Dissertation im Rahmen des Doktorats der Sozial- und Wirtschaftswissenschaften an der Wirtschaftsuniversität Wien

Regional knowledge creation and R&D collaboration in Europe

Martina Neuländtner

Erster Beurteiler:

Priv. Doz. Dr. Thomas Scherngell

Zweiter Beurteiler:

Univ. Prof. Dr. Jürgen Essletzbichler

Institut für Wirtschaftsgeographie und Geoinformatik

Welthandelsplatz 1, 1020 Wien

Abstract

Knowledge creation is the essential fundament for innovative regional activities, and therefore, is widely acknowledged as the driving force for regional socio-economic development. Hence, modelling the complexities of regional knowledge creation processes and analysing their determinants stimulates current scientific debates in the field of economic geography. Increasingly, particular interest is drawn to the role of inter-regional R&D networks to create, access and diffuse knowledge. In the vein of this research, the dissertation takes on new perspectives to advance the understanding of regional knowledge creation and R&D networks and their interplay – while particularly accounting for different kinds of heterogeneity.

The overall aim of this dissertation is to *identify and systematically characterise how R&D networks drive regional knowledge creation*, accounting for (i) *technological heterogeneity* expressing technology-specific forms of knowledge creation, (ii) *heterogeneity in modes of knowledge creation* and *knowledge output* that refers to specific characteristics of the knowledge creation processes, and (iii) *heterogeneity of research actors*, reflected by, e.g. actor-specific knowledge endowments, collaboration rationales and research interests.

Given the embedding in the state-of-the-art literature, the explicit methodological focus, and the strong empirical focus, the dissertation substantially contributes to the current scientific debate in economic geography, particularly to the stream exploring the geography of innovation. *First*, the dissertation provides statistical evidence that network connectivity is able to compensate for geographical barriers to R&D collaboration, though this compensation effect differs in magnitude across technological fields. *Second*, the results confirm a generally positive impact of R&D networks on regional knowledge creation, but for the first time unveil important differences across different modes of knowledge creation; network embeddedness is particularly important for knowledge creation in science-based fields. *Third*, spatial spillovers of network effects arise, i.e. being spatially proximate to highly networked regions is conducive for a region’s knowledge creation capability, in particular for more explorative modes of knowledge creation. *Fourth*, a novel empirical agent-based simulation model of regional knowledge creation demonstrates the potential of taking a simulation approach to entering new grounds in the investigation of regional knowledge production mechanisms.

Kurzfassung

Die Schaffung neuen Wissens gilt als unabdingbare Voraussetzung für Innovationstätigkeit und als eine treibende Kraft für regionale sozioökonomische Entwicklung. Die Modellierung komplexer regionaler Prozesse zur Schaffung von Wissen sowie die Analyse ihrer Determinanten stehen daher im Zentrum aktueller Fragestellungen der Wirtschaftsgeographie. Diesbezüglich steigt zunehmend das Interesse an interregionalen Forschungs- und Entwicklungsnetzwerken (F&E-Netzwerke) und deren Rolle für die Schaffung, den Zugang zu neuem Wissen und dessen Verbreitung. Diese Dissertation folgt diesem Forschungsstrang und nimmt neue Perspektiven hinsichtlich regionaler Wissensgenerierung und F&E-Netzwerke sowie zu deren Zusammenspiel ein. Dabei werden insbesondere Unterschiede aufgrund verschiedener Arten von Heterogenität berücksichtigt und hervorgehoben.

Das Ziel dieser Dissertation ist die *Identifikation und die systematische Charakterisierung der Rolle von F&E-Netzwerken für die regionale Wissensgenerierung*, wobei im Wesentlichen drei Arten von Heterogenität berücksichtig werden: Heterogenität (i) in *Technologien*, (ii) bei *Modi der Wissensgenerierung* und (iii) von *Forschungsakteuren* (u.a. hinsichtlich ihrer Wissensausstattung und Forschungsstrategien). Die Dissertation ist durch einen expliziten methodischen und empirischen Fokus gekennzeichnet und in den aktuellen Stand der Forschung eingebettet. Sie kann daher wesentlich zum aktuellen wissenschaftlichen Diskurs der Wirtschaftsgeographie, insbesondere im Bereich der *Geography of Innovation*, beitragen.

Erstens zeigt die Dissertation auf, dass Netzwerkverbindungen dazu beitragen können, geografische Barrieren bei der Generierung neuer F&E-Netzwerkbeziehungen zu überwinden. *Zweitens* deuten die Ergebnisse auf einen generellen positiven Effekt von F&E-Netzwerken auf die regionale Wissensschaffung hin. Dieser Zusammenhang zeigt sich besonders deutlich bei der Generierung von Wissen im wissenschaftlichen Bereich. *Drittens* sind für explorative Wissensgenerierung zusätzlich positive externe Effekte – *räumliche spillover* – durch Netzwerkverbindungen benachbarter Regionen beobachtbar. *Viertens* zeigt die Dissertation – mittels eines empirischen agentenbasierten Modells zur Simulation interregionaler Wissensgenerierung – neue Wege zur Analyse von regionalen Wissensgenerierungsmechanismen auf.

Acknowledgements

Being aware of the enforcing and accelerating role of networks for research and innovation activities, networks have also served as a locus for creating something new throughout the writing of this dissertation. I consider myself very grateful to be embedded in a network of colleagues, family and friends from whom I have received a great deal of assistance and support.

First of all, I would like to express my deepest gratitude to my supervisor Thomas Scherngell for his unwavering support and guidance. I have very much benefited from his constructive criticism, constant advise and appreciative joint scientific discussions, for which I am sincerely thankful. His positive attitude and belief in my work inspired me with confidence during more cumbersome and challenging times. Moreover, I would like to extend my gratitude to my committee members, especially to my co-supervisor, Jürgen Essletzbichler; their valuable and helpful suggestions during the proposal phase and their scientific and professional assistance is very much appreciated.

Furthermore, I would like to thank Matthias Weber, Head of Center for Innovation Systems & Policy at the AIT Austrian Institute of Technology, for the institutional support and the opportunity to write this dissertation surrounded by skilled and ambitious colleagues, making it an invaluable experience that I do not take as granted. Especially, I am deeply indebted to my colleagues Barbara Heller-Schuh, Georg Zahradnik, Astrid Unger, Manfred Paier, Michael Barber and Xheneta Bilalli Shkodra. Without their undeniable passion for data, modelling, and programming (as well as cake and ice cream), writing this dissertation would have been much more difficult.

Finally, I am very thankful to my family and longtime friends who have always been with a sympathetic ear. Especially, I owe deepest gratitude and admiration to my husband Johannes for his unconditional support and understanding.

Contents

List of Figures

List of Tables

1 Introduction

New knowledge resulting from Research and Development (R&D) activities is widely considered a key driver of productivity and economic growth (see, e.g. Romer 1990, Lucas 1988). In this context, the creation and accumulation of new knowledge and technologies are understood as a prerequisite for innovation and hence, constitutes a crucial factor for organisations, regions and countries to be resilient and competitive in times of converging technologies and increasing market pressure (see, e.g. Fischer 2001, Boschma 2004). In the vein of economic geography and regional science, since the early 1990s, scholars have stressed the geographical distribution of new knowledge as one central element to explain the divergent economic and socio-economic development of regions and countries (see, e.g. Feldman 1994, Grossman and Helpman 1994). For a long time, the presumably local character of knowledge creation and its spatial distribution dominated the debates in economic geography (Asheim 1996, Markusen 1996, Porter 1998), emphasising the role of spatial proximity in the process of creating and diffusing new knowledge, often described by the notion of regional innovation systems (Cooke 2001, Asheim and Isaksen 2002, Tödtling and Trippl 2005). More recently, knowledge creation has been increasingly seen as the result of interactive, collaborative learning processes among actors of different types located in other geographical spaces (regions). Especially in a primarily knowledge-based economy, collaborative knowledge creation is gaining importance, not only within, but increasingly also across regions (see, e.g. Maggioni et al. 2007, Hoekman et al. 2010, Scherngell and Lata 2013).

Consequently, attention has been shifted to the investigation and modelling of regional knowledge creation processes focusing on inter-regional knowledge interactions in the form of R&D collaborations. In the process of collaborative knowledge creation, networks of R&D relationships between R&D actors in science and industry enable knowledge flows between these actors, as well as access to external, new sources of knowledge that are most often located further away in geographical space (for an overview see Scherngell 2013). Accessing region-external knowledge sources may allow incorporating new knowledge components and capabilities into intra-regional knowledge creation processes, accordingly enhancing and diversifying the existing regional knowledge base (e.g. Bathelt et al. 2004). Potentially, this reduces regional disparities regarding the innovatory potential and may prevent technological lock-ins by an increased technological diversification within the region (see, e.g. Boschma and Ter Wal 2007).

In the recent past, specifically, the relationship between regional knowledge creation and inter-regional R&D networks has attracted increasing interest (e.g. Varga et al. 2014, Wanzenböck and Piribauer 2018). In these studies, the focus lies on investigating determinants of knowledge creation while explicitly accounting for the role of R&D networks. In a slightly different vein, the networked character of knowledge creation is acknowledged by analysing drivers and barriers for the constitution of R&D networks (e.g. Scherngell and Barber 2009, Hoeckman et al. 2010). Whereas, these studies predominantly focus on geographical barriers, recent works aim to advances this perspective by a relational aspect investigating network structural effects as drivers for R&D network creation (e.g. Bergé 2017).

Overall, these studies provide initial systematic evidence – using novel datasets to observe R&D networks at the regional level – on the importance of inter-regional R&D network linkages and a region's network position for knowledge creation. However, they fall short in three essential aspects: they (i) neglect potential idiosyncrasies of network effects across different technological domains, (ii) consider knowledge creation as a homogeneous process, not differentiating between different modes of knowledge creation and types of knowledge output, and (iii) capture knowledge creation as an aggregate of knowledge creation processes on the regional level, not accounting for heterogeneity in knowledge creation processes on an actor-level, e.g. in terms of their knowledge endowment and research strategies. These gaps constitute the main entry points for this dissertation.

The objective of the dissertation and focus of the research articles

Against this background, the dissertation aims *to identify and systematically characterise how R&D networks drive regional knowledge creation, accounting for different kinds of heterogeneity*. Considering various forms of heterogeneity in the analysis of regional knowledge creation promises an advanced understanding of the complex knowledge creation process per se, and the forces driving divergent regional innovative capabilities and accordingly, inter-regional economic and socio-economic disparities. Specifically, the dissertation contributes to the understanding of regional knowledge creation and R&D collaboration networks by focusing on heterogeneity reflected in different *technologies*, *modes of knowledge creation* and *knowledge outputs*, and knowledge-specific *characteristics of R&D actors*. The various aspects of such heterogeneities that shape inter-regional collaborative knowledge creation processes in different ways constitute the entry points for the four research articles. Moreover, to adopt different perspectives on regional knowledge creation, three different methodological approaches are employed in the research articles, namely spatial interaction modelling (SIM), econometric modelling (non-spatial and spatial), and agent-based modelling (ABM). Moreover, all four articles feature an explicit network-analytical angle and rely on Social Network Analysis (SNA) concepts.

Article I and **Article II** initially examine two different angles of *technological heterogeneity,* reflecting technology-specific forms of knowledge creation. While the first article investigates determinants of distinct collaboration patterns of different technological R&D collaboration networks, the second article compares the effects of such networks for various technological fields. In both cases, explicitly accounting for diversity in technological conditions, on the one hand, overcomes the statistical issue of unobserved heterogeneity, but on the other hand, also allows for novel insights on the importance of selected determinants of knowledge interactions for specific technologies, given the different knowledge bases and rationales of collaboration. The **aim** of **Article I** is to *identify determinants of the knowledge interactions within technology-specific inter-regional R&D networks*. Specifically, it is analysed how geographical and relational distance between regions influence the inter-regional knowledge flows within R&D collaboration networks of Key Enabling Technologies (KETs). In this article, a spatial interaction model is estimated, taking the perspective of knowledge interactions that are a vehicle for the creation and diffusion of new knowledge. Spatial interaction models refer to a class of models that allow investigating interactions between origin and destination locations based on origin-, and destination-specific characteristics as well as spatial separation information. As an empirical basis, the collaborative R&D projects of the EU Framework Programmes (FPs) for European NUTS 2 regions, as a proxy for inter-regional knowledge flows, are used. The identification of the distinct technological R&D networks follows the classification of Key Enabling Technologies (KETs).

Widening the angle of technological heterogeneity as to the first article, **Article II aims** to estimate *how regional embeddedness in R&D networks affects regional knowledge creation across different technological fields*. The second article employs an augmented regional

Knowledge Production Framework (KPF), estimating a negative binomial regression model. Again, empirically, KET-specific R&D collaboration networks of the EU FPs are used to derive the measure of regional embeddedness. To quantify regional knowledge output, the number of regional patent applications in different KET fields is used.

Article III then shifts attention to issues of *heterogeneity in modes of knowledge creation and knowledge output* that refers to specific characteristics of the knowledge creation processes shaped, e.g. by different knowledge bases and technological competences, research rationales and collaboration strategies, as well as different spatial scales of knowledge creation. This type of heterogeneity is made explicit by distinguishing between exploitation-oriented and exploration-oriented knowledge creation, as well as quantity-driven versus quality-driven knowledge output. Accordingly, **Article III aims** to *disentangle the effects of R&D networks on regional knowledge creation of different forms,* namely knowledge exploitation and knowledge exploration, as well as to *analyse the differences in effects between knowledge quantity and quality.* This article takes the perspective of spatial dependence, making neighbourhood relations among regions and the role of spatial spillovers explicit. Here, spatial dependence is incorporated into the model specification by applying spatial lag operators to the dependent and the independent variables, i.e. a spatial Durbin model (SDM). Including a spatial lag operator integrates neighbouring observations, allowing for the interpretation of local and global spatial externalities. Once more, empirically, the R&D networks are constructed based on inter-regional R&D projects of the EU FPs observed for the European NUTS 2 regions.

In **Article IV**, the *heterogeneity in knowledge creation capabilities of R&D actors*, reflected by, e.g. actor-specific knowledge endowments, collaboration rationales and research interests, motivates the analysis of the interplay between the micro-dynamics of knowledge creation processes and the macro-structure of inter-regional R&D networks. Hence, **Article IV** takes a simulation perspective **aiming** to *model the complex nature of multi-regional knowledge creation processes of European regions*. Specifically, an empirical multi-regional agent-based model (ABM) is presented that comprises a micro-level with interacting agents following specific (collaborative) research strategies and a macro-level of knowledge diffusion across empirically initialised regions. In general, ABMs are used to simulate heterogeneous agents' behaviour and interactions within a given environment to reflect an image of real-world systems. The developed model demonstrates a way to investigate drivers for regional knowledge creation of different kinds, such as inter-regional networks and agglomeration factors, while accounting for agent heterogeneity and the non-linearity of knowledge creation processes.

The dissertation is cumulative, comprising four research articles represented by the four main sections of the thesis (*Section 2 – Section 5*). *Section 6* summarises the main conclusions from the research articles and provides some implications in a regional, national and European policy context. Moreover, it discusses their limitations and raises ideas for future research on regional knowledge creation and R&D networks.

2 Geographical or relational: What drives technology-specific R&D collaboration networks? (Article I)

This section is based on the study *"Geographical or relational: What drives technology-specific R&D collaboration networks?"* (joint work with Thomas Scherngell, published in *Annals of Regional Science*, 2020)

Abstract: R&D collaboration networks enable rapid access to global sources of knowledge, especially in strongly knowledge-based and technology-driven industries. However, technological idiosyncrasies require a refined picture, particularly when explaining the interplay between geographical and relational effects driving the constitution and dynamics of R&D collaboration networks. We employ a spatial interaction modelling approach to estimate how spatial separation and network structural effects influence technology-specific R&D collaborations between European regions. Results underline both the significance of geographical barriers and network structural effects and confirm that network effects can compensate for geographical barriers – throughout all technologies investigated, although the effects differ in magnitude. However, when two regions are dissimilar in their network centrality, the potential to reduce negative geographical effects is relatively lower.

Keywords: R&D collaboration networks, spatial and network-structural effects, technological heterogeneities, Key Enabling Technologies, European Framework Programme, spatial interaction modelling

2.1 Introduction

Collaborative Research and Development (R&D) activities between firms, universities and research organisations are generally recognised to constitute an essential element for the successful generation of innovation. The notion of *R&D collaboration networks* has come into fairly wide use for describing such collaborative research endeavours and has become a fascinating research domain in various aspects (see Scherngell 2013 for an overview). With knowledge creation inevitably linked to innovation (Popadiuk and Choo 2006, among others), such R&D collaboration networks are considered to play an essential role from a regional perspective, moderating and structuring knowledge creation and diffusion processes within and across regions (Wanzenböck et al. 2014).

