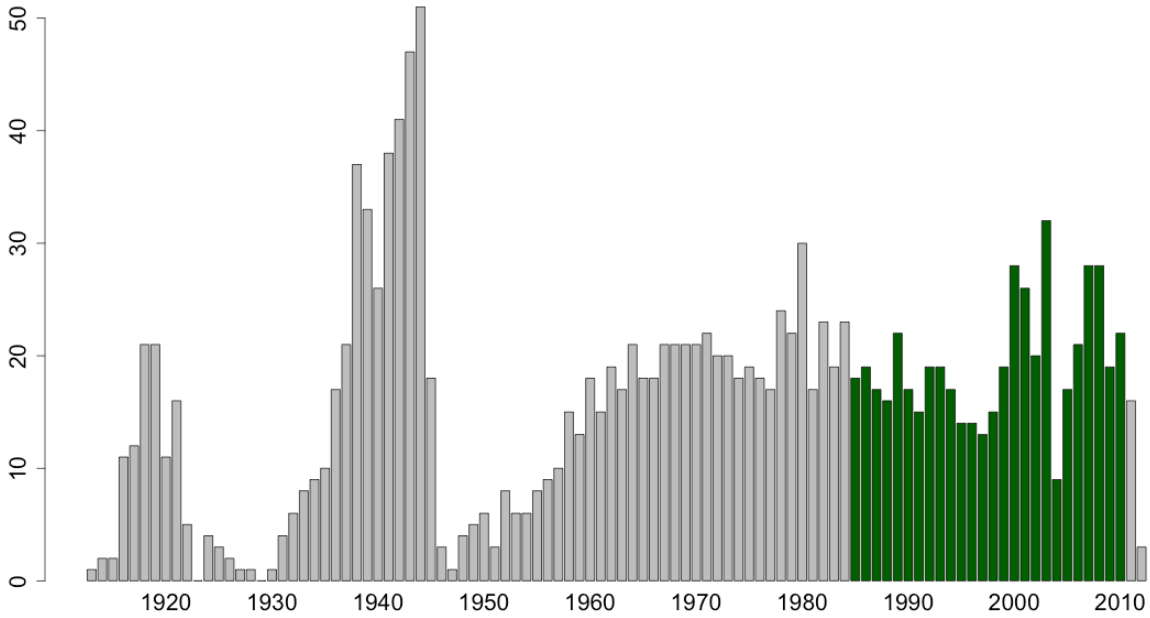
The qualities of creative products themselves are also important to incorporate better in future research of core/periphery brokerage. Looking at award winning as dependent variable always includes a potential risk that assignment of awards for artistic achievements is driven more by commercial or political rather than pure artistic considerations (Holbrook, 1999). Alternatively, the level of novelty might be used as dependent variable (de Vaan et al., 2015), which might also help us better understand how exploration of novel ideas and exploitation of free capacities characterize individual achievements of core brokers who work with peripheral artists. Another promising approach to evaluate the success of movies could be the citation among feature films (Spitz and Horvát, 2014; Bioglio and Pensa, 2018), however, the availability of such data is limited in the context of the Hungarian film production.

Finally, the dynamics of individual career trajectories needs to be considered more in depth in the approach we provide in this chapter. Because success bread success, individual creators are more attractive for colleagues if they have worked with prestigious collaborators or after they win an award (Liu et al., 2018). Therefore, it would be important to look at the differences regarding the effect of core/periphery brokerage on individual success across those individual artists who started as peripheral artists but become part of network core at

some point in their career versus those creators who have initially belonged to the core and developed peripheral connections later.

5.7 Appendix

Table 5.4 Number of movies produced in Hungary, 1912-2011



Source: Author's own construction based on the Hungarian National Film Archive.

Note: The examined period is highlighted by dark green colour.

Table 5.5 Collaboration patterns in movie creation based on 7-year moving windows

Year	Movies	Creators	Average creators per movie
1990	124	254	4.19
1991	124	259	4.19
1992	124	276	4.40
1993	123	291	4.62
1994	121	295	4.75
1995	113	295	5.07
1996	109	310	5.37
1997	109	316	5.45
1998	108	360	5.81
1999	117	412	5.81
2000	127	454	5.84
2001	133	466	5.78
2002	151	510	5.65
2003	147	493	5.62
2004	149	507	5.65
2005	153	491	5.47
2006	153	486	5.56
2007	155	498	5.54
2008	154	493	5.53
2009	144	465	5.49

Source: Author's own construction based on the Hungarian National Film Archive.

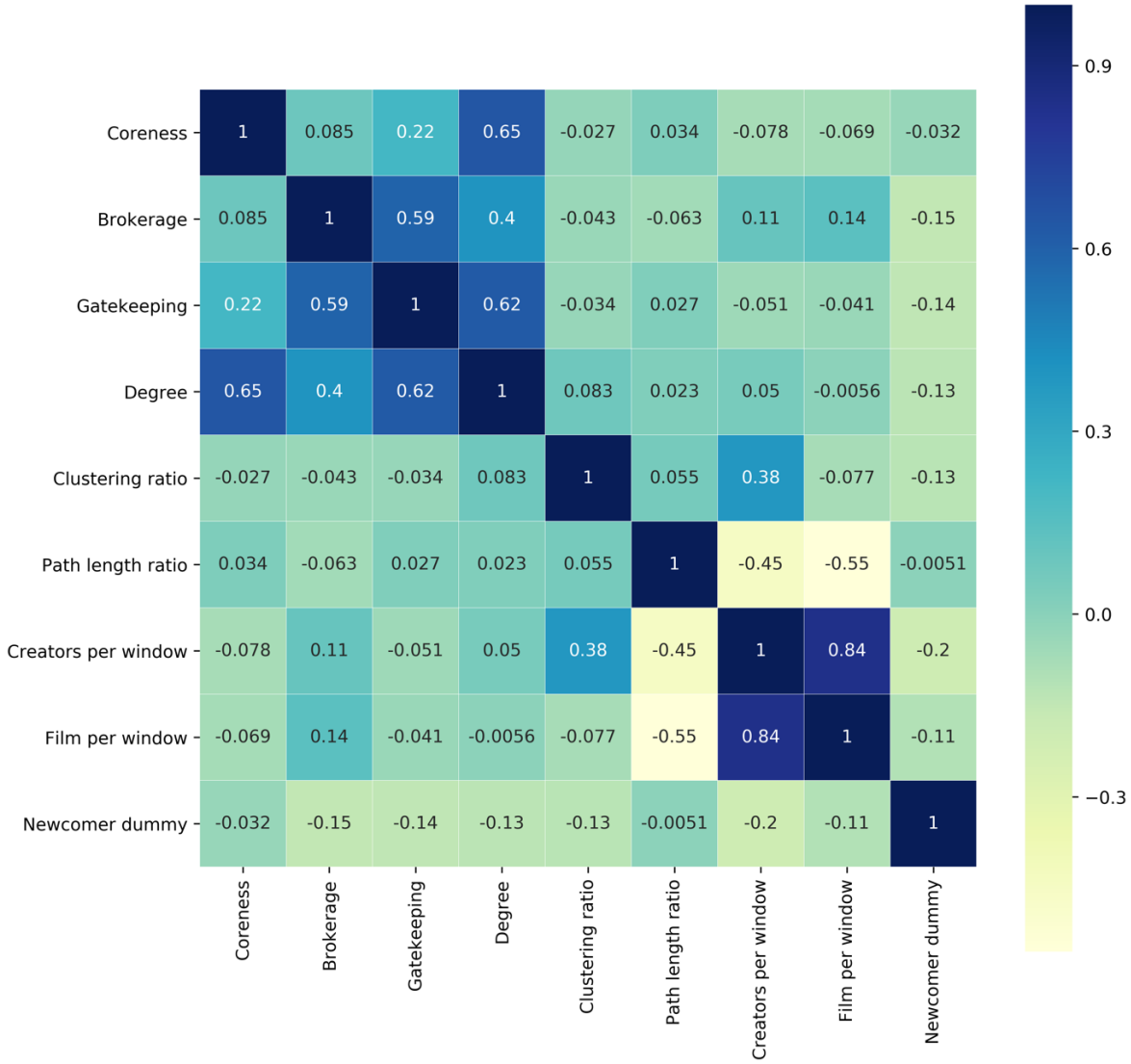
Note: Number of movies and number of movie creators per year are based on 7-year periods.

Table 5.6 Descriptive statistics of our continuous variables

	Observations	Min	Max	Mean	Std. dev.
Coreness	7672	0	0.470	0.019	0.046
Brokerage	7672	0	6.696	4.206	1.207
Gatekeeping	7672	0	1	0.170	0.310
Clustering ratio	7672	0.727	0.926	0.841	0.058
Path length ratio	7672	0.388	0.621	0.503	0.054
Creators Per Window	7672	254	510	425	87.222
Films Per Window	7672	108	155	136	16.883

Source: Author's own construction based on the Hungarian National Film Archive.

Figure 5.7 Correlation matrix of the variables



Source: Author's own construction.