Recently, scholars started combining the relational and the geographical perspective, acknowledging the interrelation between space and networks in creating knowledge (Glückler et al. 2017). In this context, the general importance of networks is stressed but also the role of different network structures and topologies. On an organisational level, it is quite well acknowledged that the position of single nodes, e.g. representing firms, and the network structure as a whole have great impact on the creation of new knowledge and its diffusion (e.g. Ahuja 2000, Zaheer and Bell 2005, Giuliani 2007); also, seen from a spatial perspective understanding regions as network nodes (e.g. Whittington et al. 2009, Maggioni and Uberti 2011). Studies in the vein of geography of R&D collaboration networks, focusing on the identification and estimation of determinants affecting structures and dynamics of such R&D collaboration networks, are often accomplished at the regional level of analysis (see Scherngell and Barber 2009, Hoekman et al. 2010, Scherngell and Lata 2013, Lata et al. 2015, Morescalchi et al. 2015, Bergé 2017, Marek et al. 2017, among others). However, these works capture R&D collaboration networks, and accordingly, the underlying R&D activities in a relatively aggregated manner, neglecting technology-specific peculiarities of knowledge creation and interactions, such as knowledge properties and different modes of (collaborative) knowledge creation.

Reviewing the theoretical and empirical literature on R&D collaboration networks over the past two decades, we find an emphasis in the debate on how *geographical* characteristics affect their dynamics and on the role of *relational* drivers, also referred to as network structural effects. These two groups of determinants have been often discussed under the notion of local buzz (spatial proximity) versus global pipelines (region-external network relations) in R&D collaboration (see, e.g. Bathelt et al. 2004). There are several studies that separately address geographical or network structural factors when analysing cross-region R&D collaboration networks (see Scherngell 2019 for an overview). Nevertheless, there are only very few, but usually geographically and/or technologically quite limited studies, addressing both factors in an integrated modelling framework (see, e.g. Broekel and Boschma 2012, Broekel and Hartog 2013 or Bergé 2017).

This study intends to address this research gap by shifting attention to the role of geographical versus relational effects when explaining the constitution and dynamics of R&D collaboration networks in one integrated modelling framework and for a larger geographical area while accounting for technological idiosyncrasies. Accordingly, the objective is to estimate determinants of technology-specific R&D collaboration networks, shifting particular attention – as in previous works – to *geographical* effects, such as geographical distance or country borders, but also to network structural, i.e. *relational*, effects, such as central positioning, influencing the collaboration probability between two regions. To address this objective, we employ a spatial interaction modelling approach at the regional level. The R&D collaboration network under consideration is a network of organisations that collaborate in projects funded by the EU Framework Programme (FP). This network is partitioned into different technological domains and aggregated from the organisational to the regional level, using 505 European metropolitan and remaining non-metropolitan regions. The technological disaggregation is attained by assigning collaborative projects to specific relevant technologies.

In the latter context, we use the so-called Key Enabling Technologies (KETs), considered by the EU as specifically relevant in the global innovation competition. We use semantic techniques developed in an EU funded research project to assign data items to these technologies, and by this, go beyond standard classification systems that cannot capture these technologies. With our focus on networks of KETs, we propose – in contrast to previous research – a more nuanced grained and policy-relevant perspective when identifying determinants of R&D collaboration networks.

The study departs from previous related research in at least three significant aspects: *First*, and most importantly, we include – additionally to geographical effects – network structural effects as a major additional set of determinants, while previous research mainly focused on spatial and technological barriers for R&D collaboration networks. Such network structural effects, e.g. the central positioning of regions in the network, are assumed to play a crucial role for overall dynamics (see Wanzenböck et al. 2014) and the probability of establishing additional network links between region pairs (Barthélemy 2011). *Second*, we introduce technological heterogeneities in our investigation of determinants affecting structures and dynamics of R&D collaboration networks, going beyond existing works that remain at an aggregated level of technological fields (see Morescalchi et al. 2015 for an overview). *Third*, we introduce an innovative set of regions and distinguish – in contrast to previous research – between metropolitan and non-metropolitan regions in our regional system. This enables to disentangle urbanisation effects from other effects (e.g. geographical proximity or country borders) in a more robust way.

The remainder of the study is organised as follows. The following subsection recaps the main elements of the theoretical and empirical debate on the determinants of R&D networks, explicitly highlighting the relevance of the focus on geographical and relational characteristics in different technologies. Subsection 2.3 shifts specific attention to the role of technological heterogeneities in such networks that have been largely neglected so far in the empirical literature. Subsection 2.4 describes the spatial interaction approach used to identify determinants of collaboration, followed by Subsection 2.5 that sets out the empirical setting. Subsection 2.6 discusses the estimation results before Subsection 2.7 closes with a summary and some future research ideas.

2.2 The theoretical and empirical debate on determinants of R&D networks

The investigation of R&D collaboration networks has attracted much attention in the recent past. In regional science, this stems from the comprehensive agreement that spatial and network dimensions are crucial for moderating and structuring knowledge creation and diffusion processes within and across regions (see, e.g. Autant-Bernard et al. 2007a, Bathelt and Glückler 2003). Recently, this research interest is mainly motivated by the seminal work by Bathelt and Glückler (2003) that suggest a '*relational turn*' in economic geography, highlighting the interrelation between networks, geography, and knowledge (Bathelt and Glückler 2003, Glückler et al. 2017). From the angle of '*proximity'*, various contributions acknowledge the reinforcing role of non-spatial proximity dimensions, such as organisational, institutional, social and cognitive factors, for networks of knowledge creation and innovation (e.g. Kirat and Lung 1999, Boschma 2005, Torre and Rallet 2005, Mattes 2012).

This theoretical debate has paved the way for the increasing empirical interest in the analysis of R&D collaboration networks, also driven by new large-scale datasets on collaborative R&D, and the advancement of methodological instruments, e.g. in spatial interaction modelling (see Scherngell 2019 for an overview). Meanwhile, a large and diverse body of empirical literature on determinants of R&D collaboration networks in different technological fields and different geographical areas exists. Despite their differences, spatial proximity turns out to be an essential factor for the constitution of R&D collaboration in all these studies, also in times of increasing globalisation and new communication technologies (see e.g. Scherngell and Barber 2009, Lata et al. 2015, Marek et al. 2017). This is usually explained by the specific characteristics of the knowledge elaborated on in such collaborations, considering that more complex knowledge requires exchanging more tacit knowledge elements. Accordingly, face-to-face interaction in inter-organisational learning processes makes spatial proximity (still) a crucial factor in establishing and maintaining R&D network links (Rallet and Torre 1998, Storper and Venables 2004)[1]. Given the high costs of transmitting tacit knowledge in geographical space, complex knowledge is more immobile. Accordingly, network effects may become more critical for such fields to overcome geographical barriers. In contrast, with more explicit (codified) knowledge elements being involved in the knowledge creation process, e.g. in very science-based and open technological fields (e.g. nanotechnology or biotechnology), the spatial scale of the collaboration may increase, pointing to a less critical role of geographical space as a driver for network dynamics.

However, apart from being geographically close to create and exchange complex and tacit knowledge, being part of the same professional community – such as a research network – may facilitate knowledge creation and transfer; i.e. '*organisational proximity*' (Kirat and Lung 1999, Boschma 2005) or '*organised proximity*' (Torre and Rallet 2005). This type of relational proximity is characterised by shared knowledge and knowledge bases (e.g. same scientific community) and interacting actors that enable interaction and accelerate

[1] In recent literature, differing spatial and collaboration network structures across different technologies are often related to the complexity of knowledge creation in different technological fields (see Fleming and Sorensen 2001, Balland and Rigby 2017).

knowledge creation (Boschma 2005, Torre and Rallet 2005). The EU Framework Programmes (FP), for instance, feature such kind of 'organised proximity', where firms, universities and research organisations located in various European regions collaborate in all sorts of topics aiming for excellence throughout the European Research Area (ERA; see Breschi and Cusmano 2004, followed by many others). Moreover, fostering inter-regional collaboration through funding opportunities, such as the EU FPs, has facilitated long-distance collaborations (Scherngell and Lata 2013), highlighting the potential of networks as an organisational arrangement to overcome geographical barriers.

In this vein, a region's position in global R&D networks has been increasingly considered important in recent years, particularly for regions with less local knowledge endowments and R&D capabilities (Wanzenböck and Piribauer 2018). This has shifted attention to the conditioning role of networks, i.e. relational effects, moderating and structuring collaboration, compared to geographical ones (Glückler et al. 2017).

Inspired by network science, we can derive relevant arguments in this context. A first key aspect concerns the accessibility to new knowledge, referring to the position of regions in networks and their network embeddedness in terms of the number of collaboration links[2]. However, the quantity of a region's collaboration arrangements matters and their quality indicates, on the one hand, access to reliable information itself and, on the other hand, linkages to other partnering organisations holding reliable information themselves (e.g. Uzzi and Lancaster 2003). A second key aspect stresses that regions may more likely increase collaborations to other regions showing similar network attributes, e.g., their number and quality of collaboration links[3].

The importance of relational or network structural effects on R&D collaborations have been partly addressed in only a few empirical studies up to now, such as for social proximity (Autant-Bernard et al. 2007a), institutional proximity (Ponds et al. 2007), network proximity (Bergé 2017), as well as relational dependence (Maggioni et al. 2007). Moreover, there are only very few and geographically and/or technologically quite limited studies addressing

[2] Studies on the effect of an actor's embeddedness in a knowledge network on its innovative performance are manifold and exist for different industries such as biotechnology and chemicals (Salman and Saives 2005, Gilsing et al. 2008). Driven by the debate on "local buzz" and "global pipelines" as two forms of interactive knowledge creation (Bathelt et al. 2004), the spatial dimension of the actor's embeddedness in networks of knowledge creation gained attraction in regional science.

[3] In social network analysis, this is usually referred to as homophily, i.e. social actors are more likely to inter-link when they have similar attributes (McPherson 2001).

both geographical and network structural factors in one framework, e.g. the study of Broekel and Boschma (2012) for the Dutch aviation industry, Broekel and Hartog (2013) for Germany, or Bergé (2017) for the field of Chemistry in Europe. For the case of publicly funded R&D collaboration, such as the EU FP, this would be of specific interest given the policy interest in fostering collaboration across geographical distances by manifesting sustainable network links. Against the background of these theoretical and empirical debates, we pose our first set of hypotheses:

> **Hypothesis 1a:** Network structural effects drive collaboration patterns in publicly funded cross-region R&D collaboration networks

> **Hypothesis 1b:** Network connectivity compensates for geographical barriers in the constitution of publicly funded cross-region R&D collaboration networks

Hence, we assume that network channels are able to reduce negative effects on cross-region R&D collaboration probabilities stemming from geographical barriers (e.g. distance). In a similar vein, studies by e.g. Bell and Zaheer (2007), Glückler (2006), and Hansen and Løvås (2004), find evidence to support this hypothesis for the case of knowledge transfer, flow, and spillovers.

2.3 Technological heterogeneities in R&D collaboration networks

While technological heterogeneities in terms of differences in knowledge bases and knowledge creation processes have been subject to a long-lasting debate among evolutionary scholars (Nelson and Winter 1982, Pavitt 1984, Breschi et al. 2000, Malerba 2002), they have been rarely addressed in the context of R&D collaboration networks, in particular in empirical terms. Conceptually, the role of different knowledge domains – originally referred to as technological regimes[4] – has been stressed for explaining differences across sectors in patterns of innovation. Accordingly, it can be considered highly relevant for R&D collaboration networks as a major input for innovation. Malerba (2002) identifies three key dimensions of knowledge related to the notion of technological regimes: the degree of *accessibility* (i.e. opportunities of gaining knowledge, e.g. through cross-regional network

[4] The term '*technological regime*' originates in the work by Nelson and Winter (1982) and characterizes the knowledge environment in which organisations within the same industry are argued to be subject to same technological and knowledge conditions, such as e.g. the degree of accessibility, the sources of technological opportunities, the cumulativeness of knowledge (Freeman 1982, Malerba and Orsenigo 2000), and the nature of knowledge (e.g. specificity, tacitness, complexity; Winter 1997).

links), sources of technological *opportunity*, and *cumulativeness of knowledge* (i.e. the degree by which the generation of new knowledge builds upon current knowledge). Each dimension is assumed to differ among sectors and technologies due to the knowledge base's specific properties, which is determined by differences in technological knowledge itself, involving varying degrees of *specificity*, *tacitness*, *complementarity*, and *independence* (Winter 1987).

We assume that such heterogeneities in terms of regional knowledge bases, knowledge types and attributes relate to differing structural properties of R&D collaboration networks, as well as varying underlying mechanisms that drive the constitution thereof. This motivates our second hypothesis:

> **Hypothesis 2a:** Technological R&D collaboration networks differ with respect to their estimated network and geographical effects.

Existing empirical studies investigating differences across technologies have a relatively limited geographical and sectoral coverage, not allowing for a systematic and comprehensive interpretation of determinants of R&D collaboration (e.g. Broekel and Graf 2012 for the case of ten German technologies). Moreover, they mostly also disregard technological heterogeneities that may influence the relevance and spatial scale of R&D collaboration (see Ponds et al. 2007, Martin and Moodysson 2013, Tödtling et al. 2006, Trippl et al. 2009).

However, the pure observation of heterogeneities does not explain why they exist. Considerations on the manifold nature of knowledge and different knowledge bases may provide useful anchor points in this context. For instance, Asheim and Coenen (2005) emphasise the existence of two types of knowledge bases: *analytical* and *synthetic*, each linked to a different technological environment; whereas in technologies with analytical knowledge bases, scientific knowledge is predominant, a synthetic knowledge base alludes to industrial settings where innovation often occurs through the application and/or new combination of existing knowledge, such as engineering-oriented fields (Asheim and Coenen 2005). Moreover, Pavitt (1984) categorises sectors according to their sources of technology used, the institutional sources and nature of the technology produced, as well as the characteristics of innovating firms (e.g. size, principal activity). Thereof, Pavitt derives four types of sectors: *supply-dominated* (e.g. clothing, furniture), *scale-intensive* (e.g. food, cement), *specialised supplier* (e.g. engineering, software, instruments), and *science-based* *producers* (e.g. chemical industry, biotechnology, electronics). Derived from this discussion, we pose an additional hypothesis related to hypothesis 2a:

> **Hypothesis 2b:** Geographical effects are assumed to have stronger negative impacts on engineering-oriented fields, while science-oriented fields are more driven by negative network structural effects.

From an empirical perspective, the question arises which technological breakdowns are to be chosen for observing technological heterogeneities. Here we can see that especially novel and fast-growing technologies that spur innovation and technological progress of countries, regions and industries have gained anew interest, both in academia (see, e.g. Evangelista et al. 2018, Montresor and Quatraro 2017) and in the policy realm. At the European policy level, this is reflected by the new emphasis on so-called Key Enabling Technologies (KETs), bringing technologies into focus that are considered crucial for the EU's development towards a sustainable, knowledge-based economy (EC 2009, EC 2012)[5]; these are *Nanotechnology*, *Microelectronics*, *Photonics*, *Advanced Materials (AM)*, *Advanced Manufacturing Technology (AMT)* and *Industrial Biotechnology* (EC 2009)[6].

Despite the common specificities of KETs (by which they identify as *'key enabling'*), we argue that these distinct technologies differ concerning their geographical and network impacts on inter-regional R&D collaboration. Empirically, KETs are strongly spatially concentrated on specific regions (Montresor and Quatraro 2017, Evangelista et al. 2018). Regarding cross-region R&D collaborations, Wanzenböck et al. (2020) observe noticeable differences between KETs in the spatial distribution of regional network effects. While network effects are more spatially concentrated in the engineering-based fields (such as *Photonics* or *AMT*), inter-regional network linkages tend to be more equally distributed across regions in the science-based sectors (Wanzenböck et al. 2020).

[5] In a line of efforts towards the initiation and implementation of a coherent European Strategy for KETs, the European Commission set up two High Level Expert Groups (in 2010 and 2013) to advice on the elaboration of a KETs strategy and to ensure its successful implementation (EC 2012, EC 2015a).

[6] KETs are understood as generic technologies that are characterised by relatively rapid pervasiveness and growth, high knowledge and R&D intensity, and highly skilled employment etc. (EC 2009). Due to their specific characteristics, R&D collaboration networks are considered of particular importance in a KET context in order to cope with the high demand for R&D in these technological fields, and to gain rapid access to nation-wide and global state of the art knowledge. Moreover, KETs are claimed to affect the regional capacity of developing new technological specialisations (Montresor and Quatraro 2017). Specifically, in such globally relevant technologies like KETs, R&D networks may serve as channels for transmitting knowledge over larger geographical distances (see e.g. Autant-Bernard et al. 2007), and hence be of particular importance for innovation and regional growth processes (Huggins and Thompson 2014).

With respect to the generally uneven spatial distribution of knowledge creation, especially in technology-specific knowledge environments, these findings strongly point at KET-specific differences in terms of accessibility of new and external knowledge determined by different degrees of spatial and network proximity across KETs. Moreover, regional disparities regarding the specialisation in certain KETs, suggests disparate technological opportunities as well as varying degrees of cumulativeness of knowledge, resulting in differing regional innovation paths and potentials for cross-sectoral and cross-regional spillovers. Considering KETs in light of Pavitt's taxonomy (1984), they can be characterized as either *specialised suppliers* – generally engineering-oriented – carrying out frequent innovations often in collaboration with customers, or *science-based producers* that develop new products and processes often in collaboration with universities. Hence, KETs potentially differ with respect to their sectoral and institutional sources of technology used; in particular, in terms of the degree to which new knowledge is created within the sector or comes from outside, as well as to which extent intramural and extramural knowledge sources are used (Pavitt 1984).

Against this background, this study shifts attention to R&D collaboration networks in different technologies – proxied by KET fields – and focuses on the debate of the differing role of geographical and relational characteristics in such distinct technological domains that follow particular rationales and aims in knowledge creation. This is addressed with a novel dataset and for the first time in an integrated modelling framework for a larger geographical area, namely the whole European territory.

2.4 Methodological approach and model

For the estimation of spatial and network structural determinants of technology-specific R&D collaboration networks, we follow earlier research and employ a spatial interaction modelling approach. In general, spatial interaction models can be used to describe interactions (e.g. flows, collaborations) between actors distributed over some geographic space. In contrast, the interactions are a function of the attributes of the locations of origin, the destination's characteristics, and the friction (*separation*) between the respective origin and destination. The purpose of such models is to explain the interaction frequencies between two spatial entities.

Specifically, spatial interaction models refer to a class of models applied to identify determinants – particularly separation effects – of interactions between discrete spatial entities (Roy and Thill 2003), such as in our case, interactions in R&D collaboration networks between regions. The general form of the model can be written as

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij} \quad \text{with} \quad i, j = 1, \dots, N \tag{2.1}$$

where $\mu_{ij} = E(Y_{ij})$ is the expected mean interaction frequency between locations i and j, and ε_{ij} is an error about the mean (Fischer and Wang 2011). In this study, locations correspond to European regions, where each location is both the origin and destination of interactions.

In general, these models comprise three types of factors to explain mean interaction frequencies between spatial locations i and j. Accordingly, the model class distinguishes: (i) *origin-specific* factors characterising the ability of the origins to generate R&D network links, (ii) *destination-specific* factors indicating the attractiveness of destinations, and (iii) *separation factors* that represent the way different forms of *separation* between origins and destinations constrain or impede the interaction, most basically geographical distance (LeSage and Fischer 2016). Hence, mean interaction frequencies between origin i and destination j are modelled by

$$\mu_{ij} = O_i \, D_j \, S_{ij} \quad \text{with} \quad i, j = 1, \dots, N \tag{2.2}$$

where O_i and D_j are the origin-specific and destination-specific factors, respectively, and S_{ij} denotes a multivariate function of separation between locations i and j.

While there are different functional forms to specify origin, destination and separation functions (see Fischer and Wang 2011), studies investigating R&D networks usually employ univariate (i.e. with only one variable) power functional forms for origin and destination functions, and multivariate (i.e. with several separation variables) exponential functional forms for the separation function. We follow these lines and define

$$O_i = O(o_i, \alpha_1) = o_i^{\alpha_1} \tag{2.3}$$

$$D_j = D\big(d_j, \alpha_2\big) = d_j^{\alpha_2} \tag{2.4}$$

$$S_{ij} = \exp\left[\sum_{k=1}^{K} \beta_k s_{ij}^{(k)}\right] \tag{2.5}$$

Here, o_i and d_j are measured in terms of variables controlling for the mass in the origin and the destination, respectively. In the context of R&D networks, these are often captured by the number of firms or researching organisations in a region. Accordingly, α_1 and α_2 are scalar parameters to be estimated so that the product of the functions $O_i D_j$ can be simply interpreted as the number of cross-region R&D collaborations which are possible. The core of the spatial interaction model is the separation function as defined by Equation (2.5), with K (k = 1, …, K) separation measures to be estimated that will show the relative strengths of the separation measures, and β_k denoting the respective k^{th} estimate for separation measure k.

The model applied takes the specific form of a spatially filtered, negative binomial spatial interaction model (see Scherngell and Lata 2013 in a similar context)[7]. The primary motivation for this is given by the true integer nature and distributional assumptions on the dependent variable, cross-region R&D collaborations. Further, the proposed model specification accounts for the spatial dependence of the data used (participation in European Framework Programme (FP) projects) in the empirical application, as well as for a high degree of variation (overdispersion) and a large number of zero counts. Hence, it is assumed that the dependent variable Y_{ij} follows a negative binomial distribution with expected values, as stated in (2.2).

Compared to the Poisson model that assumes equidispersion (i.e. conditional mean equals the conditional variance), the negative binomial model explicitly corrects for overdispersion[8] by adding a dispersion parameter θ. Hence, the negative binomial spatial interaction model takes the form (Long and Freese 2006)

$$\Pr(Y_{ij} = y_{ij} | \mu_{ij}, \gamma) = \frac{\Gamma(y_{ij} + \theta)}{\Gamma(y_{ij} + 1)\Gamma(\theta)} \left(\frac{\theta}{\theta + \mu_{ij}}\right)^{\theta} \left(\frac{\mu_{ij}}{\theta + \mu_{ij}}\right)^{y_{ij}} \tag{2.6}$$

[7] Although the data used has excess zeroes, we did not opt for a zero-inflated version of the negative binomial model, since we argue that each region possibly has the chance to engage in a collaboration (no structural zeroes).

[8] Not accounting for overdispersion would result in incorrect standard errors, leading to possibly wrong significances of parameters (Cameron and Trivedi 1998).

where $\mu_{ij} = E\left[y_{ij} \middle| O_i, D_j, S_{ij}\right] = \exp\left[O_i(\alpha_1)\, D_j(\alpha_2)\, S_{ij}(\beta)\right]$ and Γ denotes the gamma function with a model parameter θ accounting for overdispersion in predictors (see Cameron and Trivedi 1998 for a more detailed derivation).

To take the spatial dependence of flows into account, spatial filtering using eigenvectors (ESF) is employed[9] (see appendix to this section for details on ESF). In this study, six separate – one for each KET – regression models are estimated via the spatially filtered negative binomial spatial interaction model. We include the first ten eigenvectors from the set of κ eigenvectors with MI/MI_{max} larger than 0.25 (see, e.g. Scherngell and Lata 2013), where MI denotes the Moran's I value and MI_{max} its maximum value, as additional explanatory variables in the model (see, e.g. Fischer and Wang (2011) for details).

Recalling the negative binomial specification of the model in Equation (2.6), the final empirical model to be estimated is specified by setting

$$\mu_{ij} = \exp(\alpha_0 + \alpha_1 \ln(o_i) + \alpha_2 \ln(d_j) + \sum_{k=1}^{K} \beta_k s_{ij}^{(k)} + \sum_{q=1}^{Q} \phi_q E_q + \sum_{r=1}^{R} \varphi_r E_r + \xi_{ij}) \quad (2.7)$$

where E_q denotes the selected subset of eigenvectors expanded by means of the Kronecker product associated with the origin variable, and E_r the respective eigenvectors for the destination variable; ϕ_q and φ_r are the corresponding coefficients. Note that the explanatory variables enter the regression in their logged form (except the dummy variables). Since the assumption of the dependent variable – the R&D interactions between region i and j – being independent and normally distributed does not hold, the parameters of the model are estimated by means of Maximum Likelihood (ML) estimation (see Cameron and Trivedi 1998 for estimation details).

2.5 Data and variables

This study's main interest is to estimate determinants of technology-specific R&D collaboration networks, focusing on spatial separation and network structural effects. The geographical coverage comprises the current 27 EU member states (excluding Malta and Cyprus) plus the United Kingdom, Switzerland and Norway, corresponding to 505 regions.

[9] In the context of spatial interactions, spatial autocorrelation of flows is understood as correlation between R&D collaboration flows from the same origin or destination, to neighbouring origins or destinations, respectively. Not accounting for spatial autocorrelation leads, similar to overdispersion, to incorrect inferences and hence, wrong significances (Chun 2008).

Going beyond previous research, we distinguish 270 metropolitan regions as well as 235 remaining non-metropolitan regions. Whereas, metropolitan regions are NUTS 3 regions or a combination thereof integrating neighbouring urban areas to one spatial entity[10], the remaining non-metropolitan regions are either original NUTS 2 regions, or adapted NUTS 2 regions with respective NUTS 3 regions – belonging to a metropolitan region – removed (see Figure A2.1 in the appendix of this section for a map of metropolitan regions)[11].

Dependent variable

As dependent variable EU funded KET R&D collaboration links are used (see Table A2.1 in the appendix of this section some descriptive statistics). Data is extracted from the EUPRO database[12] comprising systematic information on collaborative research projects of FP1-FP7 as well as Horizon 2020 (until 2016), including details on respective participating organisations, e.g. name, type, and their geographical location in the form of organisation addresses (see Heller-Schuh et al. 2015 for details). Clearly, projects carried out under the EU FPs constitute a specific type of R&D collaboration network subject to certain governance rules (e.g. each project must have partners from at least two different countries). However, these rules are less relevant for forming collaboration than their behaviour driven by strategic, technological, geographical, cultural, and institutional conditions.

To construct the dependent variable, we consider the 7th FP and H2020 with a time horizon of 2007-2016. For each KET, a technology-specific symmetric regional collaboration matrix is constructed, where the elements indicate the number of joint EU funded research projects[13]. This matrix is then transformed into a vector with rows representing all possible combinations of links between the regions; this results in a vector of length n^2-by-1

[10] Metropolitan regions represent all agglomerations of at least 250,000 inhabitants; whereas each agglomeration is represented by at least one NUTS 3 region. If in an adjacent NUTS 3 region more than 50% of the population also lives within this agglomeration, it is included in the metropolitan region. This is based on the 2013 NUTS version and the 2010 Geostat population grid defined by Eurostat.

[11] Although the NUTS 2 level perspective is widely used in previous related empirical literature (e.g. Scherngell and Barber 2009, Hoekman et al. 2012), we opt for metropolitan regions as units of analysis. Metropolitan regions are a quite recently introduced classification on a European level based on agglomeration (EC 2008, Dijkstra 2009), which by definition is an urban core including the surrounding catchment area. Hence, this classification corrects for distortions created by e.g. the NUTS classification that separates these two geographical spaces in most cases.

[12] The EUPRO database is maintained by AIT Austrian Institute of Technology and is accessible via RISIS (risis2.eu). It has been advanced within RISIS, in particular in terms of geolocalisation, standardisation and integration with other datasets.

[13] The number of collaborations between regions results from the aggregate of collaborations (full count) between the participating organisations located within these regions.

containing the inter- and intra-regional collaboration activities of all region pairs. Figure A2.2 in the appendix (Subsection 2.8) illustrates the networks' spatial distribution showing the characteristic star-shaped backbone structure, revealing the Paris region as dominating hub in all networks. Nevertheless, the R&D networks differ with respect to density, variance in the number of collaborations, spatial scales and importance of certain regions (e.g. London in *Nanotechnology* and *Biotechnology*; see Table A2.2 in the appendix of this section).

Independent variables

As described in the previous section, the independent variables comprise three types: origin, destination and separation variables. The origin variable o_i and the destination variable d_j are specified as the number of organisations participating in joint EU funded FP projects in region i and j in a distinct KET field. Empirically, these variables represent the potential of regions to engage in collaborative R&D activities. Statistically, they control for the different sizes of the regions (see Figure A2.3 in the appendix of this section for their spatial distribution). For the separation variables, we distinguish between (i) spatial separation variables, and (ii) network structural separation variables (see the appendix of this section for Table A2.3 with descriptive statistics and Table A2.4 providing correlation measures between explanatory variables).

Clearly, this study's focus lies on the separation variables capturing the friction between two regions assumed to influence their collaboration intensity. Concerning our research question, we shift attention to geographical versus relational, i.e. networks structural separation variables:

- As variables accounting for spatial separation effects, *first*, the geographical distance $s_{ij}^{(1)}$, measured as the great circle distance, indicating the shortest distance between two regions i and j, *second*, $s_{ij}^{(2)}$ a dummy variable indicating the presence of a common national border of regions (set to one, if two regions are located in different countries, zero otherwise), and *third*, $s_{ij}^{(3)}$ a dummy variable indicating links between two metropolitan regions (set to one, if the link is between two metropolitan regions, zero otherwise) are included in the model.

- As network structural separation effects, *first*, the gap in degree centralities $s_{ij}^{(4)}$ and *second*, $s_{ij}^{(5)}$ the gap in the hub score between the two regions i and j, are included[14]. Whereas the degree centrality simply measures the number of collaboration links of a region, the hub score (Kleinberg's authority centrality[15]) is defined as the principal eigenvector of $A * t(A)$, where A is the adjacency matrix of the KET-specific R&D network. It hence, indicates whether a region maintains KET-specific collaboration links and is at the same time linked to other regions, that themselves are well-connected to access KET-specific knowledge. Together, the two variables account for differences in the *quantity* of collaboration links and difference in the *quality* of these interactions.

We refrain from including a measure for technological separation, such as a technological distance included in previous works to isolate geographical from technological effects since the units of analysis are distinct technological fields, with relatively homogenous subclasses.

Assignment of data items to KETs

The meaningful delimitation of KETs is essential to address the research objectives of this study. However, KETs are usually cross-cutting technological domains and are not pre-defined categories in the data. Thus, we employ the classification approach developed in the EU funded project KNOWMAK that provides a publicly available ontology for KETs, comprising a hierarchical system of topical classes for each KET characterised by a set of weighted keywords. First, using natural language processing techniques, the data items, i.e. FP projects, are assigned to these topical classes. The underlying fundament of the assignment is an advanced ontology of the KET knowledge domains that describes the substantive contents of each KET by sets of topics and subtopics that are characterised by hundreds of keywords (Maynard et al. 2017). The population of the ontology with meaningful keywords is of crucial importance for a proper assignment of projects to the specific KETs. Maynard et al. (2017) employ a solution with multiple layers of keyword extraction from policy and other relevant documents on KETs and a mixture of automated techniques interspersed with expert knowledge at key junctures[16]. Second, projects are

[14] We refrain taking other centrality concepts here that are e.g. not defined for weighted graphs (betweenness) and/or fragmented ones (closeness).

[15] Equals the authority score for undirected graphs (see Kleinberg 1999).

[16] Different Natural Language Processing (NLP) techniques are used to refine sets of keywords and explore inter-relations between them (e.g. two generic keywords are marked as stop-words, combinations of

tagged and mapped to specific KET subtopics aggregated to the six main KETs to extract the KET-specific collaboration networks for the analysis at hand. The mapping of projects to a KET is based on a similarity score between the project description and the specific keyword sets of the particular KET subtopics. The similarity score depends basically on the overlap in keywords from the ontology and the text of the project description, whereas their representativeness weights the keywords for a specific topic using pointwise mutual information (PMI) procedures (see Blei 2012). Note that the assignment of projects is subject to a series of robustness and sensitivity analyses (including manual checking of individual cases) to guarantee a sufficiently meaningful and robust result (see also Maynard et al. 2017)[17]. This development has led to a public standard where different knowledge creation activities are mapped to KETs and used to produce indicators on regional knowledge creation in Europe, including the number of regional FP participations (accessible and reproducible under knowmak.eu).