Table 5.8 VIF-values in our final continuous model

	VIF	1/VIF
Coreness	4.93	0.203
Coreness^2	4.86	0.206
Brokerage	1.69	0.593
Gatekeeping	1.88	0.531

Source: Author's own construction

Table 5.9 Robustness check for k-core-based coreness measure

	Dependent variable – award winning			
	(1)	(2)	(3)	(4)
K-core	1.149***	1.126***	0.735***	0.369***
(z-score)	(0.105)	(0.106)	(0.118)	(0.138)
K-core^2	-0.239***	-0.212***	-0.110***	-0.083**
(z-score)	(0.034)	(0.034)	(0.035)	(0.035)
Brokerage		0.471***	0.185***	0.042
(z-score)		(0.037)	(0.061)	(0.073)
Gatekeeping			0.604***	0.477***
(z-score)			(0.079)	(0.092)
Degree				0.396***
(z-score)				(0.093)
Clustering ratio	-6.595	-5.966	-5.766	-5.298
	(25.295)	(25.220)	(25.416)	(25.546)
Path length ratio	23.066	22.749	25.348	25.254
	(24.448)	(24.309)	(24.499)	(24.609)
Creators Per Window	0.011	0.011	0.017	0.018
	(0.018)	(0.018)	(0.018)	(0.018)
Films Per Window	0.046	0.048	0.031	0.028
	(0.103)	(0.102)	(0.103)	(0.103)
Newcomer (dummy)	2.007***	2.409***	2.795***	2.796***
	(0.128)	(0.135)	(0.147)	(0.145)
Constant	-22.296	-23.332	-24.956	-25.129
	(15.499)	(15.441)	(15.566)	(15.639)
YEAR FE	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes
BIC	2719.977	2616.452	2552.608	2541.553
Log likelihood	-1243.699	-1187.464	-1151.069	-1141.069
Observations	7672	7672	7672	7672

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01

Table 5.10 Robustness check for constraint-based Brokerage variable and the influence of Degree

	Dependent variable – award winning			
	(1)	(2)	(4)	(3)
Brokerage (betweenness)		0.569***		0.055
(z-score)		(0.090)		(0.079)
Brokerage (constraint)	-0.543***		-0.466***	
(z-score)	(0.105)		(0.100)	
Coreness	0.337***	-0.033**	0.222***	0.336***
(z-score)	(0.088)	(0.014)	(0.102)	(0.112)
Coreness^2	-0.010	0.124*	-0.003	-0.016
(z-score)	(0.012)	(0.073)	(0.012)	(0.014)
Gatekeeping	0.569***	0.670***	0.496***	0.573***
(z-score)	(0.065)	(0.083)	(0.081)	(0.096)
Degree			0.177**	0.263***
(z-score)			(0.085)	(0.086)
Clustering ratio	-1.614	-2.255	-2.359	-2.940
	(25.135)	(24.964)	(25.150)	(25.104)
Path length ratio	23.206	24.561	23.465	24.579
	(24.770)	(24.558)	(24.758)	(24.652)
Films Per Window	0.028	0.026	0.029	0.027
	(0.107)	(0.106)	(0.106)	(0.106)
Newcomer (dummy)	2.988***	2.853***	2.971***	2.849***
	(0.150)	(0.146)	(0.149)	(0.146)
Creators Per Window	0.098	0.020	0.019	0.019
	(0.018)	(0.018)	(0.018)	(0.018)
Constant	-28.443*	-28.364*	-27.646*	-27.437*
	(15.591)	(15.481)	(15.592)	(15.552)
YEAR FE	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes
BIC	2507.845	2532.006	2511.940	2530.508
Log likelihood	-1128.688	-1140.766	-1126.263	-1135.547
Observations	7676	7672	7676	7672

Source: Author's own construction

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01

Table 5.11 Robustness check for only individual awards as dependent variable

	Dependent variable – only individual award winning			
	(1)	(2)	(3)	(4)
Coreness	1.244***			-0.125
(z-score)	(0.417)			(0.216)
Coreness^2	-0.279*			0.032
(z-score)	(0.152)			(0.035)
Brokerage		0.545***		0.008
(z-score)		(0.062)		(0.123)
Gatekeeping			1.194***	1.210***
(z-score)			(0.085)	(0.153)
Clustering ratio	82.496	69.658	77.901	76.602
	(57.348)	(52.723)	(54.288)	(54.198)
Path length ratio	-54.174	-42.995	-51.270	-50.098
	(52.252)	(49.030)	(50.297)	(50.300)
Films Per Window	0.257	0.227	0.234	0.229
	(0.207)	(0.196)	(0.203)	(0.203)
Newcomer (dummy)	3.549***	3.704***	4.653***	4.659***
	(0.273)	(0.263)	(0.307)	(0.309)
Creators Per Window	-0.027	-0.023	-0.022	-0.021
	(0.036)	(0.034)	(0.035)	(0.035)
Constant	-69.796**	-62.854**	-67.756**	-66.950**
	(34.427)	(31.671)	(32.952)	(32.923)
YEAR FE	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes
BIC	1154.968	1120.291	1007.158	1032.943
Log likelihood	-464.002	-451.028	-394.462	-394.260
Observations	6182	6182	6182	6182

Source: Author's own construction

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01

Table 5.12 Robustness check with variables cut-off at 0.75 percentile

	Dependent variable – award winning						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Core	1.638***			1.482***	1.582***		1.210***
	(0.125)			(0.150)	(0.181)		(0.171)
Broker		1.418***		1.202 ***		0.525***	-0.143
		(0.143)		(0.195)		(0.294)	(0.356)
Gatekeeper			2.110***		2.126***	1.865***	1.520***
			(0.156)		(0.205)	(0.196)	(0.278)
Core X Broker				2.740***			3.255***
				(0.173)			(0.400)
Core X Gatekeeper					2.959***		2.309***
					(0.171)		(0.231)
Broker X Gatekeeper						2.333 ***	1.909***
						(0.167)	(0.196)
Core X Broker X Gatekeeper							2.632***
							(0.165)
Constant	-48.028***	-47.770***	-55.393***	-53.470***	-58.656***	-55.571***	-15.455***
	(17.035)	(16.925)	(16.955)	(17.129)	(17.201)	(17.007)	(1.518)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Role FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
BIC	2680.38	2741.69	2623.65	2610.93	2546.40	2631.13	2631.78
Log likelihood	-1228.37	-1259.03	-1200.01	-1184.70	-1152.44	-1194.80	-1243.60
Observations	7672	7672	7672	7672	7672	7672	7672

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01 Further control variables that are not reported in the table include Clustering, Average Path Length, Creator Movie Window, Films Per Window, Newcomer. Core, Broker and Gatekeeper variables are constructed by the 0.75 percentile cut-off of the continuous measures.

Table 5.13 Robustness check with random networks and random award winners

	Dependent variable – award winning				
	(1) FULL	(2) Random awards	(3) Random rewire	(4) Randomized creators	(5) Random network
Coreness	0.569***	-0.099	1.072***	0.211	-0.002
(z-score)	(0.090)	(0.116)	(0.110)	(0.155)	(0.081)
Coreness^2	-0.033**	0.017	-0.088***	-0.063	0.047*
(z-score)	(0.014)	(0.021)	(0.029)	(0.040)	(0.027)
Brokerage	0.124*	-0.016	-0.069	-0.112	-0.053
(z-score)	(0.073)	(0.070)	(0.073)	(0.086)	(0.107)
Gatekeeping	0.670***	0.009	0.060	-0.011	-0.002
(z-score)	(0.083)	(0.073	(0.103)	(0.072)	(0.065)
Constant	-28.364*	-17.295**	-17.026***	-13.231***	-15.488***
	(15.481)	(7.555)	(5.247)	(4.997)	(5.079)
Controls	Yes	Yes	Yes	Yes	Yes
YEAR FE	Yes	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes	Yes
BIC	2532.006	3074.756	2584.146	3068.464	2888.347
Log likelihood	-1140.766	-1412.144	-1166.838	-1408.997	-1318.939
Observations	7672	7672	7672	7672	7672

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01 Further control variables that are not reported in the table include Clustering, Average Path Length, Creator Movie Window, Films Per Window, Newcomer. Model (1) FULL represent our original model. Model (2) Random awards represents results in case we randomly redistribute awards among individuals in each year. Model (3) Random rewire shows the results in case of a randomly rewired network, preserving the original graph's degree distribution. Model (4) Randomized creators keeps the original structure of the network, but creators are randomly signed to different positions. Model (5) Random network is based on randomly generated networks with similar properties. The table strengthen the argument that attributes of individual and network properties indeed influence the success of individuals.

Table 5.14 Robustness checks based on rare event logit regression models

	Dependent variable – award winning			
	(1)	(2)	(3)	(4)
Coreness	0.842***			0.558***
(z-score)	(0.087)			(0.090)
Coreness^2	-0.080***			-0.031**
(z-score)	(0.018)			(0.016)
Brokerage		0.463***		0.125**
(z-score)		(0.041)		(0.058)
Gatekeeping			0.880***	0.664***
(z-score)			(0.055)	(0.070)
Constant	36.455***	36.054***	33.429***	34.665**
	(9.812)	(13.885)	(9.153)	(15.356)
Controls	Yes	Yes	Yes	Yes
YEAR FE	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes
BIC	2634.881	2690.742	2521.088	2448.863
Log likelihood	-1190.935	-1223.384	-1143.074	-1088.890
Observations	8401	8401	8401	8401

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01 Further control variables that are not reported in the table include Clustering, Average Path Length, Creator Movie Window, Films Per Window, Newcomer.

Table 5.15 Robustness checks with individual level fixed effect models

	Dependent variable – award winning				
	(1)	(2)	(3)	(4)	(5)
Coreness	0.651***		0.639***		0.291*
(z-score)	(0.139)		(0.143)		(0.151)
Coreness^2	-0.064***		-0.062***		0.002
(z-score)	(0.021)		(0.022)		(0.023)
Brokerage		0.437***	0.431***		0.196**
(z-score)		(0.078)	(0.078)		(0.091)
Gatekeeping				0.754***	0.673***
(z-score)				(0.090)	(0.104)
Controls	Yes	Yes	Yes	Yes	Yes
YEAR FE	Yes	Yes	Yes	Yes	Yes
INDIVIDUAL FE	Yes	Yes	Yes	Yes	Yes
BIC	1272.019	1263.027	1250.515	1214.709	1213.200
Log likelihood	-545.350	-544.795	-530.656	-520.637	-508.057
Observations	2653	2653	2653	2653	2653

Source: Author's own construction.

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01 Further control variables that are not reported in the table include Clustering, Average Path Length, Creator Movie Window, Films Per Window, Newcomer.

Table 5.16 Robustness check based on 5-year long moving windows

	Dependent variable – only individual award winning			
	(1)	(2)	(3)	(4)
Coreness	0.930***			0.598***
(z-score)	(0.169)			(0.117)
Coreness^2	-0.150***			-0.095***
(z-score)	(0.054)			(0.029)
Brokerage		0.473***		0.264***
(z-score)		(0.049)		(0.058)
Gatekeeping			0.643***	0.363***
(z-score)			(0.051)	(0.072)
Constant	-16.221***	-15.867***	-17.542***	-17.814***
	(3.967)	(3.939)	(4.013)	(4.015)
Controls	Yes	Yes	Yes	Yes
YEAR FE	Yes	Yes	Yes	Yes
ROLE FE	Yes	Yes	Yes	Yes
BIC	2664.797	2653.229	2573.424	2564.400
Log likelihood	-1218.837	-1217.421	-1177.518	-1159.903
Observations	6220	6220	6220	6220

Source: Author's own construction

Note: Standard errors in parentheses * p<0.10, ** p<0.05, *** p<0.01 Further control variables that are not reported in the table include Clustering, Average Path Length, Creator Movie Window, Films Per Window, Newcomer.