2.6 Estimation results

Table 2.1 displays the estimation results of the spatial interaction models. While the first column reports the ML estimates for a basic spatial interaction model (model 1), including the origin and destination variables as well as the spatial separation measures: geographical distance, country border effect, and the metropolitan region; the second column comprises the results for the full model (model 2) expanding the purely spatial model by including two network structural separation measures. Estimating the two models separately allows us to test our hypotheses (see Subsection 2.2 and Subsection 2.3) since we can observe the changes in the spatial effects directly, when accounting for network structural effects. Each of the two model specifications was executed for all six KETs to compare the effect sizes of the determinants of technology-specific R&D collaboration networks. For all models, the significance of the θ-parameter suggests the preference of a negative binomial model over the Poisson specification without heterogeneity. Moreover, for all models, a likelihood ratio test shows the preference of the spatially filtered negative binomial model against the non-filtered version. Note that we aggregate over the whole period (i.e. summing up FP7 and H2020) due to the too high number of zeros challenging a reasonable estimation.

keywords and multi-term keywords are constructed that are specifically relevant for a topic to get a better discrimination (Maynard et al. 2017).

[17] Details on the semantic approach and also the technical tools are given at knowmak.eu.

Our discussion shifts attention to the separation variables given our focus on geographical versus network structural effects. The origin and destination variables that just control for the mass in the origin and the destination region are significant and higher than one (see Table 2.1), i.e. the number of organisations active in a KET in a region naturally increases the likelihood for R&D collaboration in this KET with other regions.

Turning to the results of the separation effects for model (1), it can be seen that the geographical distance between two regions has a negative impact on the expected collaboration frequency between these two regions for all KETs – as indicated by the negative and significant estimates; this result coincides with findings in previous studies (Scherngell and Barber 2009, Scherngell and Lata 2013). Whereas, the effects are the highest (the most negative) for *Photonics* (for a coefficient of -0.25 this equals to a change of -22% given by its exponential[18], closely followed by *Nanotechnology* (with a factor change of 0.78; i.e. a change of -22%). The effects for *Microelectronics, Advanced Materials* and *AMT* are the smallest – all three within a small range of change of -13% to -14%. The coefficients for the country border effects are also significantly negative for all KETs, suggesting that a national border between any two regions decreases the expected collaboration frequency for participating organisations located in these regions.

This finding is a somewhat pessimistic outcome in a European integration and policy context. While country border effects seem to diminish in the FP networks as a whole (Scherngell and Lata 2013), in KETs – that are considered the most important technological domains for economic competitiveness – they are still a significant barrier for collaboration. Here the negative effects are the lowest for *Nanotechnology* and *Photonics*, while *Microelectronics* shows the highest negative effect. For region pairs located in different countries, the expected number of collaborations is hypothetically decreased by -22% in the case of *Microelectronics.*

The estimates for the metropolitan region dummy are positive and significant for all KETs except *Advanced Materials*. This implies that two metropolitan regions ‘increase’ the expected number of collaborations of their organisations by +0.7% in the field of *Microelectronics*, which exhibits the smallest effect and +23% in *Nanotechnology* with the

[18] A change of one kilometre in geographical distance results in an expected count decrease by a factor of $\exp(-0.250) = 0.779$ which implies a change of -22% (see Long and Freese 2006).

most considerable effect (compared to links between non-metropolitan regions and links between metropolitan and non-metropolitan regions).

Foremost, we can distinguish two groups of KETs with respect to their geographical effects: (i) *Nanotechnology* and *Photonics*, and (ii) *Microelectronics*, *Advanced Materials* and *AMT*, that each share common characteristics but are complementary to each other in terms of the importance of geographical effects. Whereas, the geographical distance is the most restrictive force for *Nanotechnology* and *Photonics* for inter-regional collaboration, the country border shows the weakest effect (across all KETs), in the case of *Microelectronics* and *AMT*, this relation is reversed, showing a strong impact of the country border and the weakest effects of geographical distance. Hence, R&D collaborations in *Nanotechnology* and *Photonics* are much more localised but still inter-regional. This may be related to the resource and infrastructure intensive character of these technological fields, with many countries having only one scientific centre, which is, therefore 'forced' to collaborate across countries (or even at a global scale). In contrast, *Microelectronics* and *AMT*, on the one hand, are relatively global in their collaboration behaviour, but on the other hand, are to a more considerable extent negatively affected by country borders. Moreover, they are to a lesser extent confined to collaboration between metropolitan regions, as evidenced by the relatively lower estimate for the metropolitan region dummy.

Model (2) adds network structural separation variables, enabling us to infer on our main research question, namely whether network structural effects are at stake at all, and whether they are more important than geographical ones, able to compensate for geographical barriers under certain network structural conditions (hypothesis 1). We find a significantly negative impact of the gap in degree centralities between two regions on their expected collaboration frequency – in all KETs. That is, the number of collaborations is expected to be higher between similar regions in terms of the *quantity* of existing collaboration links. This is regardless of the actual number of collaboration links unless they are similar, i.e. two regions with many links but also two regions with each only a few links. In terms of KET-specific differences, for the gap in degree centralities, i.e. the quantity of the links, we find some notable differences: the highest negative effect is found for *AMT* with a change of -24% and *Microelectronics*, whereas *Advanced Materials* exhibits the smallest effect (change of -0.6%).

Table 2.1. *Estimation results of the spatially filtered negative binomial spatial interaction models*

	Model (1)						Model (2)					
	Nano	Micro	Photonics	AM	AMT	Ind. Biotech.	Nano	Micro	Photonics	AM	AMT	Ind. Biotech.
Origin and destination variable [$\alpha_1 = \alpha_2$]	1.316*** (0.005)	1.436*** (0.008)	1.303*** (0.006)	1.611*** (0.012)	1.455*** (0.007)	1.303*** (0.005)	1.352*** (0.006)	1.565*** (0.009)	1.318*** (0.006)	1.712*** (0.014)	1.525*** (0.008)	1.339*** (0.005)
Geographical distance [β_1]	-0.245*** (0.008)	-0.144*** (0.012)	-0.250*** (0.008)	-0.145*** (0.014)	-0.148*** (0.010)	-0.189*** (0.007)	-0.213*** (0.008)	-0.097*** (0.012)	-0.222*** (0.008)	-0.123*** (0.015)	-0.083*** (0.010)	-0.157*** (0.007)
Country border effects [β_2]	-0.153*** (0.020)	-0.250*** (0.033)	-0.162*** (0.021)	-0.199*** (0.039)	-0.212*** (0.026)	-0.213*** (0.018)	-0.185*** (0.020)	-0.298*** (0.033)	-0.192*** (0.021)	-0.233*** (0.039)	-0.281*** (0.026)	-0.235*** (0.018)
Metropolitan region [β_3]	0.210*** (0.010)	0.071*** (0.017)	0.144*** (0.011)	-0.021 (0.020)	0.110*** (0.013)	0.153*** (0.009)	0.186*** (0.010)	0.034* (0.017)	0.135*** (0.011)	-0.033 (0.020)	0.120*** (0.013)	0.131*** (0.009)
Gap in degree centralities [β_4]	-	-	-	-	-	-	-0.137*** (0.005)	-0.178*** (0.009)	-0.117*** (0.006)	-0.062*** (0.012)	-0.272*** (0.007)	-0.148*** (0.005)
Gap in hub score [β_5]	-	-	-	-	-	-	-0.170** (0.054)	-1.291*** (0.084)	0.238*** (0.059)	-0.962*** (0.095)	-0.156* (0.066)	-0.280*** (0.048)
Constant [α_0]	-5.906*** (0.056)	-5.999*** (0.090)	-5.638*** (0.061)	-6.473*** (0.106)	-6.055*** (0.075)	-6.154*** (0.052)	-5.701*** (0.058)	-5.997*** (0.089)	-5.421*** (0.062)	-6.625*** (0.107)	-5.663*** (0.074)	-5.883*** (0.054)
Dispersion [θ]	1.117*** (0.015)	0.733*** (0.015)	0.921*** (0.012)	0.747*** (0.020)	0.806*** (0.013)	1.238*** (0.015)	1.144*** (0.015)	0.760*** (0.015)	0.933*** (0.012)	0.755*** (0.021)	0.852*** (0.014)	1.293*** (0.016)
Likelihood ratio test	1449.4***	592.2***	666.2***	372.7***	733.8***	1666.5***	1469.0***	610.5***	655.7***	320.4 ***	702.0***	1633.8***

Notes: The dependent variable is the number of EU funded R&D collaborations between two regions; for each model ten origin and destination spatial filters as specified in the text are included as explanatory variables; the number of observations is 255,025; standard errors are given in parentheses; *** indicates significance at the 0.001 level, ** indicates significance at the 0.01 level, *indicates significance at the 0.05 level; due to the symmetry of the origin and destination variable, α_1 equals α_2 up to numerical precision; the Likelihood ratio test compares tests the spatial filtered model against the non-filtered equivalent (Chi-squared with 20 degrees of freedom); Nano = Nanotechnology, Micro = Microelectronics, AM = Advanced Materials, AMT = Advanced Manufacturing Technologies, Ind. Biotech. = Industrial Biotechnology

The effects of the gap in hub score point towards the same direction, being negative and significant for all KETs (except *Photonics),* i.e. regions with a similar hub position in the networks tend to be linked to regions in similar central positions, indicating that also the difference in the *quality* of the links matters. In *Microelectronics,* the hub score effect is by far highest, suggesting a distinguished authority- and hub-structured network for this KET. In other words, the collaboration probability between two regions decreases when their difference in terms of quantity (degree) and quality (hub-score) of links increases, i.e. hubs are more likely to connect with other hubs than to connect with peripheral regions, which is described as homophily from a network science perspective. Interestingly, in the case of *Photonics,* the coefficient of the gap in hub score is significantly positive, indicating the presence of a 'hub and spoke' structure, where outlying regions are connected to a central hub-region; described as preferential attachment in a context of social networks.

Looking at both network structural effects – the gap in degree and the gap in hub score – they both point towards the affirmative role of *similarity* of two regions (regarding quantity and quality of research links) for the number of R&D collaborations between them. Torre and Rallet (2005) refer to this as the '*logic of similarity'* of organised proximity. Although this conception initially refers to the organisational level, it may also be applied to the regional level. In the context of our results, this could be interpreted insofar as regions that are similar in terms of research infrastructure, types of researching organisations, technological profiles, etc., share the same frameworks and systems of representation, which facilitate the ability for organisations located in these regions to interact. This holds for research-intensive regions with large numbers of organisations but also for more peripheral regions.

Interestingly, including the additional network structural separation variables, does not change the interpretation of the coefficients for the variables already included in model (1) in terms of significance and direction; however, the effects of geographical distance and the metropolitan region dummy moderately decrease in magnitude when adding these variables, i.e. these spatial effects may partly be a proxy for the other effects reflected by them, hence, not accounting for network structural variables leads to an overestimation of the geographical separation.

However, in the case of the country border effects, this relation is reversed, resulting in higher coefficients, meaning that accounting for network structural measures, country borders have an increasingly hindering effect on the expected emergence of R&D collaborations. This finding shows that, when looking for similar partners in terms of quantity and quality of their collaborations (small gap in degree centrality and hub score), national partners are more likely to be chosen, i.e. the country border gains significance. This is especially the case for the large number of small- and medium-sized organisations with a mediocre amount of network links; in contrary to large technology hubs and industry clusters in need of equivalent partners to engage in cross-regional R&D activities.

Strikingly, looking at the changes in the geographical effects, when accounting for relational effects in model (2), we again find similarities for the KETs *Microelectronics*, *AMT* and partly *Advanced Materials* as they show the largest differences, indicating relatively strong proxy effects between geographical and relational effects. Both geographical distance and the country border effect change in opposite directions, increasing the impact of the country border and decreasing the negative effect of geographical distance. Hence, within-country collaborations gain even more importance when looking for similar partners in terms of their embeddedness and connectivity. However, at the same time, the probability of long-distance collaborations increases as well.

Resuming these results in the context of our hypotheses, we conclude that hypothesis 1a and hypothesis 1b can be supported, i.e., network structural effects are indeed highly relevant for the description of R&D collaboration networks and that geographical effects change when accounting for such network structural effects. This indicates – to a certain extent – a proxy structure between these separation measures. A region's position in global R&D collaboration networks, as promoted by the EU FPs, is of tremendous importance to overcome geographical barriers such as spatial distance. Moreover, we can observe that similar regions in terms of their network centrality (both degree and hub score) show a higher probability for collaboration. This indicates that the substitution effect of networks for geographical barriers is moderated by the similarity in the network centrality between two regions. When two regions are dissimilar in their network centrality, the potential to reduce negative geographical effects is relatively lower.

Turning to the second set of hypotheses, we find that geographical and relational effects – though at stake for all technologies under consideration – are found to vary in magnitude across them, confirming hypothesis 2a. With respect to hypothesis 2b, we cannot – at least with our focus on six KETs in this study – verify our assumption that geographical effects have a stronger negative impact in engineering-oriented fields, whereas network structural effects are more critical for science-oriented fields. In fact, *Advanced Materials* and *AMT* – both being characterised as more engineering-oriented – are relatively less influenced by the negative effect of geographical distance. Moreover, the two science-oriented fields *Microelectronics*, as well as *Biotechnology* are relatively strongly hampered by country borders. Both findings contradict hypothesis 2b. However, looking at the network structural effects, we indeed find that *Microelectronics* is considerably driven by the negative effects of network structural effects but still, engineering-oriented fields, such as *Advanced Materials* and *AMT* are found to be highly affected as well. This makes it especially difficult for regions to link to hubs in terms of networks structural characteristics in these technologies.

2.7 Concluding remarks

The investigation of the spatial dynamics of R&D collaboration networks has become one of the most important research domains in regional science, accounting for their essential influence for successfully generating new knowledge, and accordingly, innovation. In the recent past, attention has been shifted to get more comprehensive and statistically robust insights into R&D network dynamics by systematically identifying and estimating determinants and drivers of real-world, observed network structures. The number of empirical works embedded in this research vein has faced an upsurge over the past ten years, related to methodological advancements, but more importantly to the recent establishment of large-scale databases enabling to trace R&D collaboration networks in space and time, covering increasingly large geographical areas and periods[19].

Empirical studies investigating determinants of R&D collaboration networks – mostly done at the regional level of analysis – have so far brought interesting results (see Scherngell 2019 for an overview), pointing to the still significant role of geographical barriers (geographical distance and/or country borders). However, these studies did not look at spatial and network structural dependencies, highlighting the role of a region's network embeddedness. Moreover, they did not yet dig into technological differences that may be prevalent across these results. Such technological heterogeneities are assumed to play a major role, given the

[19] E.g. in form of the RISIS infrastructure (see risis2.eu)

different knowledge bases and knowledge creation regimes underlying various technological fields, and accordingly different collaboration behaviours.

This study has directly addressed this research gap, aiming to identify spatial and network structural determinants of technology-specific R&D collaboration networks across a set of European regions. We have employed a spatially filtered negative binomial spatial interaction model to estimate a set of determinants, explicitly focusing on spatial effects and – in contrast to previous works – on network structural effects. By technology-specific networks, we refer to collaborative R&D projects of the EU Framework Programme (FP) observed in six Key Enabling Technologies (KETs), giving rise to six cross-region European R&D networks in different relevant technologies. In our empirical strategy, we have used the EUPRO database on EU FP projects, that contains an assignment of projects to a specific KET based on semantic technologies (see Maynard et al. 2017). The spatial interaction models are applied to each KET separately and aggregated for FP7 and H2020 for a system of 505 European metropolitan and remaining non-metropolitan regions, relating the cross-region collaboration intensity to a set of exogenous variables, in particular, spatial and network structural separation variables.

The results are highly interesting, both in the context of previous research and from a European policy perspective. In general, geographical barriers, including geographical distance and country borders, are a significant hurdle for the likelihood to establish network links across regions in the six KETs. While the negative effect of geographical distance is not surprising, the significant country border effects are somewhat pessimistic in a policy context. Negative country border effects have diminished when looking at the FP as a whole (see Scherngell and Lata 2013) but are back at stake when looking at important technological fields, such as the KETs.

Specifically, we can distinguish two groups of KETs, each sharing common characteristics in terms of their geographical effects: (i) *Nanotechnology* and *Photonics*, and (ii) *Microelectronics* and *AMT*. They appear complementary in terms of the impact of geographical barriers on R&D collaborations; whereas R&D collaborations of the first pair are strongly restricted by geographical distance with only a small impact of country border effects, the latter pair is characterised by national collaborations but at the same time driven by long-distance collaborations.

In the light of our hypotheses, the results confirm that network structural effects turned out to be indeed an important additional determinant in explaining the constitution of publicly funded technology-specific cross-region R&D collaboration networks. In this sense, results underline that network effects can compensate for geographical barriers – in all technologies investigated, although the effects differ in magnitude. However, the results also point to some logic of similarity, i.e. regions of similar network embeddedness are more likely to collaborate than regions with a high gap in their network embeddedness. A similar effect is observable for the regions' connectivity in terms of their hub score. Thus, two regions dissimilar in their network centrality have limited potential to reduce negative geographical effects. Accordingly, lagging regions in terms of network centrality face statistically significant barriers to attach to more prominent regions in the network.