Chapter 6

Conclusion

6.1 Main findings and implications

Broadly, this dissertation has studied tie formation in different networks of knowledge sharing and collaboration and has emphasized the role geography plays in this process. Modelling how and why different pieces of economies get connected is a fundamental challenge to understand society as a complex system. All the above chapters contribute to the literature on networks in economic geography, with a common aim to better understand the drivers and outcomes of network ties that channel ideas, knowledge and technologies in agglomeration. The thesis has two specific objectives. First, the introduction of new factors, approaches and contexts to study tie formation processes in spatial networks. Second, connecting network tie formation in agglomeration to the performance of nodes. These objects are articulated in four specific research questions:

RQ 1) Do spinoff background matters for tie formation in cluster knowledge networks?
RQ 2) What mechanisms drive the creation and persistence of ties in cluster knowledge networks?
RQ 3) How do co-location, complexity and collaboration determine technological relatedness ties?
RQ 4) How do bridging ties in core/periphery networks of creative production influence individual success?

Each question sets up different perspectives to study tie formation in spatial networks, while emphasizing the importance of geography in learning, knowledge sharing and technological combination. In the following, the main conclusions and implications of the thesis are summarised in the order of the empirical chapters and then at the thesis level.

Do spinoff background matters for tie formation in cluster knowledge networks? The chapters' general contribution is that it introduces a new factor, spinoff background to consider while studying cluster knowledge networks. It takes the first steps to combine two streams of literature, namely the one that emphasizes the importance of knowledge network ties to exploit positive agglomeration externalities in industry clusters (Giuliani and Bell, 2005), and the one that highlights the importance of spinoff companies in the persistence and prosperity of clusters (Klepper, 2007; 2010; Buenstorf and Klepper, 2009). The chapter shows that spinoffs are more capable to collaborate in cluster knowledge networks. After controlling for firm characteristics, dyad level proximities of companies and network structural properties, being a spinoff company still determines how effectively firms form ties in cluster networks.

What mechanisms drive the creation and persistence of ties in cluster knowledge networks? The main objective of the chapter is to present a novel approach to study the drivers of knowledge network formation by differentiating between the creation and persistence of ties. Previous studies that empirically investigate the dynamics of networks behind industry clusters (Giuliani, 2013; Balland et al., 2016) suggest that on the macro-level cluster knowledge networks are relative stable over time, but they are rapidly changing on the micro-level. These studies identify the key drivers of network formation in clusters, however, they do not differentiate between the creation of new ties and the maintenance of previously existing relationships. Therefore, the chapter proposes a new framework to study the micro-processes and possible determinants of tie creation and tie persistence in knowledge networks separately, in order to better understand the evolution of clusters. It suggests to consider related costs, uncertainties and possible knowledge redundancy of forming relationships. The general argument is that the differentiation of tie creation from tie persistence helps us to capture forces that determine the organization of clusters over time. Findings indeed demonstrate that forces of selection, retention and variation are also present during the evolution of cluster knowledge networks, which suggests that network relations are organized in a way to avoid the lock-in of the cluster. As findings show significant differences in the determinants of tie creation and tie persistence, it outlines a stream of future research to discuss in the next chapter.

How do co-location, complexity and collaboration determine technological relatedness ties? The general objective of the chapter is to extend the literature of economic geography on tie formation by a new context, namely the network of technologies. The chapter also contributes to the flourishing research area on technological relatedness by turning the usual argumentation around and use relatedness as a dependent variable. Results show that geography, complexity and collaboration influence both the formation of technological relatedness ties and the strength of these linkages. Research on the determinants of technological relatedness also shed light on how agglomeration and complexity influence the perceived proximity of technologies. It also has important consequences for recent regional development strategies as they heavily build on concepts such as relatedness, complexity and comparative advantage in technologies (Balland et al., 2018), however, these issues are visited in detail in the next chapter.

How do bridging ties in core/periphery networks of creative production influence individual success? The objective of the last empirical chapter is to combine the literature on agglomeration, social network formation and creative performance. Economic geographers argue that unlike manufacturing industries, creative production is very much related to urban agglomerations as local buzz or the geographically concentrated network of actors is essential to effectively update each other, exchange ideas, information and negotiate decisions (Scott,

2004; Grabher, 2001; Storper and Venables, 2004). The chapter in relation contributes to the discussion on the spatial networks of creative industries (Balland et al., 2013) by focusing on collaboration ties inside an agglomeration and show how the formation of specific linkages matters for the success of individuals. The main findings uncover that creators are more likely to achieve individual success if they have a position in the industry core and form ties that bridge the core and the periphery of the network. Besides the relevant conclusion for the economic geography of creative industries, the chapter aims to contribute to the more general discussion on brokerage in core/periphery networks in the social network literature. Moreover, in contrast to previous studies in economic geography on creative industries, this study analysis individual level network relations to understand the relationship between specific ties and performance.

Overall the dissertation introduces four novel aspects to consider while studying networks in order to better understand innovativeness and high performance of economic actors in agglomeration. Besides searching for new drivers and approaches to explain network tie formation, the thesis reinforces the argument that geography is a key determinant of knowledge exchange and technological combination. The dissertation contributes to the literature on industry clusters by partly confirming the results of previous studies and introducing a novel factor and a new approach to consider in relation. Moreover, it emphasizes that taking ideas and perspectives from fields such as entrepreneurship, management studies or network science can further deepen our understanding on the development of industry clusters. It establishes that besides geographical proximity of firms, spinoff background matters for collaboration in industry clusters. An important implication is that entrepreneurial capabilities and background determine firms' involvement in cluster knowledge networks. Geographical proximity is also proved to be more important to establish cooperation in clusters than to maintain linkages. This finding extends the literature by showing that physical proximity of actors indeed creates opportunities for cooperation, however, its influence on social ties is questionable in case we consider the creation and persistence of ties separately in a dynamic setting.

The dissertation also extends the literature of economic geography on networks by showing that geography does not only enhances tie formation between firms, but also between technologies. More precisely, in the context of technological relatedness co-location also determines the presence and strength of linkages. A key implication of this is that further research on relatedness should consider it as a dynamically evolving feature and should also be aware that it co-evolves with agglomeration.

The thesis also outlines possibilities to study individual performance by dynamic networks inside agglomeration. More specifically it demonstrates that forming structurally important ties in urban social networks determine the success of individuals. By this conclusion the dissertation also connects the literature on networks in agglomeration and node performance.

In sum, every chapter of the thesis is built around the existing literature of economic geography on networks. All the empirical cases challenge the argument that agglomeration or the geographical proximity of actors is key for collaboration, knowledge sharing and technological combinations. Each of the chapters introduces novel aspects to consider the heterogeneity of nodes while studying spatial network relations that channel ideas, knowledge and technology. Besides identifying new determinants, context and approaches, each of the chapters reinforce that geography is an important component of network tie formation.

6.2 Limitations and future research

A number of limitations as well as related future research questions must be highlighted regarding each studies of the thesis and also in general terms.

An explicit aim of the thesis is to introduce new factors as drivers of spatial network formation. To do so, it suggests that spinoff background matters for tie formation in cluster knowledge networks. As the thesis only takes the first steps to combine the two streams of literature, namely the literature on cluster knowledge networks and the literature on spinoffs in clusters, there are several opportunities for future research in this direction. A key limitation in relation is that the present study does not have detailed information on spinoff-parent relations and on the actual inherited capabilities and routines of spinoff companies. The precise measurement of entrepreneurial capabilities and routines that enable spinoffs to effectively collaborate in cluster networks is a key challenge for future research. A central question in relation is, for what extent spinoffs inherit connections and network ties from their parent company. Bagley (2019) takes the first steps in relation and shows that the personal ties of founders to parent companies support the high performance of spinoffs in clusters. However, the description of tie-inheritance in cluster networks is still an important subject to study. It is also in line with the other objective of the thesis as tie-inheritance could be seen as a form of tie persistence.

The separation of tie creation and tie persistence in spatial networks of various kinds offers numerous opportunities for future research. As it by definition requires a micro-perspective, it can help us to more accurately describe network evolution and the evolution of regional economies themselves. There are only a few studies in economic geography (notable exceptions are Broekel and Bednarz, 2018; Giuliani et al., 2018) that search for differences in creation, maintenance and dissolution of relationships. The distinction is important in many contexts. In clusters it helps to understand how firms behave to avoid lock-in situations. In larger and more sparse geographical knowledge networks – such as Europe width scientific or co-inventor networks – the approach to identify newly created and repeated inter-regional collaboration has potentials to inform policy about the backbone of R&D collaboration across space or the potential isolation of regions in the long run. As the

literature on strong ties in social networks suggest that repeated interaction is crucial for the transfer of complex knowledge, studying persistent collaboration in inter-regional settings is an important challenge for the future to understand the evolutionary processes behind the European Research Area.

The dissertation also aims to search for the determinants of tie formation in the context of technological relatedness. A key limitation in relation is that the analysis is only at the tie level and it does not take into account the structural patterns of the technology space. To better understand how do unrelated technologies turn to be related over time, a network perspective similar to the studies on cluster networks is necessary regarding the dynamics of the technology space. Separating tie creation from tie persistence in this context could also help to gain further insights, as new tie creation represents a shift from unrelated to related technologies. As technological relatedness has become a key component to develop strategies for regional economies, a key issue for the future of regional diversification and development strategies is to understand how regions should adapt to the dynamics of the technology space itself.

Besides searching for the drivers of spatial network tie formation, the thesis also aims to connect the performance of nodes to their network ties. The tested methodology and indicators on brokerage in core/periphery networks could also be useful to explore the different attitudes and performances of European regions in networks of research and innovation. As research on spinoffs and industrial agglomeration also highlights that spinoff companies tend to perform better and have higher survive rate than other firms in clusters (e.g. Carias and Klepper, 2010; Klepper, 2010; Boschma and Wenting, 2007; Buenstorf and Guenther, 2011; Wenting, 2008), it is a straightforward next step to study how network ties of spinoffs contribute to their economic performance and long-term survival. Moreover, it is also a promising future research direction to connect the performance of nodes to their behavior in terms of new tie creation and tie maintenance.

A common limitation to all the research of this thesis comes from data availability. Datasets that combine information on network relations with detailed geographical information are still rare, however, they are necessary to more precisely track the evolution of clusters, industries, technologies or regional economies. For example, employer-employee datasets, online social network data (such as Twitter, Foursquare or LinkedIn datasets), communication data (e-mail or phone call record datasets) or tax data (value added tax flow datasets) are possible sources that contain detailed, longitudinal, relational and geographical information to more precisely track the formation of ties in regions, industries or technology fields in the future.

As the products of creative industries are more and more related to the world wide web, many sources of data became available on project-based cultural and creative production during the past decade. However, a key limitation for many datasets and also for the related study of this thesis is that most available data sources do not have detailed geographical information on the production process, involved companies, individuals or institutions. The

combination of individual collaboration networks with the micro-level geography of creative workers is a necessity for future research to understand the spatial patterns of creative production. More detailed datasets could also allow contributions from a geographical point of view to the emerging field of 'science of science' or 'science of success' (Fortunato et al., 2018).

All the empirical exercises of this thesis are case studies with specific settings, therefore the generality of the results should be further tested. Three chapters focus specifically on Hungarian cases. In the post socialist country of Hungary inward FDI has played a major role in the economic transition after the fall of the Iron Curtain (Lengyel and Leydesdorff, 2011; Radosevic, 2002; Resmini, 2007). Even until now, there is a remarkable difference in the productivity of domestic and foreign-owned firms in Hungary (Elekes and Lengyel, 2016). Employees who collect experience at more productive multinational enterprises (MNEs) and then move to domestic firms significantly improve the productivity of Hungarian companies (Csáfordi et al., 2018). Moreover, foreign-owned firms are also more likely to introduce innovations (Halpern and Muraközy, 2012). This context provides great opportunities for future research, especially to study the role MNEs and FDI play in less developed regions of Europe. The dissertation also provides tools and approaches to study the differences in network tie formation between foreign and domestic companies in Central and Eastern European countries.

None of the objectives of this thesis are explicitly related to policy issues or to economic development strategies. However, the studies of the dissertation have important, yet implicit implications for cluster policy, policy on creative and cultural sectors or regional economic development programs.

Looking at the formation of linkages in cluster knowledge networks made clear that spinoff companies are key actors in the organization of knowledge sharing and collaboration in clusters. This conclusion suggests that by convincing spinoffs to actively participate in the creation of local cluster development strategies could also help the acceptance of these plans (Caniels and Romijn, 2008). Firms in clusters also consciously create and maintain their ties to other companies and these local networking mechanisms help to avoid cluster lock-in. However, as these mechanisms could be different during the life-cycle of clusters, cluster policy in an earlier stage should still focus on creating opportunities for interaction at the first place. The above results show that firms are capable to manage their local ties deliberately, therefore, it is reasonable to suggest that cluster policy should rather focus on providing novel and high quality (external) technological knowledge to the cluster.

Relatedness is also shown to be a dynamically evolving feature influenced by co-location, complex and collaboration. Because relatedness of technologies changes over time, long-term development strategies could consider unrelated diversification as a reasonable alternative (Pinheiro et al., 2018). Given the concentration and complexity of technologies in a region, linking previously unrelated technologies for the first time seems to be a risky, but valid long-term goal for regional economic development.

Linking performance to network tie formation provides evidence that individuals are more successful in urban creative industries in case they connect central and peripheral actors. As public founding is still key in many of these industries (especially in the European film industry), grant applications should contain information on the diversity of the applicants' previous collaborators. Programs or workshops that enhance the meeting and communication of young and already established creators can also have an indirect impact on future success. Moreover, the connection of nodal performance and network ties are important to consider during policy evaluation, especially in cases where the explicit aim of the policy is to connect different regions to combine their knowledge.

Bibliography

Abramson, E. and Rosenkopf, L. (1997): Social network effects on the extent of innovation diffusion: a computer simulation. *Organization Science*, 8 (3), 289-309.

Agarwal, R., Echambadi, R., Franco, A. M. and Sarkar, M. B. (2004): Knowledge transfer through inheritance: spin-out generation, development, and survival. *Academy of Management Journal*, 47 (5), 501-522.

Allen, T. J. (1977): *Managing the Flows of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*. MIT Press, Cambridge, MA.

Almeida, P. and Kogut, B. (1999): Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45 (7), 905-917.

Amin, A. (2000): Industrial districts. In Sheppard, E., Barnes T.J. (eds.) *A Companion to Economic Geography*. Oxford: Blackwell Publishing, 149-168.

Anheier, H. K., Gerhards, J. and Romo, F. (1995): Forms of Capital and Social Structure in Cultural Fields: Examining Bourdieu's Social Topography. *American Journal of Sociology*, 100 (4), 859—903.

Aral, S. and van Alstyne, M. (2011): The Diversity-Bandwidth Trade-off. *American Journal of Sociology*, 117 (1), 90-171.

Aral, S. (2016): The future of weak ties. *American Journal of Sociology*, 121 (6), 1931-1939.

Armstrong, J. S. and Overton, T. S. (1977): Estimating non-response bias in mail surveys. *Journal of Marketing Research*, 14 (3), 396-402.

Asheim, B. (1996): Industrial districts as learning regions: a condition for prosperity. *European Planning Studies*, 4 (4), 379-400.

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

Bagley, M. J. O. (2019): Small worlds, inheritance networks and industrial clusters. *Industry and Innovation*, 26 (7), 741-768.

Balassa, B. (1965): Trade Liberalization and Revealed Comparative Advantage. *The Manchester School*, 33 (2), 99-123.

Balland, P. A. (2012): Proximity and the Evolution of Collaboration Networks: Evidence from Research and Development Projects within the Global Navigation Satellite System (GNSS) Industry. *Regional Studies*, 46 (6), 741-756.

Balland, P-A., De Vaan, M. and Boschma, R. (2013): The dynamics of interfirm networks along the industry life cycle: The case of the global video game industry, 1987-2007. *Journal of Economic Geography*, 13 (5), 741-765.

Balland, P. A., Boschma, R. and Frenken, K. (2015): Proximity and Innovation: From Statics to Dynamics. *Regional Studies*, 49 (6), 907-920.

Balland, P.-A., Belso-Martínez, J. A. and Morrison, A. (2016): The dynamics of technical and business networks in industrial clusters: embeddedness, status or proximity? *Economic Geography*, 92 (1), 35-60.

Balland, P. A. and Rigby, D. (2017): The Geography of Complex Knowledge. *Economic Geography*, 93 (1), 1-23.

Balland, P. A., Boschma, R., Crespo, J. and Rigby, D. (2018): Smart Specialization Policy in the European Union: Relatedness, Knowledge Complexity and Regional Diversification. *Regional Studies*, 1-17.

Barabási, A-L. and Albert, R. (1999): Emergence of scaling in random networks. *Science*, 286 (5439), 509-512.

Barabási, A-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T. (2002): Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311, (3-4), 590-614.

Barabási, A.-L. (2018a): *Network Science*. Cambridge University Press, Cambridge, United Kingdom.

Barabási, A-L. (2018b): *The Formula: The Universal Laws of Success*. Hachette Book Group. New York, NY.

Bastos, M., Piccardi, C., Levy, M., McRoberts, N. and Lubell, M. (2018): Core-periphery or decentralized? Topological shifts of specialized information on Twitter. Social *Networks*, 52 (1), 282-293.

Bathelt, H. and Glückler, J. (2003): Toward a relational economic geography. *Journal of Economic Geography*, 3 (2), 117-144.

Bathelt, H., Malmberg, A. and Maskell, P. (2004): Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28 (1), 31-56.

Beckman, C. M., Haunschild, P. R. and Phillips, D. J. (2004) Friends or strangers? Firm-specific uncertainty, market uncertainty, and network partner selection. *Organization Science*, 15 (3), 259-275.

Bielby, W. T. and Bielby, D. D. (1999): Organizational mediation of project-based labor markerts: talent agencies and the careers of screenwriters. *American Sociological Review*, 64, 64-85.

Bioglio, L. and Pensa, R. G. (2018): Identification of key films and personalities in the history of cinema from a Western perspective. *Applied Network Science*, 3-50.

Borgatti, S. P. and Everett, M. G. (1999): Models of core/periphery structures. *Social Networks*, 21 (4), 375-395.

Borgatti, S. P., Everett, M. G. and Freeman, L. C. (2002): *Ucinet 6 for Windows*. Analytic Technologies, Harvard.

Borgatti, S. P., Mehra, A., Brass, D. J. and Labiance, G. (2009): Network Analysis in the Social Sciences, *Science*, 323 (5916), 892-895.

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

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

Boschma, R. and Wenting, R. (2007): The spatial evolution of the British automobile industry: Does location matter*? Industrial and Corporate Change*, 16 (2), 213-238.

Boschma, R. and Frenken, K. (2010): The spatial evolution of innovation networks. A proximity perspective. In Boschma, R. and Martin, R. (eds.) *The Handbook of Evolutionary Economic Geography*. Cheltenham, Edward Elgar, 120-135.

Boschma, R. and Fornahl, D. (2011): Cluster evolution and a roadmap for future research. *Regional Studies*, 45 (10), 1295-1298.

Boschma, R. (2015): Do spinoff dynamics or agglomeration externalities drive industry clustering? A reappraisal of Steven Klepper's work. *Industrial and Corporate Change*, 24 (4), 859-873.

Boschma, R., Balland, P. A. and Kogler, D. (2015): Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24 (1), 223-250.

Boschma, R. (2017): Relatedness as driver behind regional diversification: a research agenda. *Regional Studies*, 51 (3), 351-364.

Boschma, R. and Frenken, K. (2018): Evolutionary Economic Geography. In Clarke, G., Feldman, M. and Wojcik, D. (eds.): *New Oxford Handbook of Economic Geography*. Oxford, Oxford University Press, 213-229.

Breschi, S., Lissoni, F. and Malerba, F. (2003): Knowledge-relatedness in firm technological diversification. *Research Policy*, 32 (1), 69-87.

Breschi, S. and Lissoni, F. (2009): Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*, 9 (4), 439-468.

Broekel, T. and Binder, M. (2007): The regional dimension of knowledge transfers – a behavioral approach. *Industry and Innovation*, 14 (2), 151-175.

Broekel, T. and Boschma, R. (2012): Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography*, 12 (2), 409-433.

Broekel, T. and Hartog, M. (2013a): Explaining the Structure of Inter-Organizational Networks using Exponential Random Graph Models*. Industry and Innovation*, 20 (3), 277-295.

Broekel, T. and Hartog, M. (2013b): Determinants of Cross-Regional R&D Collaboration Networks: An Application of Exponential Random Graph Models. In Scherngell, T. (eds.) *The Geography of Networks and R&D Collaborations*. Springer, Cham, 49-70.