Additionally, we find indeed significant and very relevant differences between the KETs under consideration, not in terms of direction and significance of the effects, but in terms of their relative importance. *Advanced Materials*, *AMT*, and *Microelectronics* seem to be less affected by geographical barriers than *Nanotechnology* and *Photonics.* For the latter, network structural effects seem to be of relatively lower importance, i.e. these KETs may be more open to non-conventional network partners than in other KETs. Hence, the assumption of engineering-oriented technological fields being more affected by geographical effects, while science-oriented fields are more driven by network structural effects, is not supported by the findings.

Some ideas for a future research agenda come to mind. *First*, the results presented in this study are static, mainly relating to the problem of the high number of zeros when going to a panel with annual observations, leading to severe estimation issues. However, advancement to a dynamic perspective to look at changes of the estimates over time is crucial and needs consideration in the future. *Second*, looking at other forms of technology-specific R&D networks should complement the results of this study that clearly focuses on a specific form of policy-induced networks. *Third*, investigating the underlying micro-dynamics of collaboration – e.g. by utilising the effect estimates from this study in a simulation approach – may provide a better understanding of the results presented here, in particular as what concerns the different determinants and their magnitude in different technological fields.

2.8 Appendix to Article I

Eigenvector spatial filtering (ESF)

Eigenvector spatial filtering (ESF) is based on the mathematical relationship between the Moran's I, as a measure for spatial autocorrelation, and spatial weight matrices. Following Griffith and Chun (2014), the purpose is to obtain a set of synthetic proxy variables by extracting them as eigenvectors from a standard spatial matrix W (see, e.g. Fischer and Wang (2011) on the construction of spatial weight matrices), and then add these vectors as control variables to the regression model. This set of variables is obtained from the transformed spatial weight matrix

$$W' = \left(I - \iota\iota'\frac{1}{N}\right) W \left(I - \iota\iota'\frac{1}{N}\right) \tag{A2.1}$$

where I is an N-by-N identity matrix, ι is an N-by-1 vector of ones and ι' is its transpose. The decomposition generates N eigenvectors $E_n = (E_1, E_2, \ldots, E_N)$ and their associated N eigenvalues $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$. As shown by Tiefelsdorf and Boots (1995), all obtained eigenvalues relate to distinct Moran's I values. Whereas the first eigenvector E_1 measures the maximum global spatial autocorrelation, the second eigenvector E_2 measures the maximum residual spatial autocorrelation after extracting the first, and so on. Generally, only a set of κ eigenvectors with MI/MI_{max} larger than 0.25 is selected as additional control variables, where MI denotes the Moran's I value (see, e.g. Fischer and Wang (2011) for details) and MI_{max} its maximum value, respectively (Fischer and Wang 2011). To apply the eigenvectors within the spatial interaction framework, it is necessary to expand them by means of the Kronecker product, which yields $E_n \otimes \iota$ in the case of the destination, and $\iota \otimes E_n$ for the origin vectors.

Descriptive statistics and additional figures

Table A2.1. *Descriptive statistics on R&D collaborations in six KETs*

	Nano	Micro	Photonics	AM	AMT	Biotech
# All links	255,025	255,025	255,025	255,025	255,025	255,025
# Positive links	38,822	16,480	35,092	11,451	24,785	46,229
% Zero links	84.78	93.54	86.24	95.51	90.28	81.87
# Intra-regional collaborations	2,774	1,820	2,464	323	1,076	3,534
# Inter-regional collaborations	77,245	23,940	64,506	10,364	38,678	109,329
# Organisations	5,189	1,820	4,559	1,298	2,363	5,912

Notes: # denotes 'number' Nano = Nanotechnology, Micro = Microelectronics, AM = Advanced Materials, AMT = Advanced Manufacturing Technologies, Ind. Biotech. = Industrial Biotechnology

Table A2.2. *R&D collaboration network characteristics of six KETs*

	Nano	Micro	Photonics	AM	AMT	Biotech
Number of edges	19,510	8,278	17,754	5,709	12,467	23,295
Number of vertices	453	333	449	341	382	463
Density	0.19	0.15	0.18	0.10	0.17	0.22
Degree centralisation	0.66	0.69	0.68	0.57	0.58	0.65
Mean degree	86.16	49.72	79.08	33.48	65.27	100.63
Maximum degree	384	278	383	227	285	403
Betweenness centralisation	0.05	0.10	0.06	0.10	0.05	0.04
Transitivity	0.49	0.43	0.47	0.35	0.53	0.52

Notes: Nano = Nanotechnology, Micro = Microelectronics, AM = Advanced Materials, AMT = Advanced Manufacturing Technologies, Ind. Biotech. = Industrial Biotechnology

Table A2.3. *Descriptive statistics of regression variables*

Number of R&D collaborations (dependent variable)						
	Nano	Micro	Photonics	AM	AMT	Biotech
minimum	0	0	0	0	0	0
mean	0.62	0.19	0.52	0.08	0.31	0.87
median	0	0	0	0	0	0
maximum	485	276	552	42	186	619
Origin/ destination						
	Nano	Micro	Photonics	AM	AMT	Biotech
minimum	0	0	0	0	0	0
mean	10.51	3.79	9.25	2.62	4.80	11.90
median	5	1	4	1	2	5
maximum	204	116	201	49	101	222
Geographical distance						
	Nano	Micro	Photonics	AM	AMT	Biotech
minimum	0	0	0	0	0	0
mean	1090.1	1090.1	1090.1	1090.1	1090.1	1090.1
median	1007.7	1007.7	1007.7	1007.7	1007.7	1007.7
maximum	3942.8	3942.8	3942.8	3942.8	3942.8	3942.8
Gap degree centralities						
	Nano	Micro	Photonics	AM	AMT	Biotech
minimum	0	0	0	0	0	0
mean	81.72	43	74.6	29.67	60.62	92.16
median	61	28	55	18	43	73
maximum	382	278	379	228	285	401
Gap in hub score						
	Nano	Micro	Photonics	AM	AMT	Biotech
minimum	0	0	0	0	0	0
mean	0.040	0.030	0.033	0.051	0.048	0.045
median	0.013	0.009	0.011	0.020	0.016	0.014
maximum	1	1	1	1	1	1

Notes: Nano = Nanotechnology, Micro = Microelectronics, AM = Advanced Materials, AMT = Advanced Manufacturing Technologies, Ind. Biotech. = Industrial Biotechnology

Table A2.4. *Correlations between dependent variables in six KETs*

	Origin/ destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
Nanotechnology						
Origin/destination	1.000	-0.020	-0.019	0.189	0.323	0.368
Geogr. distance	-0.020	1.000	-0.548	-0.063	0.064	-0.015
Country border	-0.019	-0.548	1.000	0.052	-0.055	-0.010
Metro region	0.189	-0.063	0.052	1.000	0.106	0.145
Gap degree centralities	0.323	0.064	-0.055	0.106	1.000	0.500
Gap in hub score	0.368	-0.015	-0.010	0.145	0.500	1.000
Microelectronics						
	Origin/ destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
Origin/destination	1.000	-0.004	-0.010	0.189	0.462	0.394
Geogr. distance	-0.004	1.000	-0.548	-0.063	0.044	0.015
Country border	-0.010	-0.548	1.000	0.052	-0.037	-0.018
Metro region	0.189	-0.063	0.052	1.000	0.151	0.126
Gap degree centralities	0.462	0.044	-0.037	0.151	1.000	0.477
Gap in hub score	0.394	0.015	-0.018	0.126	0.477	1.000
Photonics						
	Origin/ destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
Origin/destination	1.000	-0.021	-0.004	0.170	0.329	0.348
Geogr. distance	-0.021	1.000	-0.548	-0.063	0.064	-0.009
Country border	-0.004	-0.548	1.000	0.052	-0.048	-0.006
Metro region	0.170	-0.063	0.052	1.000	0.108	0.128
Gap degree centralities	0.329	0.064	-0.048	0.108	1.000	0.473
Gap in hub score	0.348	-0.009	-0.006	0.128	0.473	1.000
Advanced materials (AM)						
	Origin/ destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
Origin/destination	1.000	-0.012	-0.012	0.148	0.451	0.402
Geogr. distance	-0.012	1.000	-0.548	-0.063	0.043	0.005
Country border	-0.012	-0.548	1.000	0.052	-0.034	-0.007
Metro region	0.148	-0.063	0.052	1.000	0.123	0.132
Gap degree centralities	0.451	0.043	-0.034	0.123	1.000	0.603
Gap in hub score	0.402	0.005	-0.007	0.132	0.603	1.000
Advanced manufacturing technology (AMT)						
	Origin/ destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
Origin/destination	1.000	-0.022	-0.012	0.152	0.411	0.405
Geogr. distance	-0.022	1.000	-0.548	-0.063	0.035	-0.008
Country border	-0.012	-0.548	1.000	0.052	-0.031	-0.008
Metro region	0.152	-0.063	0.052	1.000	0.118	0.145
Gap degree centralities	0.411	0.035	-0.031	0.118	1.000	0.534
Gap in hub score	0.405	-0.008	-0.008	0.145	0.534	1.000
Industrial Biotechnology						
	Origin/ destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
Origin/destination	1.000	0.001	-0.027	0.179	0.292	0.380
Geogr. distance	0.001	1.000	-0.548	-0.063	0.075	-0.013
Country border	-0.027	-0.548	1.000	0.052	-0.054	-0.016
Metro region	0.179	-0.063	0.052	1.000	0.091	0.139
Gap degree centralities	0.292	0.075	-0.054	0.091	1.000	0.487
Gap in hub score	0.380	-0.013	-0.016	0.139	0.487	1.000

Figure A2.1. *Metropolitan and non-metropolitan regions*

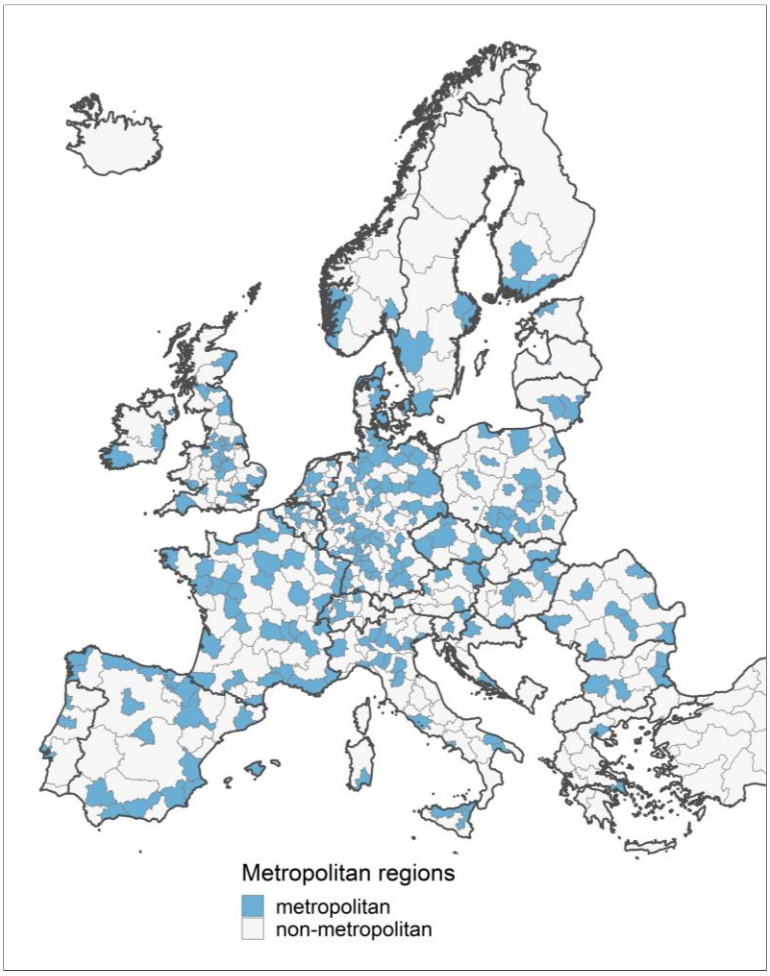


Figure A2.2. *Spatial R&D networks of KETs (2007-2016)*

(a) Nanotechnology	(b) Microelectronics

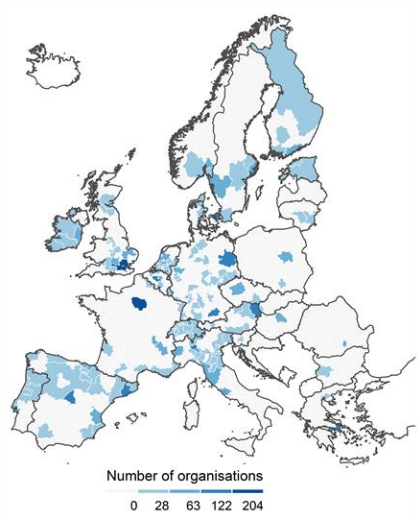
(c) Photonics	(d) Advanced Materials

(e) Advanced Manufacturing Technology	(f) Industrial Biotechnology

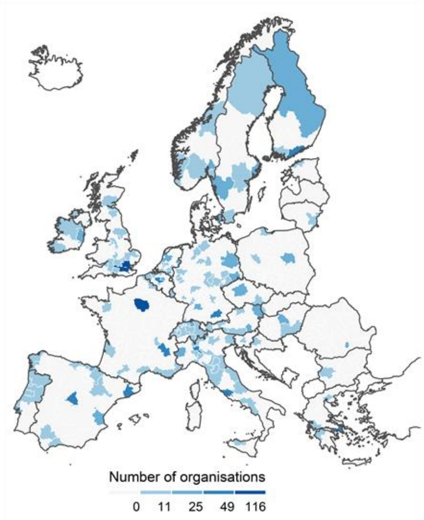
Note: Only top 95% of links in terms of collaboration frequency are displayed

Figure A2.3. *Spatial distribution of organisations in KETs (2007-2016)*

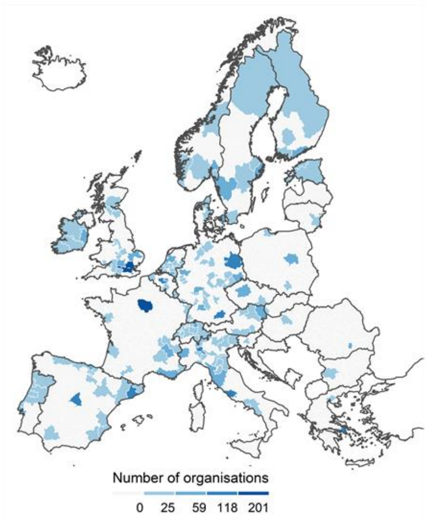
(a) Nanotechnology



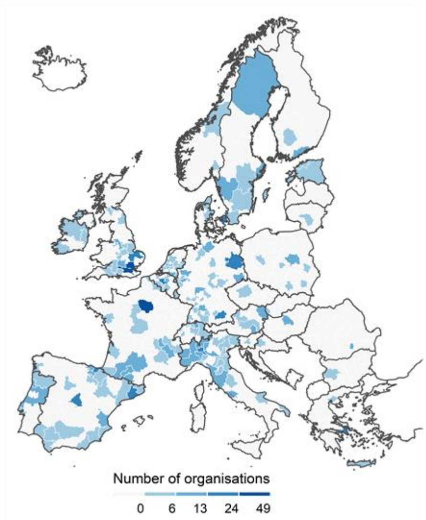
(b) Microelectronics



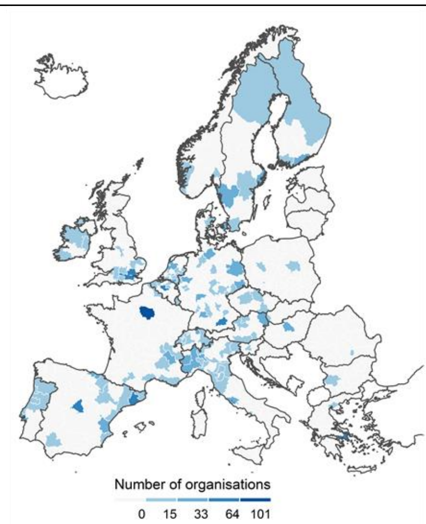
(c) Photonics



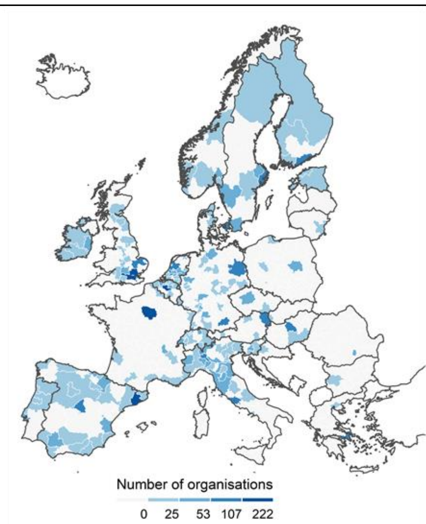
(d) Advanced Materials



(e) Advanced Manufacturing Technology



(f) Industrial Biotechnology



3 Impacts of EU funded R&D networks on the generation of Key Enabling Technologies: Empirical evidence from a regional perspective (Article II)

This section is based on the study *"Impacts of EU funded R&D networks on the generation of Key Enabling Technologies: Empirical evidence from a regional perspective"* (joint work with Iris Wanzenböck and Thomas Scherngell, published in *Papers in Regional Science*, 2020)

Abstract: Cross-regional R&D collaborations are essential for regional innovativeness. However, we lack insights into technology field-specific effects of a region's network connectivity. This study investigates Key Enabling Technologies (KETs) to compare knowledge creation effects of EU funded R&D networks for different technological fields. We apply a spatially filtered regression model together with marginal effect interpretations for non-linear models to quantify and compare network effects. The generally positive network effects differ depending on region-internal endowments and the nature and development stage of the underlying technologies. Policy implications arise for the interrelations between EU research, industrial and regional policy.