Broekel, T., Balland, P.-A., Burger, M. and van Oort, F. (2014): Modeling knowledge networks in economic geography: a discussion of four methods. *The Annals of Regional Science*, 53 (2), 423–452.

Broekel, T. (2015): The co-evolution of proximities – a network level study. *Regional Studies*, 49 (6), 921-935.

Broekel, T. and Bednarz, M. (2018): Disentangling link formation and dissolution in spatial networks: An application of a two-mode STERGM to a project-based R&D network in the German biotechnology industry. *Networks and Spatial Economics*, 1-28.

Broekel, T. (2019): Using structural diversity to measure the complexity of technologies. *PLOS One*, 14 (5), ee0216856.

Buenstorf, G. and Klepper, S. (2009): Heritage and agglomeration: the Akron tyre cluster revisited. *The Economic Journal*, 119 (537), 705-733.

Buenstorf, G. and Klepper, S. (2010): Why does entry cluster geographically? Evidence from the US tire industry. *Journal of Urban Economics*, 68 (2), 103-114.

Buenstorf, G. and Guenther, C. (2011): No Place Like Home? Relocation, Capabilities, and Firm Survival in The German Machine Tool Industry After World War II. *Industrial and Corporate Change*, 20 (1), 1-28.

Buerger, M., Broekel, T. and Coad, A. (2012): Regional dynamics of innovation – Investigating the co-evolution of patents, R&D, and employment. *Regional Studies*, 46(5), 565–582.

Burger, M., Van Oort, F. and Linders, G. J. (2009): On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4 (2), 167-190.

Burt, R. (1992): *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.

Burt, R. (2000): Decay functions. *Social Networks*, 11, 1-28.

Burt. R. S. (2004): Structural holes and good ideas. *American Journal of Sociology*, 110 (2), 348-399.

Burt, R. (2005): *Brokerage and closure. An introduction to social capital*. Oxford University Press, New York.

Caniels, M. C. J. and Romijn, H. A. (2003): Firm-level accumulation and regional dynamics. *Industrial and Corporate Change*, 12 (6), 1253-1278.

Caniels, M. C. J. and Romijn, H. A. (2008): Actor networks in strategic niche management: insights from social network theory. *Futures*, 40 (7), 613-629.

Caniels, M. C. J., Kronenberg, K. and Werker, C. (2014): Conceptualizing proximity in research collaborations between universities and firms. In Rutten, R., Benneworth, P., Irawati, D. and Boekema, F. (eds.): *The social dynamics of innovation networks*. Routledge, 221-238.

Cantner, U. and Graf, H. (2006): The network of innovators in Jena: an application of social network analysis. *Research Policy*, 35 (4), 463-480

Capone, F. and Lazzeretti, L. (2016): Interorganisational networks and proximity: an analysis of R&D networks for cultural goods. *Sinergie*, 34 (101), 53-70.

Capone, F. and Lazzeretti, L. (2018): The different roles of proximity in multiple informal network relationships: evidence from the cluster of high technology applied to cultural goods in Tuscany. *Industry and Innovation*, 25 (9), 897-917.

Carias, C. and Klepper S. (2010): Entrepreneurship, The Initial Labor Force, and the Location of New Firms. Paper Presented at the International Schumpeter Society Conference, Aalborg, 2010/0621-24.

Carley, K. (1991): A theory of group stability. *American Sociological Review*, 56 (3), 331-354.

Castaldi, C., Frenken, K. and Los, B. (2015): Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies*, 49 (5), 767-781.

Cattani, G. and Ferriani, S. (2008): A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organization Science*, 19 (6), 924-844.

Cattani, G., Ferriani, S. and Allison, P. D. (2014): Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields: A Core-Periphery Perspective. *American Sociological Review*, 79 (2), 258-281.

Caves, R. E. (2000): *Creative Industries; Contracts Between Art and Commerce*. Cambridge, MA, Harvard University Press.

Cheadle, J. E., Stevens, M., Deadric, T. W. and Bridget, J. G. (2013): The differential contributions of teen drinking homophily to new and existing friendships: An empirical assessment of assortative and proximity selection mechanisms. *Social Science Research*, 42 (5), 1297-1310.

Clauset, A., Arbesman, S. and Larremore, D. B. (2015): Systematic inequality and hierarchy in faculty hiring networks. *Science Advances*, 1, 1, e1400005.

Cohen, W. M., and Levinthal, D. A. (1990): Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly,* 35 (1), 128-153.

Coleman, J. S. (1988): Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95-S120.

Cooke, P. (2002): *Knowledge Economies. Clusters, Learning and Cooperative Advantage*. London, Routledge.

Cowan, R., Jonard, N. and Zimmermann, J.B. (2007): Bilateral collaboration and the emergence of innovation networks. *Management Science*, 53 (7), 1051–1067.

Crespo, J., Suire, R. and Vicente, J. (2014): Lock-in or lock-out? How structural properties of knowledge networks affect regional resilience? *Journal of Economic Geography*, 14 (1), 199-219.

Crespo, J., Suire, R. and Vicente, J. (2016): Network structural properties for cluster long-run dynamics: evidence from collaborative R&D networks in the European mobile phone industry. *Industrial and Corporate Change*, 25 (2), 261-282.

Csáfordi, Zs., Lőrincz, L., Lengyel, B. and Kiss K. M. (2018): Productivity spillovers through labor flows: productivity gap, multinational experience and industry relatedness. *The Journal of Technology Transfer*, 1-36.

Csermely, P., London, A., Wu, L.-Y. and Uzzi, B. (2013): Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1 (2), 93-123.

Cusmano, L. and Morrison, A. (2015): Spin-off and clustering: a return to the Marshallian district. *Cambridge Journal of Economics*, 39 (1), 49-66.

Dahlander, L. and Frederiksen, L. (2012): The Core and Cosmopolitans: A Relational View of Innovation in User Communities. *Organizational Science*, 23 (4), 988-1007.

Dahl, M. S. and Pedersen, C. O. R. (2003): Knowledge flows through informal contacts in industrial clusters. Myth or reality? *Research Policy*, 33 (10), 1673–1686.

Dahl, M. S. and Sorenson, O. (2012): Home Sweet Home: Entrepreneurs' Location Choices and The Performance of Their Ventures. *Management Science*, 58 (6), 1059-1071.

Dahlander, L. and McFarland, D.A. (2013): Ties that last: tie formation and persistence in research collaborations over time. *Administrative Science Quarterly,* 58 (1), 69-110.

Dalmazzo, A. (2002): Technological complexity, wage differentials and unemployment. *Scandinavian Journal of Economics*, 104(4), 515-530.

De Dreu, C.K.W and Weingart, L. R. (2003): Task versus relationship conflict, team performance, and team members satisfaction: a meta-analysis. *Journal of Applied Psychology*, 88 (4), 741-749.

Desrochers, P. (1998): On the Abuse of Patents as Economic Indicators. *The Quarterly Journal of Austrian Economics*, 1(4), 51–74.

Desrochers P. – Leppälä S. (2011): Opening up the 'Jacobs spillovers' black box: local diversity, creativity and the processes underlying new combinations. *Journal of Economic Geography*, 11 (5), 843–863.

De Stefano, D. and Zaccarin, S. (2013): Modelling Multiple Interactions in Science and Technology Networks. *Industry and Innovation*, 20 (3), 221-240.

De Vaan, M., Boschma, R. and Frenken, K. (2013): Clustering and firm performance in project-based industries: the case of the global video game industry, 1972-2007. *Journal of Economic Geography*, 13 (6), 965-991.

De Vaan, M., Stark, M. and Vedres, B. (2015): Game changer: the topology of creativity. *American Journal of Sociology*, 120 (4), 1-51.

Diodato, D., Neffke, F. and O'Clery, N. (2018): Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, 106, 1-26.

Dosi, G. (1982): Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change. *Research Policy,* 11 (3), 147–162.

Duranton, G. and Puga, D. (2004): Micro-foundations of urban agglomeration economies. In Henderson, J. V. and Thisse, J.-F. (eds.): *Handbook of regional and urban economics, Vol 4*. Amsterdam: North-Holland, 2063-2117.

EC European Commission (2013): *A blueprint for the EU forest-based industries*. Brussels, Belgium. Online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX: 52013SC0343&from=EN

Ejermo, O. (2003): Patent diversity as a predictor of regional innovativeness in Sweden. *CESPRI Working Paper*, 140.

Elekes, Z. and Lengyel, B. (2016): Related trade linkages, foreign firms and employment growth in less developed regions. *Papers in Evolutionary Economic Geography*, Nr. 16.20.

Ellison, G., Glaeser, E. L. and Kerr, W. R. (2010): What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100 (3), 1193-1213.

Emmert-Streib, F. and Dehmer, M. (2012): Exploring statistical and population aspects of network complexity. *PLoS ONE*, 7(5).

Eriksson, R. and Lengyel, B. (2019): Co-worker Networks and Agglomeration Externalities. *Economic Geography*, 91 (1), 65-89.

Ethridge, F., Feldman, M., Kemeny, T. and Zoller, T. (2016): The economic value of local social networks. *Journal of Economic Geography*, 16 (5), 1101-1122.

Everett, M.G. and Borgatti, S. (1999): Peripheries of cohesive subsets. *Social Networks*, 21, 397-407.

Everett, M.G. and Valente, T. (2016): Bridging, brokerage and betweenness. *Social Networks*, 44 (1), 202-208

Eurostat (2008): *Nace Rev. 2. Statistical Classification of Economic Activities in the European Community*. Luxembourg, European Communities.

Faulkner, R. R. (1983): *Music on Demand*. Transaction Books, Rutgers, NJ.

Feldman, M., Francis, J. and Bercovitz, J. (2005): Creating a cluster while building a firm: Entrepreneurs and the formation of industrial clusters. *Regional Studies*, 39 (1), 129-141.

Ferriani, S., Fonti, F. and Corrado, R. (2013): The social and economic bases of network multiplexity: Exploring the emergence of multiplex ties. *Strategic Organization*, 11 (1), 7-34.

Fitjar, R.D. and Rodriguez-Pose, A. (2011): When local interaction does not suffice: sources of firm innovation in urban Norway. *Environment and Planning A*, 43 (6), 1248-1267.

Fitjar, R. D. and Timmermans, B. (2019): Relatedness and the Resource Curse: Is There a Liability of Relatedness? *Economic Geography*, 95 (3), 231-255.

Fleming, L. and Sorenson, O. (2001): Technology as a complex adaptive system: evidence from patent data. *Research Policy*, 30 (7), 1019-1039.

Fleming, L., Mingo, S. and Chen, D. (2007): Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly*, 52 (9), 443-475.

Fleming, L. and Waguespack, D. M. (2007): Brokerage, boundary spanning, and leadership in open innovation communities. *Organization Science*, 18 (2), 165-180.

Fornahl, D. and Brenner, T. (2003): *Cooperation, Networks and Institutions in Regional Innovation Systems*. Cheltenham: Edward Elgar.

Fornahl, D., Broekel, T. and Boschma, T. (2011): What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location. *Papers in Regional Science*, 90 (2), 395-418.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. S., Helbing, D., Milojevic, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D. and Barabási, A-L. (2018): Science of science. *Science*, 359 (6379), eaao0185.

Fraiberger, S. P., Sinatra, R., Resch, M., Riedl, C. and Barabási, A.-L. (2018): Quantifying reputation and success in art. *Science*, 362 (6416), 825-829.

Frenken, K., Van Oort, F. and Verburg, T. (2007): Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies*, 41 (4), 685-697.

Galaskiewicz, J. and Shatin, D. (1981): Leadership and networking among neighborhood human service organizations. *Administrative Science Quarterly*, 26 (3), 434-448.

Gertler, M. S. (2003): Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3 (1), 75–99.

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G. and Vanden Oord, A. (2008): Network embeddedness and the exploration of novel technologies. Technological distance, betweenness centrality and density. *Research Policy*, 37 (10), 1717–1731.

Giuffré, K. (1999): Sandpiles of Opportunity: Success in the Art World. *Social Forces*, 77, 815-932.

Giuliani, E. and Bell, M. (2005): The micro-determinants of meso-learning and innovation: evidence from a Chilean wine cluster. *Research Policy*, 34 (1), 47-68.

Giuliani, E. (2007): The selective nature of knowledge networks in clusters: evidence from the wine industry. *Journal of Economic Geography*, 7 (2), 139-168.

Giuliani, E. (2010): Clusters, networks and economic development: an evolutionary economics perspective. In Boschma, R. and Martin, R. (eds.) *The Handbook of Evolutionary Economic Geography*. Cheltenham: Edward Elgar.

Giuliani, E. (2011): Role of technological gatekeepers in the growth of industrial clusters: Evidence from Chile. *Regional Studies*, 45 (10), 1329-1348.

Giuliani, E. (2013): Network dynamics in regional clusters: evidence from Chile. *Research Policy*, 42 (8), 1406-1419.

Giuliani, E., Balland, P.-A. and Matta, A. (2018): Straining but not thriving: understanding network dynamics in underperforming industrial clusters. *Journal of Economic Geography*, 19 (1), 147-172.

Glückler, J. (2007): Economic geography and the evolution of networks. *Journal of Economic Geography*, 7 (5), 619-634.

Gómez-Herrera, E. (2012): Comparing alternative methods to estimate gravity models of bilateral trade. *Empirical Economics*, 44(3), 1087–1111.

Gordon, I. R. and McCann, P. (2000): Industrial Clusters: Complexes, Agglomeration and/or Social Networks? *Urban Studies*, 37 (3), 513-532.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T. and Morris, M. (2008): A statnet tutorial. *Journal of Statistical Software*, 24 (9), 1–26.

Grabher, G. (1993): The weakness of strong ties – the lock-in of regional development in the Ruhr area. In Grabher, G. (ed) *The embedded firm*. London: Routledge. 255-277.

Grabher, G. (2001): Ecologies of creativity: the village, the group, and the heterarchic organisation of the British advertising industry. *Environment and Planning A*, 33 (2), 351-374.

Grabher, G. (2002): The project ecology of advertising: tasks, talents, and teams. *Regional Studies*, 36 (3), 245-262.

Granovetter, M. (1973): The strength of weak ties. *American Journal of Sociology*, 78 (6), 1360-1380.

Granovetter, M. (1985): Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91 (3), 481-510.

Greene, W. H. (1994): Accounting for excess zeros and sample selection in Poisson and negative binomial models. Stern School of Business, New York University, Working Paper 94-10.

Greve, H. R., Baum, J. A. C., Mitsuhashi, H. and Rowley, T. (2010): Built to last but falling apart: cohesion, friction, and withdrawal from interfirm alliances. *Academy of Management Journal,* 53 (2), 302-322.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, 28, 1661–1701.

Halpern, L. and Muraközy, B. (2012): Innovation, productivity and exports: the case of Hungary. *Economics of Innovation and New Technology*, 21 (2), 151-173.

Hanaki, N., Peterhansl, A., Dodds, P. S. and Watts, D. J. (2007): Cooperation in evolving social networks. *Management Science,* 53 (7), 1036-1050.

Hansen, M. T. (1999): The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, 44 (1), 82-111.

Heebels, B. and Boschma, R. (2011): Performing in Dutch Book Publishing 1880-2008. The Importance of Entrepreneurial Experience and the Amsterdam Cluster. *Journal of Economic Geography*, 11 (6), 1007-1029.

Hidalgo, C. A., Klinger, B., Barabási, A. L. and Hausmann, R. (2007): The product space conditions the development of nations. *Science*, 317 (5837), 482.

Hidalgo, C. A. (2009): The dynamics of economic complexity and the product space over a 42 year period. *CID Working Papers*, 189.

Hidalgo, C. A. and Hausmann, R. (2009): The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106 (26), 10570-10575.

Hidalgo, C. A. (2015): *Why Information Grows: The Evolution of Order, from Atoms to Economies*. Basic Books, New York.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S. and Zhu, S. (2018): The Principle of Relatedness. *International Conference on Complex Systems*, Springer, Cham, 451-457.

Hojman, D.A. and Szeidl, A. (2008): Core and periphery in networks. *Journal of Economic Theory*, 139, 295-309.

Holbrook, M. B. (1999): Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research*, 26 (2), 144-155.

Huber, F. (2012): On the role of interrelationship of spatial, social and cognitive proximity: personal knowledge relationships of R&D workers in the Cambridge Information Technology Cluster. *Regional Studies*, 46 (9), 1169-1182.

Hunter, D. R., Goodreau, S. M. and Handcock, M. S. (2008): Goodness of fit for social network models. *Journal of the American Statistical Association*, 103 (481), 248–258.

Iammarino, S. and McCann, P. (2006): The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. *Research Policy*, 35 (7), 1018-1036.

Ibert, O. (2004): Projects and firms as discordant complements. Organizational learning within the Munich software ecology. *Research Policy*, 33 (10), 1529-1546.

Jackson, M. O. and Wolinsky, A. (1996): A strategic model of social and economic networks. *Journal of Economic Theory*, 71 (1), 44-74.

Jackson, M. O. (2008): *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jacobs, J. (1969): *The Economy of Cities*. Vintage, New York, NY.

Jaffe, A.B., Trajtenberg, M. and Henderson, R. (1993): Geographic localization and knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics*, 108 (3), 577–598.

Johns, J. (2006): Video games production networks: value capture, power relations and embeddedness. *Journal of Economic Geography*, 6 (2), 151-180.

Jones, C. (1996): Careers in project-networks: The case of film industry. In Arthur, M. B. – Russeau, D. M. (eds.): *The Boundaryless Career*. Oxford University Press, Oxford, UK.

Juhász, S. and Lengyel, B. (2018): Creation and persistence of ties in cluster knowledge networks. *Journal of Economic Geography*, 18 (6), 1203-1226.

Klepper, S. (2007): Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry. *Management Science*, 53 (4), 616-631.

Klepper, S. (2009): Spinoffs: A review and synthesis. *European Management Review*, 6 (3), 159-171.

Klepper, S. (2010): The origin and growth of industry clusters: the making of Silicon Valley and Detroit. *Journal of Urban Economics,* 67 (1), 15-32.

Knoben, J. and Oerlemans, L. A. G. (2006): Proximity and Inter-Organizational Collaboration: A Literature Review. *International Journal of Management Reviews*, 8 (2), 71-89.

Kogler, D. F., Rigby, D. L. and Tucker, I. (2013): Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, 21 (9), 1374–1391.

Kojaku, S. and Masuda, N. (2018): Core-periphery structure requires something else in the network. *New Journal of Physics,* 20, 043012.

König, M. D., Tessone, C. J. and Zenou, Y. (2010): From assortative to dissortative networks: the role of capacity constraints. *Advances in Complex Systems*, 13 (4), 483-499.

Kossinets, G. and Watts, D. J. (2009): Origins of homophily in an evolving social network. *American Journal of Sociology*, 115 (2), 405-450.

Krafft, J., Quatraro, F. and Saviotti, P.P. (2011): The knowledge-base evolution in biotechnology: a social network analysis. *Economics of Innovation and New Technology*, 20(5), 445-475.

Krafft, J., Quatraro, F. and Saviotti, P.P. (2014): The dynamics of knowledge-intensive sectors' knowledge base. Evidence from biotechnology and telecommunications. *Industry and Innovation*, 21(3), 215-242.

Krugman, P. (1991): *Geography and Trade.* MIT Press, Cambridge, MA.

Kogler, D., Rigby, D. and Tucker, I. (2013): Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, 21 (9), 1374-1391.

Lambert, D. M. and Harrington, T. C. (1990): Measuring nonresponse in customer service mail surveys. *Journal of Business Logistics*, 11 (2), 5-25.

Lambiotte, R., Blondel, V. D., de Kerchove, C., Huens, E., Prieur, C., Smoreda, Z. and Van Dooren, P. (2008): Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387 (21), 5317-5325.

Lane, P. J. and Lubatkin, M. (1998): Relative absorptive capacity and interorganizational learning. *Strategic Management Journal*, 19 (5), 461-477.

Lavie, D. and Rosenkopf, L. (2006): Balancing exploration and exploitation in alliance formation. *Academy of Management Journal*, 49 (4), 797-818.

Lazzeretti, L. and Capone, F. (2016): How proximity matters in innovation network dynamics along with the cluster evolution. A study of the high technology applied to cultural goods. *Journal of Business Research*, 69 (12), 5855-5865.

Lengyel, B. and Leydesdorff, L. (2011): Regional innovation systems in Hungary: The failing synergy at the national level. *Regional Studies*, 45 (5), 677-693.

Lengyel, B., Varga, A., Ságvári, B., Jakobi, A. and Kertész, J. (2015): Geographies of an online social network. *PLOS ONE*, 10: e0137248.