Keywords: R&D networks, Key Enabling Technologies (KETs), regional knowledge production, network embeddedness, interaction effect

3.1 Introduction

Following the theoretical debate in the first decade of the 21st century on benefits of collaborative research and development (R&D) for knowledge creation and innovation, first works providing respective empirical evidence have appeared in the recent past (Ponds et al. 2009, Fornahl et al. 2011, Varga et al. 2014, Wanzenböck and Piribauer 2018). Generally, these works indeed underline that – as suggested by theory – knowledge-intensive organisations increasingly mobilise non-local collaborations and international networks to externally access the knowledge required to perform their research and innovation activities. From a regional perspective, high network interconnectivity can stimulate region-internal innovation activities due to the inflow of new knowledge via cross-regional collaborations. This enriches the local knowledge base or even supports the diversification and technological renewal of entire sectors or regions (e.g. Boschma and Frenken 2010).

To leverage the positive mechanisms associated with such cross-regional R&D collaborations, policy measures have been increasingly implemented at the regional, national and supra-national level. In the latter context, the European Framework Programmes (FPs) – the most prominent example in terms of funding assigned – is recently widely discussed in the literature (Autant-Bernard et al. 2007b, Scherngell and Barber 2009, Sebestyen and Varga 2013, Wanzenböck et al. 2015). Wanzenböck and Piribauer (2018) provide robust original evidence on the space-time impacts of regional FP network embeddedness at the level of European regions, but at an aggregated level are neglecting technological idiosyncrasies. Thus, we shift the debate to whether and how EU-wide R&D networks contribute to a region's knowledge creation capability in different technological fields. In this regard, we are interested in whether the overall positive influence of collaborative R&D holds when we investigate specific technologies, given the different modes of knowledge creation, the heterogeneities in the underlying knowledge, or the development stage of the technologies. By assuming that such heterogeneities crucially influence the relevance and spatial scale of R&D collaboration, this paper aims to estimate how regional embeddedness in EU funded R&D networks affects regional knowledge creation across different technological fields and to disentangle potential similarities and differences in terms of the estimated impacts between the technologies under consideration.

To gain this more fine-grained understanding of the role of inter-regional R&D networks, we need to discriminate and observe knowledge production in different technological fields in our empirical framework. Here, we focus on the case of Key Enabling Technologies (KETs)[20] as a specifically relevant example in a European policy context. KETs, considered generic technologies, can serve as basic technological input for various innovations adopted in a variety of different industries. Given their horizontal and systemic nature, the opportunities for cross-sectoral and cross-regional spillovers, especially between leading and lagging regions, are generally seen as high (Montresor and Quatraro 2017). For regions or countries, specialisation in KETs is associated with more sustainable innovation paths, enhanced ability to build linkages across industries and higher potentials for diversifying into new sectors EC 2009, 2012, 2015a). In line with the ideas of a new form of industrial policy approach (Rodrik 2014, Foray 2016), one of the current priorities of the EU is to foster research and capability building activities around KETs and to induce industrial change, particularly in structurally weak regions. However, the empirical studies of Montresor and Quatraro (2017) and Evangelista et al. (2018) show that the spatial distribution of KETs is highly concentrated on certain regions in Western and Central Europe with high regional disparities across Europe.

To observe the R&D network structures in different KETs, we rely on the definition and the distinction between technologies introduced by the European Commission (EC 2015b). We identify relevant projects funded by FP7 and construct KET-specific R&D networks at the regional level based on relevant keywords. Social network analytic (SNA) centrality measures are used to calculate a region's positioning in the field-specific networks. Our regional sample consists of a set of 257 European NUTS 2 regions. The empirical model we are employing assumes that a central network positioning is only conditionally useful for generating new knowledge. The resources and skills available in a region significantly might moderate how external knowledge can be absorbed and utilized. In analogy to the study of Wanzenböck and Piribauer (2018), we control for the interaction between regional R&D

[20] The six fields under consideration are 1) *Nanotechnology*, 2) *Microelectronics*, 3) *Photonics*, 4) *Advanced Materials*, 5) *Advanced Manufacturing Technologies*, and 6) *Industrial Biotechnology*. The notion of Key Enabling Technologies (KETs) has been introduced by the EU (Montresor and Quatraro 2017). From a scientific point of view, the roots of the concepts show similarities with the notion of General Purpose Technologies (GPT) (Bresnahan and Trajtenberg 1995, Lipsey et al. 2005, Qiu and Cantwell 2018), or emerging technologies (as for *Biotechnology*, for instance, see Rotolo et al. 2015). A well-defined framework for KETs, their specific characteristics and demarcation to other related concepts has not been developed yet. However, the study at hand focuses not on the semantic properties of the concept but rather on providing systematic empirical evidence with respect to differing technological fields (Note: this research has been conducted prior to the work done in Article I, where an ontology developed in the EU funded project KNOWMAK is used to delimitate the specific KET fields).

network embeddedness and own region characteristics in an augmented regional Knowledge Production Function (KPF). A methodological advancement of this paper is that we account for potential nonlinearities in a spatially filtered negative binomial model. Moreover, we introduce adjusted marginal effect interpretations to quantify potential interaction effects between region-external network participation and region-internal resources, on the one hand, and compare these effects across KETs, on the other hand. The identified R&D network effects are consistent with the differentiation into (i) science-based fields, building on scientific inputs, a more explicit knowledge base and global knowledge transmission patterns, as well as (ii) application-oriented fields in which practical experience and more localised or informal knowledge exchange process may be prevailing.

This paper is organised as follows. In Subsection 3.2, we provide the reasons for the study of R&D network structures and effects under the lens of technological heterogeneity before we present our approach to construct the technology-specific R&D networks and to calculate a region's network embeddedness therein (Subsection 3.3). In Subsection 3.4, we introduce our empirical model specification relating R&D networks and region-internal endowments to knowledge creation in KETs. In Subsection 3.5, we derive the marginal effect calculations necessary to derive comparable results for the different KET fields. Subsection 3.6 discusses the estimation results before Subsection 3.7 concludes in light of the technological heterogeneities observed.

3.2 Technological heterogeneity of R&D networks

The technological heterogeneity of R&D and knowledge interactions has been stressed intensively in the literature, initially in a sectoral systems of innovation context (see Malerba 2002). Here, knowledge creation is considered a non-linear and heterogeneous process, characterised by the specific interplay of actors and the technological knowledge they create, absorb and transmit across geographical space. To address these heterogeneities, several scholars proposed vital conceptualisations or taxonomies that enable distinguishing different 'modes' of knowledge creation across fields and over time (see, e.g. Pavitt 1984/2005, March 1991, Jensen et al. 2007). Based on such conceptualisations, heterogeneities concerning the specific rules, forms of interaction or capabilities and resources predominant in specific domains can be analytically disentangled and compared across scientific, technological or industrial fields (see, e.g. Asheim 2007, Moodysson 2008). They are useful not only to determine fundamental characteristics, similarities or differences between fields but also to derive implications for the role of region-external R&D networks in each field.

In the field of innovation economics, heterogeneities between technologies are usually investigated with respect to the complexity of knowledge combinations (Fleming and Sorenson 2001, Antonelli 2011, Balland and Rigby 2017), the learning processes according to the nature of the knowledge base (Asheim 2007, Moodysson 2008), or the development stage of a technology and its relatedness to existing knowledge (Anderson and Tushman 1990, Heimeriks and Boschma 2014). Implicitly or explicitly, each approach may bear different indications for the predominant spatial structure and relevance of collaborative networks. For instance, based on the observation that the mobility of less complex knowledge is higher than that of more complex knowledge as it requires less communication and interaction (Sorenson et al. 2006), we could conclude that long-distance collaborations are less likely in fields characterised by a higher complexity in the knowledge creation process. Furthermore, also linked to the nature of the knowledge is the differentiation between 'codified' and 'tacit' elements in knowledge or technological development processes (see, e.g. Maskell and Malmberg 1999, Howells 2002). It is assumed that the degree of informal learning based on routines and experiences determines whether knowledge or distinct capabilities can be better transmitted locally or effectively travel long distances. This view suggests that more incremental modes of technological development, tied to the domestic industrial production processes, would favour localised knowledge exchange over region-external knowledge sourcing. However, if a technology or its underlying knowledge base is more explicit and has strong scientific elements, such as in biotechnology or nanotechnology (Tamada et al. 2006, Bozeman et al. 2007), new, usually quite specific technological inputs are more often based on university research and drawn from selected partners located also outside the region (Asheim 2007). Hence, the role and spatial scale of R&D networks are likely to differ across technological fields, but a clear-cut answer is challenging to find in theory.

Recently, scholars started to take a dynamic perspective on the evolution of technologies and the role and structure of network linkages (Ter Wal 2013, Balland et al. 2015). Here, the evolution path of a technology or an industry can serve as a critical conceptual vehicle for characterising the different phases of development and, with that, the changing geographical patterns of collaboration and innovation. Generally, the spatial structure of collaboration is considered in a state of flux with advancing technological maturity. High uncertainty and need for frequent interaction may support geographically clustered R&D activities and collaboration in the early stages of a technology, while a higher degree of standardisation, the exploitation of dominant designs and diffusion of technologies may lead to geographically more dispersed linkages among actors in later stages of an industry or technology (Ter Wal 2013).

Based on this discussion, it is reasonable to assume that knowledge creation effects of R&D network embeddedness depend on the technology under consideration. However, comprehensive investigations and studies allowing for comparisons on the role of networks in different technological fields are still missing. KETs, as applied in this study, provide a new inroad to endeavours of studying technological heterogeneities. Outstanding in the conception of the six KET fields is the degree of heterogeneity between them. The fields differ considerably in the predominant modes of knowledge creation (e.g. science-based versus applied), the institutional or organisational composition of important actors (e.g. university-related versus SME), or their interweaving with domestic industrial production structures. Hence, the technological fields referring to individual KETs serve as an interesting starting point to dig deeper into the question of whether the general results, drawing a positive association between cross-regional R&D collaborations and regional knowledge creation, also hold for specific technologies. Potential differences between the technologies may be related to the debate on technological heterogeneities in R&D and differing knowledge creation modes.

3.3 Identifying technology-specific R&D networks

At this point, we describe the empirical strategy employed to observe the cross-regional R&D network activities disaggregated by KET fields. We use the KET classification to compare six different technologies in our comparative analysis of network effects on technology-specific knowledge creation in European regions, as later described in Section 3.4. This study covers a set of 257 NUTS 2 regions located in the EU-27 countries (excluding Croatia, including the UK).

Observing EU funded R&D networks in KET fields

The R&D networks we are interested in can be defined as a set of knowledge linkages between organisations jointly involved in a KET-specific research project funded by the EU FP. Due to the FP's funding requirements, all projects involve – although to a different degree – cross-regional (cross-national) knowledge linkages. Data on FP projects is drawn from the EUPRO database, which contains basic information on the project (such as duration, objective, etc.), the partners involved including the assignment to a NUTS 2 region, as well as the specific funding scheme and type of call under which the project was supported (Scherngell and Barber 2009)[21]. Since the projects are not pre-classified into technological fields, we queried the database and selected all KET-relevant projects manually. Given this study's industrial focus and time frame, we restricted the search to cooperative R&D projects funded under FP7[22] with a starting date in 2007-2013.

To identify the relevant projects for the different fields, we reviewed in a first step the available project reports containing descriptions and definitional issues concerning the six KETs (Aschhoff et al. 2010, EC 2015b). Based on this screening, we have created a list of keywords (see Table A3.1 in the appendix to this subsection) containing the defining terms for each field. In a second step, we have searched for these keywords in the project title, objective and project description and assigned the relevant FP7 projects to respective KET fields. Finally, in a third step, we have manually checked the obtained results of our queries and validated the assignment to the different fields. We have browsed the project descriptions if necessary and deleted projects without a specific R&D goal (e.g. coordinative or support projects).

Based on the retrieved information, we specify six different KET-specific R&D networks in which the nodes constitute the organisations interlinked due to their joint project participation. Table 3.1 provides some basic SNA statistics on the different networks under consideration (see Wasserman and Faust 1994 for a description of these measures). Although the networks' sizes differ considerably, basic network characteristics such as density or average clustering are relatively similar between the different KET fields. The

[21] EUPRO is publicly available for research purposes within RISIS, an integrated research infrastructure for research and innovation policy studies (risis2.eu)

[22] All projects funded under the Programmes "people", "ideas" and "capacities" were excluded from the project query. Furthermore, we exclude collaborative projects in the field of social sciences and related to the thematic area of "Socio-economic sciences and the humanities".

largest network in terms of the participating organisations and the number of projects can be observed for *Photonics*, while the smallest networks are the *Nanotechnology* and *Microelectronics* networks.

The number of participating organisations and funded projects in *Photonics* is, for instance, four times as high as in the *Nanotechnology* network, although the ratio between organisations and projects is the smallest in *Photonics* compared to all other KET networks. For all networks, the share of linkages within a region is below 10%. Interestingly, the *Photonics* network seems to be the most 'inclusive' as almost 90 % of the European regions are included with at least one participating organisation; in the *Microelectronics* and *Nanotechnology* networks, in contrast, only around 60 % of regions are represented.

Table 3.1. *The KET networks: descriptive statistics*

	Nano-technology	**Micro-electronics**	**Photonics**	**Advanced Materials**	**Advanced Manufacturing Technologies**	**Industrial Biotechnology**
Organisations (=network nodes)	563	449	2360	917	906	753
Projects	87	95	601	149	158	129
Edges	5,368	4,335	31,735	10,789	8,017	7,582
Avg. degree	19.07	19.31	26.89	23.53	17.70	20.14
Max. degree	207	242	1236	410	367	179
Std. dev. degree	17.22	23.29	49.17	28.73	18.19	19.98
Avg. clustering	0.59	0.54	0.54	0.60	0.56	0.57
Network density	0.03	0.04	0.01	0.03	0.02	0.03
Share intra-regional linkages	6%	6%	7%	6%	7%	6%
Participating regions	63%	56%	88%	76%	74%	69%

Notes: Technology-specific networks are constructed at the organisational level. The regional sample consists of 257 European NUTS 2 regions. *Degree* denotes the number of links, i.e. project participations, of an organisation. Accordingly, *Avg. degree* denotes the average degree of all organisations in the network, *Max. degree* the maximum number of participations, and *Std. dev. degree* the standard deviation of all degrees observed. *Avg. clustering* is the average clustering of the organisations, i.e. the share of closed triangles in the network, while network *density* denotes the ratio between the observed and the maximum possible number of links (see Wasserman and Faust 1994 for a formal definition and further details). *Participating regions* refers to the share of regions with at least one participating organisation in the networks.

Measuring regional embeddedness in technology-specific R&D networks

Given our interest in a region's embeddedness in KET related networks, we need to aggregate these organisation level networks to the regional level. Here we follow the approach introduced by Wanzenböck et al. (2015). We first calculate the normalised degree centrality for each organisation, and in a second step, aggregate these values based on the regional assignment of organisations. The degree centrality is a local centrality measure that takes the number of network participations into account (Wasserman and Faust 1994). We regard the degree centrality as the most useful indicator for reflecting a region's embeddedness in technology-specific R&D networks, given its simplicity of calculation and interpretation compared to other more complex global measures. Especially for the interpretation of the marginal effects as employed in this study (see Section 3.5), degree centrality is the most useful centrality measure. The correlation among typical centrality measures is usually positive and high, in particular at the regional level of analysis (Valente et al. 2008, Wanzenböck et al. 2015).