Lengyel, B. and Eriksson, R. (2017): Co-worker networks, labour mobility and productivity growth in regions. *Journal of Economic Geography*, 17 (3), 635–660.

Lengyel, I., Vas, Zs., Szakálné Kanó, I. and Lengyel, B. (2018): Spatial differences of reindustrialization in post-socialist economy: manufacturing in the Hungarian counties. *European Planning Studies*, 25 (8), 1416-1434.

Levinthal, D. A. and March, J. G. (1993): The myopia of learning. *Strategic Management Journal,* 14 (S2), 95-112.

Li, P., Bathelt, H. and Wang, J. (2012): Network dynamics and cluster evolution: changing trajectories of the aluminium extrusion industry in Dali, China. *Journal of Economic Geography*, 12 (1), 127-155.

Lin, N. (2002): *Social capital: a theory of social structure and action*. Cambridge University Press, Cambridge, UK.

Liu, L., Wang, Y., Sinatra, R., Giles, C.G., Song, C. and Wang, D. (2018): Hot streaks in artistic, cultural, and scientific careers. *Nature*, 559 (7714), 396-399.

Lusher, D., Koskinen, J. and Robins, G. (2013): *Exponential Random Graph Models for Social Networks: Theory, Methods and Applications*. New York: Cambridge University Press.

Malmberg, A. (1997): Industry geography: location and learning. *Progress in Human Geography*, 21 (4), 573-582.

March, J. G. (1991): Exploration and exploitation in organizational learning. *Organization Science*, 2 (1), 71-87.

Marmaros, D. and Sacerdote, B. (2006): How do friendships form? *Quarterly Journal of Economics,* 121 (1), 79-119.

Marshall, A. (1920): *Principles of Economics*. MacMillan, London, UK.

Martin, R. and Sunley, P. (2011): Conceptualizing cluster evolution: beyond the life cycle model? *Regional Studies*, 45 (10), 1299-1318.

Maskell, P. and Malmberg, A. (1999): Localised learning and industrial competitiveness. *Cambridge Journal of Economics*, 23 (2), 167-185.

Mcnerney, J., Farmer, J. D., Redner, S. and Trancik, J. E. (2011): Role of design complexity in technology improvement. *Proceedings of the National Academy of Sciences*, 108 (22), 9008-9013.

McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001): Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444.

Menzel, M.P. (2015): Interrelating Dynamic Proximities by Bridging, Reducing and Producing Distances. *Regional Studies*, 49(11), 1892-1907.

Menzel, M. P. and Fornahl, D. (2010): Cluster life cycles – dimensions and rationales of cluster evolution. *Industrial and Corporate Change*, 19 (1), 205-238.

Mewes, L. (2019): Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836-2010. *Economic Geography*, doi: 10.1080/00130095.2019.1567261

Molina-Morales, F. X., Belso-Martinez, J. A., Mas-Verdú, F. and Martinez-Cháfer, L. (2015): Formation and Dissolution of Inter-Firm Linkages in Lengthy and Stable Networks in Clusters. *Journal of Business Research*, 68 (7), 1557-1562.

Morgan, A. C., Economou, D. J., Way, S. F. and Clauset, A. (2018): Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science*, 7 (40), 1-16.

Morrison, A. (2008): Gatekeepers of knowledge within industrial districts: who they are, how they interact. *Regional Studies*, 42 (6), 817-835.

Morrison, A. and Rabellotti, R. (2009): Knowledge and information networks in an Italian wine cluster. *European Planning Studies*, 17 (7), 983-1006.

Morrison, A., Rabellotti, R. and Zirulia, L. (2013): When do global pipelines enhance the diffusion of knowledge in clusters? *Economic Geography*, 89 (1), 77-96.

Morrison, A. and Boschma, R. (2019): The spatial evolution of the Italian motorcycle industry (1893-1993): Klepper's heritage theory revisited. *Industrial and Corporate Change,* 28 (3), 613-634.

Neffke, F. (2009): *Productive Places: The influence of technological change and relatedness on agglomeration externalities*. Doctoral Thesis.

Neffke, F., Henning, M. and Boschma, R. (2011): How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, 87(3), 237–265.

Neffke, F.M.H., Henning, M. and Boschma, R. (2012): The impact of aging and technological relatedness on agglomeration externalities: a survival analysis. *Journal of Economic Geography*, 12 (2), 485-517.

Nelson, R. and Winter, S. (1982): *An evolutionary theory of economic change*. Belknap, Cambridge, MA.

Nooteboom, B. (2000): *Learning and Innovation in Organizations and Economies*. Oxford University Press, Oxford.

Obstfeld, D. (2005): Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly*, 50 (1), 100-130.

Owen-Smith, J. and 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.

Parra-Requena, G., Molina-Morales, F. X. and Garcia-Villaverde, P. M. (2010): The mediating effect of cognitive social capital on knowledge acquisition in clustered firms. *Growth and Change*, 41 (1), 59-84.

Peterson, R. A. and Anand, N. (2004): The Production of Culture Perspective. *Annual Review of Sociology*, 15 (30), 311-334.

Perretti, F. and Negro, G. (2007): Mixing genres, matching people: a study in innovation and team composition in Hollywood. *Journal of Organizational Behavior*, 28, 563-586.

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R. and Hidalgo, C. (2018): Shooting Low o High: Do Countries Benefit from Entering Unrelated Activities? *Papers in Evolutionary Economic Geography*, 18 (7).

Porter, M. E. (1990): *The Competitive Advantage of Nations*. London, Macmillan.

Porter, M. E. (2003): The economic performance of regions. *Regional Studies*, 37 (6-7), 545-546.

Powell, W.W., Koput, K.W. and Smith-Doerr, L. (1996): Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. *Administrative Science Quarterly*, 41 (1), 116-145.

Powell, W.W., White, D.R., Koput, K.W. and Owen-Smith, J. (2005): Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110 (4), 1132-1205.

Putnam, R. D. (2001): *Bowling alone: The collapse and revival of American community*. Simon and Schuster, New York.

Qian, H. (2018): Knowledge-Based Regional Economic Development: A Synthetic Review of Knowledge Spillovers, Entrepreneurship, and Entrepreneurial Ecosystems. *Economic Development Quarterly*, 32 (2), 163-176.

Radosevic, S. (2002): Regional innovation systems in Central and Eastern Europe: Determinants, organizers and alignments. *The Journal of Technology Transfer*, 27 (1), 87-96.

Rand, D.G., Arbersman, S. and Christakis, N.A. (2011): Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, 108 (48), 19193-19198.

Reagans, R. and McEvily, B. (2003): Network structure and knowledge transfer: the effects of cohesion and range. *Administrative Science Quarterly*, 48 (2), 240-267.

Resmini, L. (2007): Regional patterns of industry location in transition countries: does economic integration with the European Union matter? *Regional Studies*, 41 (6), 747-764.

Rigby, D. L. (2015): Technological relatedness and knowledge space: entry and exit of US cities from patent classes. *Regional Studies*, 49 (11), 1922-1937.

Rigby, D. L. and Brown, W. M. (2015): Who benefits from agglomeration. *Regional Studies*, 49 (1), 28-43.

Ripley, R., Snijders, T., Boda, Zs., Vörös, A. and Preciado, P. (2017): *Manual for RSiena (Version September 9)*. University of Oxford: Department of Statistics, Nuffield College University of Groningen: Department of Sociology.

Rivera, M.T., Soderstrom, S. B. and Uzzi, B. (2010): Dynamics of dyads in social networks: assortative, relational, and proximity mechanisms. *Annual Review of Sociology,* 36, 91-115.

Robins, R., Snijders, T., Wang, P., Handcock, M. and Pattison, P. (2007a): Recent developments in the exponential random graph (p*) models for social networks. *Social Networks*, 29 (2), 192–215.

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

Rosenkopf, L. and Padula, G. (2008): Investigating the microstructure of network evolution: alliance formation in the mobile communications industry. *Organization Science*, 19 (5), 669-687.

Rosenthal, S. S. and Strange, W. C. (2004): Evidence on the nature and source of agglomeration economies. In Henderson, V. and Thisse, J-F. (eds.) *Handbook of Regional and Urban Economics*. Amsterdam, North-Holland, 2119-2171.

Saxenian, A. (1994): *Regional Advantage Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge, MA.

Scherngell, T. and Barber, M. J. (2009): Spatial interaction modeling of cross-region R&D collaboration. Empirical Evidence from the 5th EU Framework Programme. *Papers in Regional Science*, 88(3), 531–546.

Schilling, M. A. (2005): A "small-world" network model of cognitive insight. *Creativity Research Journal*, 17 (2-3), 131-154.

Scott, A. J. (2000): *The Cultural Economy of Cities; Essays on the Geography of Image-Producing Industries*. London, Sage.

Scott, A. J. (2004): Cultural-products industries and urban economic development. *Urban Affairs Review*, 39 (4), 461-490.

Seabright, M.A., Levinthal, D.A. and Fichman, M. (1992): Role of individual attachments in the dissolution of interorganizational relationships. *The Academy of Management Journal,* 35 (1), 122-160.

Shipilov, A. V., Rowley, T. J. and Aharonson, B. S. (2006): When Do Networks Matter? A Study of Tie Formation and Decay. In Baum, J. A. C., Dobrev, S. D. and Van Witteloostuijn, A. (eds) *Ecology and Strategy, Advances in Strategic Management, Volume 23.* Emerald Group Publishing Limited.

Simon, H. A. (1962): Thee architecture of complexity. *Proceedings of the American Philosophical Society*, 106 (6), 467-482.

Simonton, D. K. (2004): Film awards as indicators of cinematic creativity and achievements: A quantitative comparison of the Oscars and six alternatives. *Creativity Research Journal*, 16 (4), 163-172.

Simonton, D. K. (2009): Cinematic Success Criteria and Their Predictors: The Art and Business of the Film Industry. *Psychology & Marketing*, 26 (5), 400-420.

Snijders, T. (2001:) The statistical evaluation of social network dynamics. *Sociological Methodology*, 31 (1), 361-395.

Snijders, T. (2002): Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure*, 3 (2), 1-40.

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

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

Sorenson, O. and Audia, P. G. (2000): The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940-1989. *American Journal of Sociology*, 106 (2), 424-462.

Sorenson, O. (2003): Social networks and industrial geography. *Journal of Evolutionary Economics*, 13 (1), 513-527.