As a first exploratory step of the empirical analysis, Figure A3.1 (in the appendix to this section) illustrates the spatial distribution of the region-level degree centrality scores for each KET. Not surprisingly, we observe the 'star' role held by the region of Île-de-France (Paris), as well as a generally strong dominance of the industrial core regions in Germany, France, Italy, Spain, and Scandinavia given their high values of degree centrality in all networks. However, we also see that degree centrality in *Industrial Biotechnology*, even though concentrated on Paris, is spread more equally among the remaining regions. In contrast, network centrality in *Microelectronics* or *Nanotechnology*, for instance, concentrates on a few hubs in Europe and is more unequally distributed over all regions.

3.4 Empirically modelling regional knowledge creation

To estimate the impact of R&D networks on regional knowledge creation in KET fields, we build on an extended Knowledge Production Function (KPF) approach as introduced in Wanzenböck and Piribauer (2018) and regard the effects of R&D network embeddedness as dependent on other regional knowledge inputs[23]. In this study, we provide measures to disentangle potential interactions and assess the effects of network embeddedness in a multiregional setting, on the one hand, and compare them between different technological fields, on the other hand.

[23] Key to the argument of including such an interaction relationship is the assumption that embeddedness in inter-regional R&D activities is driven by the skills or capabilities located within the region, determining the access and attractiveness in partnerships as well as the opportunities to exploit knowledge from external sources (Wanzenböck and Piribauer 2018). Furthermore, own capabilities influence the importance of engaging in long-distance collaboration. While regions with own well-functioning innovation systems might be less dependent on external linkages, for them the challenge is rather to find the right type of knowledge or the right partner in light of limited relational capacities. The trade-off for regions having a strong internal knowledge base is rather that maintaining a large set of network relations demands resources and causes costs, while the benefits and learning effects of many relations might be relatively small.

Formally, the basic empirical model considered can be written as

$$Y_{ik} = \alpha + \beta\, X_{ik} + \gamma Z_i + u_{ik} \tag{3.1}$$

where Y_{ik} represents the outcome variable denoting knowledge creation in region i in a specific KET field k, X_{ik} is a matrix of variables associated with a regions network embeddedness in a specific technological field and Z_i a variable matrix reflecting other region-specific characteristics that might influence regional knowledge creation propensity. α denotes a constant, β and γ are respective response coefficients, and u_{ik} captures the disturbances in our modelling relationship. The conditional dependence between network embeddedness and regional endowments is considered in the form of

$$X_{ik} = [c_{ik}, c_{ik}\ \times h_i] \tag{3.2}$$

where c_{ik} denotes the region's centrality in the KET-specific network as discussed in Section 3.3, and h_i is some measure for the skills and resources located within the region. The term $c_{ik}\ \times h_i$ then denotes the interaction between these two variables.

To measure regional knowledge output Y_{ik}, we use the number of regional patents observed in different KET fields from 2009 to 2013. According to the classification developed by Aschhoff et al. (2010) on behalf of the European Commission, we assign individual patents to KET fields based on the IPC codes listed on the patents (the list of IPC codes assigned to the KETs is provided in Table A3.2 (in the appendix of this section). Notably, we use full counting and the inventor's location to calculate the number of patents in a region. All patent information is derived from the OECD REGPAT database.

For our independent variables, we draw on the percentage of the population with tertiary education available in Eurostat, as a proxy reflecting the general quality of human resources in a region[24]. Additionally, we calculate the share of intra-regional network linkages on all FP linkages in a specific KET field to include a control variable for domain-specific capabilities within a region. Accordingly, the variable can be considered a proxy for the relative inward orientation of regional linkages, suggesting a strong regional knowledge base or innovation system in a specific field. To control for more general regional

[24] In contrast to our R&D network embeddedness variable, observations on the technology-specific skills are unfortunately not available for the 257 European NUTS 2 regions in our sample. We run robustness checks using data on the number of persons with tertiary education and/or employed in science and technology (as % of active population), which delivered similar results for the technology-specific models.

characteristics (Z_i), we consider R&D expenditures in the business sector (in % of GRP) as a control variable to reflect the R&D intensity and financial R&D inputs of the domestic industries. Given that KETs are technologies close to industrial manufacturing, the employment in the manufacturing sector (in % of the total employment) is used as an additional proxy for the size of the industry sector in the region. A detailed variable description is provided in Table A3.3 in the appendix of this section.

We estimate a negative binomial (NegBin) regression model given the non-negative count data nature of our dependent variable, that is the number of regional patents. Furthermore, we follow recent works and apply an Eigenvector spatial filtering approach to remove the spatial dependence bias from the estimated parameters (Scherngell and Lata 2013, Lata et al. 2015; see Subsection 2.8 for details on Eigenvector spatial filtering). The basic model of Equations (3.1) with (3.2) expressed as spatially filtered model version takes the form of

$$E(Y_{ik}|X_{ik}, Z_i) = \exp\ (\beta_c c_{ik} + \beta_h h_i + \beta_{c\times h}(c_{ik} \times h_i) + \sum_{z=1}^{Z} \gamma_z z_i + \sum_{m=1}^{M} \theta_m V) \tag{3.3}$$

where matrix V comprises M Eigenvectors of a first-order contiguity spatial weights matrix serving as spatial filters[25]. θ_m denotes the respective vector of coefficients for the spatial filters. As we are interested in technology-specific heterogeneities, we run individual regressions for the six KET fields under consideration.

3.5 Marginal effect measures for the role of R&D network embeddedness

The marginal effect of a variable – analytically defined as the partial derivative of the model – usually allows for comparison of the relative size and significance of the respective model parameters. In our case, however, both the interaction effect in the set of independent variables and the negative binomial model specification (Equation 3.3) induce nonlinearities, which restricts direct interpretations of regression coefficients as they were marginal effects. The fact that adjustments are needed for interactions in nonlinear models to identify the magnitude correctly, sign and significance of the interaction effect has often

[25] To avoid overfitting problems, we follow Fischer and Griffith (2008) and add not the full set of Eigenvector to our model to be estimated, but only the relevant ones that are showing a certain degree of spatial dependence. We measure the degree of spatial dependence by means of the Moran's I test and include 58 Eigenvectors.

been disregarded in applied econometrics (see, e.g. the discussions in Ai and Norton 2003, Greene 2010, Karaca-Mandic et al. 2012, Tsai and Gill 2013).

In this study, we use marginal effect calculations as introduced in Ai and Norton (2003). They derive appropriate marginal effect expressions for nonlinear models with an interaction between two explanatory variables. Only in this way can we fully disentangle the effects of R&D network linkages from the presence of internal capabilities in a region and compare these (marginal) effects across the different KET fields. It is important to note that the usual interpretations associated with the marginal effect of an explanatory variable remain despite these adjustments. In our case, that is: How much, on average, does the number of patents in a region change when we increase the network centrality (or any other independent variable) per one unit?

Following Equation (3.3), the *individual* marginal effect $\Theta_{(c_{ik})}$, here stated for a region's network embeddedness in a technology-specific network, on the conditional expected value of Y_{ik} can be expressed as

$$\Theta_{(c_{ik})} = \frac{\partial E(Y_{ik}|X_{ik},\ Z_i)}{\partial c_{ik}} = [\beta_c + \beta_{c\times h} h_i]\ E(Y_{ik}|X_{ik}, Z_i). \tag{3.4}$$

It is easy to see that the partial effect of c_{ik} in a region depends not only on the value of the interacted variable h_i but is also conditional on the expected value of Y_{ik}. Consequently, the marginal effects are not constant over the range of a variable but depend on the value of all covariates in the model and are subject to variation across regions. All marginal effects are identified for each region as region-specific (or individual) marginal effect; the mean of these individual marginal effects gives us the average marginal effect for our regional sample (Ai and Norton 2003). The individual marginal effect of a region's human resources is calculated in the same way as in Equation (3.4), with β_h being the main term and $\beta_{c\times h}$ the interacted term.

A similar argument as for the marginal effect of a main variable holds for the interpretation of the interaction term: As pointed out by Ai and Norton (2003), the interaction effect in nonlinear models is neither the regression coefficient of the interaction term nor can it be simply computed by its marginal effect. In analogy to the marginal effects of a main variable, it is conditional on all independent variables, and thus, may even show different signs for different values of the covariates. Hence, we need to demarcate the effects induced only by the interacted variables (product-term induced interaction) from those effects induced by the value of other covariates (model inherent interaction; Ai and Norton 2003, Greene 2010). For this study, the product-term induced interaction effects are of major interest as we aim at determining the conditional dependence between network embeddedness and regional skills in the generation of new knowledge in a KET field. Formally, the product-term induced interaction effect denoted as $\mathrm{I}_{c\times h}$, can be expressed as the cross-derivative of the expected value of Y_{ik} in terms of

$$\mathrm{I}_{c\times h} = \frac{\partial^2 E(Y_{ik}|X_{ik},\ Z_i)}{\partial c_{ik}\partial h_i} = \frac{\Theta_{(c_{ik})}}{\partial h_i} \tag{3.5}$$

measuring the change of the marginal effect of a region's network centrality when we change the regional skills by one unit. Details on the estimation and asymptotic details to identify the significance of the main term and the interaction effect can be found in Ai and Norton (2003).

Furthermore, when comparing KETs, we have to be aware that the marginal and interaction effects are not independent of the conditional expected value of Y_{ik}. The value of $E(Y_{ik}|X_{ik}, Z_i)$, and consequently, also the magnitude of our effect estimates, depending on the range of the dependent variable, which in our case is highly determined by the patent intensity in a technological field. Without normalisation, our marginal effect estimates for *Industrial Biotechnology*, for instance, would exceed the magnitude of those in all other fields only due to the incomparably higher patenting intensity in this field. To achieve a measure that allows valid comparison across KET fields, we normalise the predicted counts for each region with the maximum in the respective KET before calculating the marginal effects according to Equations (3.4) and (3.5).

3.6 Empirical results

We first estimate six regression models, according to Equation (3.3), one for each KET. The results for the regression models and associated test statistics can be found in the appendix to this section (Subsection 3.8). We perform LR tests to check the robustness regarding the inclusion of the interacted variables (see Table A3.5 in the appendix to this section). For all models, the NegBin specification is confirmed as shown by a significantly positive dispersion parameter. In a second step, we calculated the marginal effects based on Equations (3.4) and (3.5). The results expressed in terms of averages of the region-specific estimates are presented in Table 3.2. Note that the displayed effect sizes of both the network centrality variable and the human resource variable (i.e. our interacted variables) refer to the total effects, as they incorporate the effects of the main term and the interacted term and are calculated according to Equation (3.4). In our discussion, we focus first on the total effect of R&D network centrality on regional knowledge creation and how the effect differs by KET before we investigate the conditional dependence of regional network centrality and the human resources in more detail. For our remaining (control) variables, we present only the most striking results. While comparisons of effect sizes across fields are valid, comparisons across the independent variables need to be performed with caution and in light of how the respective variables are measured (in shares or absolute values).

We find a significantly positive effect of network centrality on regional patenting intensity in all KET fields, confirming the positive influence of cross-regional networks on a region's invention potential as found in previous studies (see, e.g. Ponds et al. 2010, Wanzenböck and Piribauer 2018). These network effects, however, differ in strength depending on the specific technology under investigation. Mean marginal effects are exceptionally high in *Nanotechnology*, meaning that a one-unit increase in a region's connectedness in the field of *Nanotechnology* increases a region's patenting intensity in this field more likely than it is the case for other fields. For instance, the effect for *Nanotechnology* exceeds by more than double the amount we observe for *Industrial Biotechnology*, the field with the second-highest marginal effect on regional patenting. These results seem to reflect the importance of networks as mechanisms for spillovers – within and across regional borders – in science-based industrial fields (see also the findings of Ponds et al. 2009). In contrast, the observed network effects are of relatively low value in *Advanced Materials, Advanced Manufacturing Technologies* and *Microelectronics*, i.e. fields that are typically linked more closely to industrial production, with engineering-based and often more informal knowledge creation.

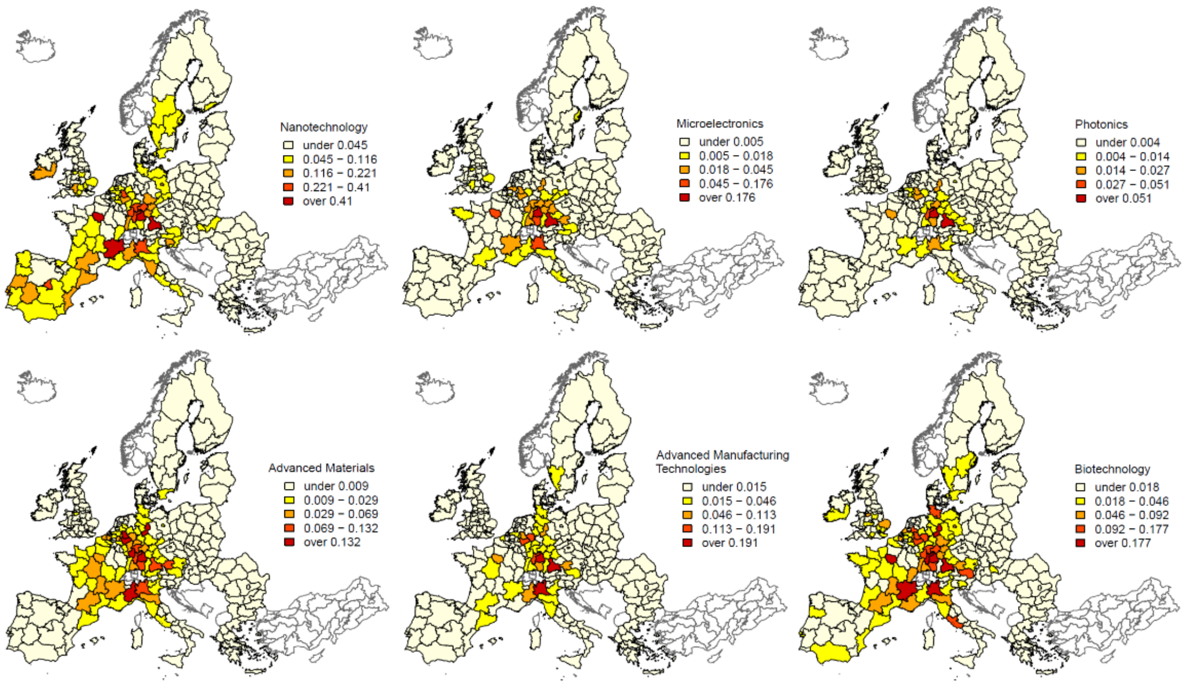
Table 3.2. *Mean of marginal effect estimates of the models with interaction*

	Nanotechnology		*Microelectronics*		*Photonics*		*Advanced Materials*		*Advanced Manufacturing Technologies*		*Industrial Biotechnology*	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Mean of marginal effects												
Network centrality	0.058	*(0.006)*	0.008	*(0.002)*	0.004	*(0.002)*	0.013	*(0.002)*	0.013	*(0.002)*	0.027	*(0.003)*
Human resources	0.138	*(0.042)*	-0.066	*(0.057)*	0.113	*(0.014)*	0.177	*(0.075)*	0.156	*(0.021)*	0.077	*(0.120)*
Business RD exp.	1.180	*(0.144)*	0.624	*(0.231)*	0.940	*(0.387)*	2.265	*(0.319)*	0.781	*(0.136)*	1.770	*(0.203)*
Empl. in industry	-0.063	*(0.008)*	-0.018	*(0.007)*	0.145	*(0.060)*	0.256	*(0.036)*	0.193	*(0.034)*	0.038	*(0.004)*
Inward orientation	-0.264	*(0.032)*	0.120	*(0.044)*	0.076	*(0.031)*	0.363	*(0.051)*	0.096	*(0.017)*	0.202	*(0.023)*
Mean of interaction effect												
Product term induced	-0.009	*(0.001)*	-0.002	*(0.001)*	0.000	*(0.000)*	-0.003	*(0.001)*	-0.002	*(0.000)*	-0.008	*(0.001)*
Model inherent	0.009	*(0.001)*	0.001	*(0.000)*	0.000	*(0.000)*	0.003	*(0.000)*	0.002	*(0.000)*	0.007	*(0.001)*

Notes: Standard error (S.E.) in brackets; marginal effect calculation according to Eq. (3.4) and interaction effect calculation according to Eq. (3.5), both based on a negative binomial model specification including spatial filters (Eq. 3.3); model inherent interaction results from the interaction of all covariates produced by the nonlinear negative binomial specification. The dependent variable is the number of regional patents in the respective field and expected count values are normalised by the maximum in each field to enable comparisons.