Sorenson, O. and Fleming, L. (2004): Science and the diffusion of knowledge. *Research Policy*, 33 (10), 1615-1634.

Spitz, A. and Horvát, E.-Á. (2014): Measuring long-term impact based on network centrality: Unraveling cinematic citations. *PloS ONE*, 9 (10), 108857.

Staber, U. (2011): Partners forever? An empirical study of relational ties in two small-firm clusters. *Urban Studies*, 48 (2), 235-252.

Stark, D. (2009): *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton: Princeton University Press.

Steglich, C. E. G., Snijders, T. A. B. and Pearson, M. (2010): Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 40 (1), 329-393.

Stiller, J and Dunbar, R.I.M. (2007): Perspective-taking and mamory capacity predict social network size. *Social Networks*, 29 (1), 93-104.

Storper, M. and Venables, A. J. (2004): Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, 4 (4), 351-370.

Suire, R. and Vicente, J. (2014): Clusters for life of life cycles of clusters: in search of the critical factors of clusters' resilience. *Entrepreneurship and Regional Development*, 26 (1-2), 142-164.

Sydow, J. and Staber, U. (2002): The institutional embeddedness of project networks: the case of content production in German television. *Regional Studies*, 36 (3), 215-277.

Tallman, S., Jenkins, M., Henry, N. and Pinch, S. (2004): Knowledge, clusters and competitive advantage. *Academy of Management Review*, 29 (2), 258-271.

Teece, D. J., Rumelt, R., Dosi, G. and Winter, S. G. (1994): Understanding corporate coherence: theory and evidence. *Journal of Economic Behaviour Organisation*, 23, 1-30.

Ter Wal, A. L. J. and Boschma, R. (2009): Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, 43 (3), 739-756.

Ter Wal, A. L. J. and Boschma, R. (2011): Co-evolution of firms, industries and networks in space. *Regional Studies*, 45 (7), 919-933.

Ter Wal, A. L. J. (2013): Cluster emergence and network evolution: a longitudinal analysis of the inventor network in Sophia-Antipolis. *Regional Studies*, 47 (5), 651-668.

Ter Wal, A. L. J. (2014): The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. *Journal of Economic Geography*, 14 (3), 589–620.

Uzzi, B. (1997): Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42 (1), 35-67.

Uzzi, B. and Spiro, J. (2005) Collaboration and creativity: The small world problem. *American Journal of Sociology*, 111 (2), 447-504.

Van Duin, M. A. J., Gille, K. J. and Handcock, M. S. (2009): A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31 (1), 52–62.

van Eck, N. J. and Waltman, L. (2009): How to normalize co-occurence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635–1651.

Vedres, B. (2017): Forbidden triads and creative success in jazz: the Miles Davis factor. *Applied Network Science*, 2 (31), 1-25.

Verspagen, B. and Duysters, G. (2004): The small world of strategic technology alliances. *Technovation*, 24 (7), 563-571.

Vicente, J., Balland, P.-A. and Brossard, O. (2011): Getting into Networks and Clusters: Evidence from the Midi-Pyrenean Global Navigation Satellite Systems (GNSS) Collaboration Network. *Regional Studies*, 45 (8), 1059-1078.

Wasserman, S. and Faust, K. (1994): *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Watts, D. J. and Strogatz, S. H. (1998): Collective dynamics of 'small-world' networks. *Nature*, 393, 440-442.

Wellman, B. (1996): Are personal communities local? A Dumptarian reconsideration. *Social Networks*, 18 (4), 347-354.

Wenting, R. (2008): Spinoff Dynamics and The Spatial Formation of the Fashion Design Industry, 1858-2005. *Journal of Economic Geography*, 8 (5), 593-614.

Wimmer, A. and Lewis, K. (2010): Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology*, 116 (2), 583-642.

Wuchty, S., Jones, B. F. and Uzzi, B. (2007): The increasing dominance of teams in production of knowledge. *Science*, 316 (5827), 1036-1039.

Zaheer, A. and Soda, G. (2009): Network Evolution: The Origins of Structural Holes. *Administrative Science Quarterly*, 42 (1), 1-31.

Acknowledgements

First, I would like to acknowledge the help of my supervisors, Ron Boschma, Andrea Morrison and Pierre-Alexandre Balland. Thank you for believing in me, I will never forget the inspiring conversations we had. I am also thankful for their helpful comments and suggestions to all the chapters including Chapter 1 Introduction and Chapter 6 Conclusion of this thesis.
Second, I would also like to acknowledge the vital role of my co-authors Balázs Lengyel, Tom Broekel and Gergő Tóth, from whom I have learned so much and who made this process enjoyable.
Third, I feel fortunate to have met such wonderful colleagues in the ANET Lab and at the University of Szeged. In particular, Zoltán Elekes, János Gyurkovics, Zsófia Vas, Eszter Bokányi, Zsófia Zádor, László Czaller and Johannes Wachs were always there to support me.
Finally, I would like to acknowledge the continuous support and endless patience of my wife Zsófi, my parents, my brother and my whole family.

Nederlandse samenvatting

Netwerken zijn essentiële hulpmiddelen gebleken voor het coderen en inzichtelijk maken hoe verschillende delen van de samenleving met elkaar verbonden raken. Hoewel wordt betoogd dat geografische nabijheid interactie en samenwerking bevordert, betekent dat niet automatisch dat alle nabijgelegen actoren met elkaar verbonden zullen zijn. Doordat handelende partijen wat capaciteiten betreft heterogeen zijn, maken zij niet eenvoudig verbinding met elkaar en leren zij evenmin makkelijk van elkaar. In dit onderzoek wordt de rol benadrukt die geografie speelt bij de vorming van netwerkverbindingen die ideeën, kennis en technologie kanaliseren. Het streven is om de relevante literatuur uit te breiden door nieuwe aspecten te introduceren als mogelijke drijvende krachten achter, en resultaten van, ruimtelijke netwerkvorming.

Dit proefschrift is opgebouwd rondom vier empirische casussen, elk met betrekking op een andere netwerkomgeving, met haar eigen omvang, complexiteit en temporele dimensies. In hoofdstuk 2 wordt de vraag gesteld of de achtergrond van spin-offs van belang is voor de vorming van netwerkverbindingen in industriële clusters. Daar worden de eerste stappen gezet om twee stromingen uit de literatuur met elkaar te combineren, namelijk de stroming waarin het belang van kennisnetwerkverbindingen voor de exploitatie van positieve agglomeratie-externaliteiten in industriële clusters wordt benadrukt en de stroming waarin het belang van spin-offbedrijven voor het in stand blijven en de welvaart van clusters wordt uitgelicht. In dit hoofdstuk wordt aangetoond dat na controle voor bedrijfskenmerken, geografische nabijheid en cognitieve nabijheid van bedrijven en de structurele eigenschappen van netwerken, of een bedrijf al dan niet een spin-off is nog steeds bepalend is voor de effectiviteit waarmee bedrijven verbindingen maken binnen clusternetwerken.

In hoofdstuk 3 wordt een nieuwe benadering gepresenteerd voor het bestuderen van de drijvende krachten achter de vorming van kennisnetwerken, waarbij onderscheid wordt gemaakt tussen de vorming en het in stand blijven van verbindingen. De algemene gedachtegang is dat onderscheid maken tussen enerzijds de vorming en anderzijds het in stand blijven van verbindingen ons helpt bij het identificeren van factoren die bepalend zijn voor de ontwikkeling van clusters in de loop van de tijd. De bevindingen laten zien dat triadische verbanden tussen verbindingen, geografische nabijheid en cognitieve nabijheid van bedrijven alsmede de interactie daartussen de vorming en het in stand blijven van verbindingen op verschillende manieren beïnvloeden. Daarnaast wordt door middel van dit onderscheid aangetoond dat factoren met betrekking tot selectie, retentie en variatie ook aanwezig zijn tijdens de evolutie van clusterkennisnetwerken.

In hoofdstuk 4 wordt de literatuur over de vorming van netwerkverbindingen verder toegepast op het complexe netwerk van technologieën. Er wordt een bijdrage geleverd aan

het bloeiende onderzoeksgebied rondom technologische verwantschap door de gangbare redenering om te keren en verwantschap als afhankelijke variabele te gebruiken. In dit hoofdstuk wordt aangetoond dat de co-locatie en complexiteit van technologieën bepalend zijn voor hun verwantschap. Daarnaast wordt aangetoond dat menselijke samenwerking mediërend werkt op de invloed van co-locatie en complexiteit op zowel de aanwezigheid als de sterkte van verwantschapsverbindingen.

Het laatste empirische hoofdstuk heeft als doel de literatuur over agglomeratie, de vorming van sociale netwerken en creatieve prestaties met elkaar te combineren. In hoofdstuk 5 wordt beschreven hoe individueel succes in een stedelijke creatieve industrie afhankelijk is van netwerkverbindingen. Er wordt zichtbaar gemaakt dat aanwezigheid in de kern van de industrie en de vorming van verbindingen tussen de kern en de periferie van een samenwerkingsnetwerk een significante positieve bijdrage leveren aan individueel creatief succes. Daarnaast toont dit onderzoek een verband aan tussen de vorming van structureel belangrijke verbindingen en succes in een stedelijke agglomeratie van creatieve industrieën.

In zijn geheel worden in de dissertatie vier nieuwe aspecten geïntroduceerd om te overwegen bij het bestuderen van ruimtelijke netwerken, teneinde meer inzicht te krijgen in de innovativiteit en topprestaties van economische actoren. Door op zoek te gaan naar nieuwe drijvende krachten en benaderingen om de vorming van netwerkverbindingen te verklaren versterkt de dissertatie de redenering dat geografische nabijheid van belang is voor kennisuitwisseling en technologische combinaties.

Curriculum Vitae

Sándor Juhász was born in Kecskemét, Hungary on September 7, 1989. He holds a Bachelor degree in Business Administration and Management and a Master degree in Regional and Environmental Economic Studies from the University of Szeged.

Before the start of his Ph.D., he worked as research assistant in the Institute of Economics and Economic Development at the University of Szeged. The research conducted during his time in Szeged resulted several scientific writings in Hungarian language. In 2017, he joined the Agglomeration and Social Networks Lendület Research Group at the Hungarian Academy of Sciences as a junior researcher. He finishes his PhD with two academic publication in English, both part of his thesis. "*Creation and persistence of ties in cluster knowledge networks*" was published in Journal of Economic Geography, in 2018, while '*Spinoffs and tie formation in cluster knowledge networks*" was published in Small Business Economics in 2019.