When we take a closer look at the spatial distribution of the individual marginal effects of network centrality for each region, we observe an unequal spatial pattern in all KET fields (Figure 3.1). The maps confirm that the network effects are higher in the traditional industrial core regions in Western Europe than in the peripheral regions in Southern, Central and Eastern Europe. Also, in the spatial distribution of network effects, differences between KET fields are noticeable: High network effects are more concentrated in the engineering-based fields of *Microelectronics*, *Photonics* or *Advanced Manufacturing Technologies*, while they are more equally distributed across regions in the science-based sectors, in particular *Nanotechnology*.

Figure 3.1. *Spatial distribution of individual marginal effects of network centrality*



Notes: Classification based on Jenks natural breaks. The spatial concentration of the effect of network centrality is confirmed by high Gini coefficients amounting to 0.86 for Microelectronics and 0.77 for Photonics, compared to 0.66 for Industrial biotechnology and 0.62 for Nanotechnology. The Moran's I is significantly positive for all fields except Nanotechnology.

To get insights into the interrelation between R&D network centrality and region-internal human resources, we indicate the interaction effects calculated for our different KET models at the bottom of Table 3.2. Based on the product term induced interaction effects, we can conclude the direction and significance of the conditional dependence between region-external networks and region-internal endowments; the model inherent interaction effects instead are the result of the nonlinear NegBin specification in Equation (3.5). Except for

Photonics, we observe small but significantly negative interaction effects for all models confirming the assumption that the availability of own-region endowments reduces the benefits of inter-regional networking or a substitution effect between internal resources and external networks (see Wanzenböck and Piribauer 2018)[26]. Generally speaking, the highest, or less negative, interaction effects can be found in Southern, Central and Eastern European regions (Figure A3.2 in the appendix of this section). Despite the higher network centrality effects in terms of magnitude in the Western European 'core', we see that the geographically more peripheral regions can generate *relatively* higher knowledge generation benefit from EU network participation. Concerning KET field-specific differences, interesting is that we found the strongest interrelation with region-internal resources in *Nanotechnology* and *Industrial Biotechnology*. The fact that know-how is more generic or codified, thus less bound to on-site industrial production, as well as the operation of large research institutes and companies in these fields could be potential explanations for the observed heterogeneities.

Regarding the other region-internal factors, the following findings are particularly interesting: In the field of *Nanotechnology*, both a stronger industrial sector (measured by high industrial employment) and dense region-internal networks (measured by the number of intra-regional linkages) seem to negatively affect the inventive activity, while for all other fields these factors contribute positively to knowledge output. It further seems that region-internal and external networks are not equally important for knowledge creation in the different KET fields. Both network-related variables seem to have a comparatively low effect in the *Advanced Manufacturing Technologies*, while the effects of intra-regional networks are exceptionally high in the field of *Advanced Materials* and *Industrial Biotechnology*. Finally, all effect estimates – except for network centrality – show the highest significance in the field of *Advanced Materials*. This finding underlines the importance of the regional context in this field. Knowledge production in the material sector seems to be more closely related to the existing manufacturing sector in a region and driven by region-internal knowledge sourcing. The dominance of industry or application-oriented

[26] An alternative modelling approach to test the moderating effect of internal endowments would be to interact the network centrality with regional R&D expenditures, our proxy for regional R&D efforts and financial resources of firms. We tested this alternative specification for robustness and achieved similar results for all KET fields, confirming the presence of a substitution effect between region-internal resources and region-external knowledge sources.

knowledge generation processes in this field might explain why region-internal factors, in particular financial inputs of the business sector, seem to be more important for inventions in material research than in other fields.

3.7 Conclusions

This paper investigates the role of regional embeddedness in EU funded R&D networks to develop KETs in European NUTS 2 regions. With the notion of KETs, we bring technologies into focus that are a major building block of industrial and innovation policy strategies at the EU level as well as in countries or regions. Given the horizontal and systemic nature of KETs, developing capabilities is considered crucial for creating new innovation and growths paths of entire regions or countries, with high possibilities for cross-sectoral or cross-regional spillovers (Montresor and Quatraro 2017, Evangelista et al. 2018). Local and global knowledge networks might play a pivotal role in the generation of KETs, notably, as they are clearly recognised as a major vehicle of knowledge spillovers (see, e.g. Breschi and Lissoni 2001, Owen-Smith and Powell 2004). However, it is not clear which role the regional network conditions play for the heterogeneous technologies and whether high inter-regional interconnectivity can stimulate regional capabilities in such key fields.

The aim of our study was twofold: First, contributing to the scarce regional literature on KETs by providing new evidence on the regional determinants of the development of KET capabilities, in particular, the effects of the EU network embeddedness for the different KET fields; second, systematically comparing the impact of regional network embeddedness in light of the different knowledge bases, dominant modes of knowledge production or spatial network structures characterising the technological fields. Our empirical model is based on the assumption that the significance of network effects is interrelated with the regional resources and skill endowments. We relied on an augmented regional knowledge production function (KPF) as in Wanzenböck and Piribauer (2018) to account for such interaction effects and introduced marginal effect interpretations applicable for non-linear model specifications. We estimated a spatially filtered negative binomial regression model based on which we derive average marginal effects from quantifying and comparing the impacts of R&D network embeddedness across technological fields.

Our technology-specific analysis confirms the results found by Wanzenböck and Piribauer (2018), supporting the assumptions of a generally positive role of network embeddedness for domestic knowledge generation, on the one hand, and the *relatively* decreasing importance of inter-regional networks for regions with high own endowments, on the other hand. Building on these two relationships, the study delivers novel and relevant insights regarding technology-specific aspects:

First, the positive role of regional network embeddedness clearly differs between the individual fields, being particularly significant for knowledge generation in science-based technological fields. This finding is well in line with previous observations that both *Nanotechnology* and *Industrial Biotechnology* draw heavily on scientific inputs, often organised in the form of collaborations organised at an inter-regional scale (Owen-Smith and Powell 2004, Bozeman et al. 2007, Ter Wal 2013, Heimeriks and Boschma 2014). In contrast, the influence of network embeddedness is lower in fields linked more closely to industrial and on-site production processes, where knowledge generation processes are typically more informal and engineering-based (EC 2015a). Network linkages, both region-internally and region-externally, seem to be of low significance for the development of *Advanced Manufacturing Technologies*.

Second, the interdependence between region-external networks and region-internal resources seems stronger for knowledge creation in *Nanotechnology* and *Industrial Biotechnology*. This finding suggests that knowledge sourcing via inter-regional networks can, particularly in the science-related fields, act as a substitute for lower levels of their own regional skills. Given the codified, more explicit, nature of the knowledge base in these fields, close network linkages to other regions may help lagging regions to develop technological capabilities in these fields. In contrast, the development of new technologies in application-oriented fields, particularly in the case of *Advanced Materials*, seems to be mainly driven by the region-internal knowledge production conditions.

Third, from a regional perspective, interesting is the finding that we observe noticeable differences in the spatial distribution of the regional network effects. While network effects are more spatially concentrated in the engineering-based fields of *Microelectronics*, *Photonics* or *Advanced Manufacturing Technologies*, the benefits of inter-regional network linkages seem to be more equally distributed across regions in the science-based sectors. However, in any case, the effects of EU funded R&D networks are higher in the 'industrial core' regions of Western Europe. It, therefore, seems that EU funded projects reinforce a highly unequal regional distribution of KET capabilities.

Some limitations and, accordingly, ideas for future research come to mind. By limiting to the case of EU funded R&D networks, we are aware that we analyse a specific, policy-driven type of knowledge networks, which limits the generalisability and transferability of results. Moreover, the classification of KETs is not untainted by problems typically arising from the application of broad typologies, such as a wide in-class variety. At the same time, however, this study is, to the best of our knowledge, the first pan-European study that systematically considers technological heterogeneities in cross-regional network structures. A valuable extension would be to further account for organisational or institutional differences or the location of critical organisations, such as universities or important research organisations, in the cross-regional networks.

Moreover, the evolution of policy-induced networks over time or their effects on innovation and the successful development of new products and processes at the regional level would be a crucial point for further analyses. This also relates to questions regarding the importance of generic or key technologies to strengthen the comparative advantages of regions, such as with regional smart specialisation strategies (Foray et al. 2009, Montresor and Quatraro 2017). In this regard, our study clearly points to technology-specific pathways which are idiosyncratic with different regional drivers.

3.8 Appendix to Article II

Observing Key Enabling Technologies

Table A3.1. *List of Keywords for retrieving KET-specific networks*

KET	Keywords
Nanotechnology	nanotechnology, nanoelectronics, nanomaterials, nanoanalytics, nanotools, nanoinstruments, nanomeasuring, nanooptics, nanomagnetics, nanostructures
Microelectronics	semiconductors, microelectronics, nanoelectronics
Photonics	solar,"lighting*" Und Wie "*photonic","laser*" Und Wie "*photonic", optical, sensor,"lens*" Und Wie "*photonic"
Advanced Materials	advanced metal, advanced polymer, advanced ceramic, superconductor, composite, biomaterial, advanced material, smart material
Advanced Manufacturing Technologies	robotics, industrial process, machine tools, computer-integrated, automation, computer integrated, transportation technology, logistic technology, computing technology, measuring technology, measurement technology, manufacturing technology
Industrial Biotechnology	enzyme, fermentation, biochemical, biomaterial, "biotechnolog*" Und Wie "industrial*"

* all FP7 projects not in: ERC, SSH, SME, PEOPLE. REGION, INFRASTRUCTURE, SIS, REGPOT

Table A3.2. *List of IPC classes of KET fields (based on Aschhoff et al. 2010)*

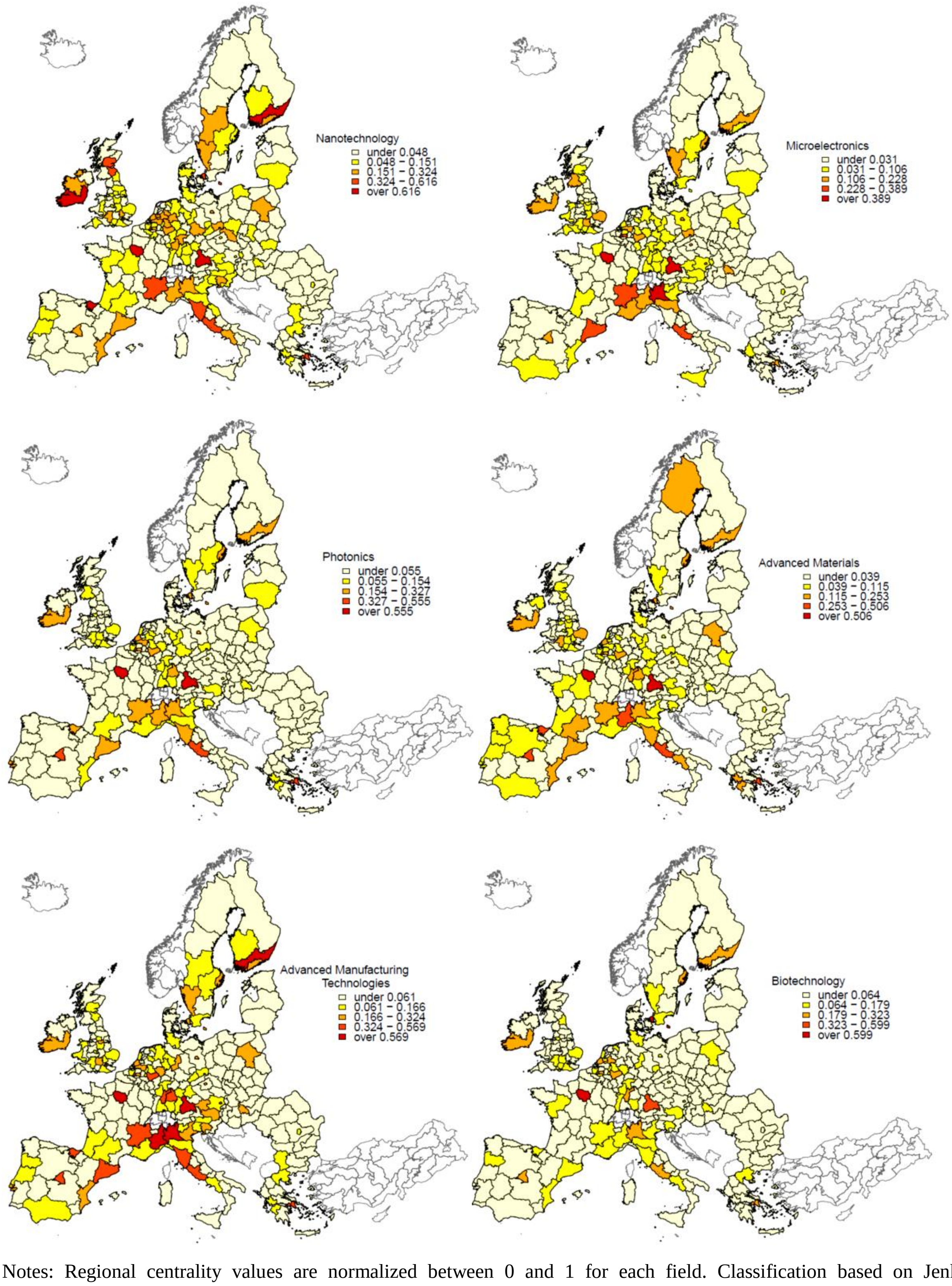
KET	IPC
Nanotechnology	Y01N, B82B
Microelectronics	H01H 57/7, H01L, H05K 1, H05K 3, H03B 5/32, Y01N 12
Photonics	F21K, F21V, G02B 1, G02B 5, G02B 6, G02B 13/14, H01L 25/00, H01L 31, H01L 51/50, H01L 33, H01S 3, H01S 4, H01S 5, H02N 6, H05B 31, H05B 33
Advanced Materials	B32B 9, B32B 15, B32B 17, B32B 18, B32B 19, B32B 25, B32B 27, C01B 31, C04B 35, C08F, C08J 5, C08L, C22C, D21H 17, H01B 3, H01F 1, H01F 1/12, H01F 1/34, H01F 1/44, Y01N
Advanced Manufacturing Technologies	B03C, B06B 1/6, B06B 3/00, B07C, B23H, B23K, B23P, B23Q, B25J, G01D, G01F, G01H, G01L, G01M, G01P, G01Q, G05B ,G05D, G05F, G05G, G06M, G07C, G08C, co-occurrence of G06 and any of A21C, A22B, A22C, A23N, A24C, A41H, A42C, A43D, B01F, B02B, B02C, B03B, B03D, B05C, B05D, B07B, B08B, B21B, B21D, B21F, B21H, B21J, B22C ,B23B, B23C, B23D, B23G, B24B, B24C, B25D, B26D, B26F, B27B, B27C ,B27F, B27J, B28D, B30B, B31B, B31C, B31D, B31F, B41B, B41C, B41D, B41F, B41G, B41L, B41N, B42B, B42C, B44B, B65B, B65C, B65H, B67B, B67C, B68F, C13C, C13D, C13G, C13H, C14B, C23C, D01B, D01D, D01G,D01H, D02G, D02H, D02J, D03C, D03D, D03J, D04B, D04C, D05B, D05C, D06B, D06G, D06H, D21B, D21D, D21F, D21G, E01C, E02D, E02F, E21B ,E21C, E21D, E21F, F04F, F16N, F26B, G01K, H05H
Industrial Biotechnology	C02F 3/34, C07C 29/00, C07D 475/00, C07K 2/00, C08B 3/00, C08B 7/00, C08H 1/00, C08L 89/00, C09D 11/04, C09D 189/00, C09J 189/00, C12M, C12N, C12P, C12Q, 12S, G01N 27/327 (excl.co-occurrence with A01), A61, C12N, C12P C12Q

Descriptive statistics for the variables

Table A3.3. *Description of variables*

Variable	Definition	Source
Regional knowledge creation in a specific KET field (dependent variable)	Number of patent applications filed to the EPO in a KET field according to IPC codes (Aschhoff et al. 2010); full counting based on inventor address, mean 2009-2013	OECD REGPAT database
Network centrality	Sum of the degree centralities of organisations located in a region; calculated for the six KET-specific networks based on projects funded by the 7th Framework Programme (FP7), 2007-2013	EUPRO database
Inward orientation	Number of intra-regional network linkages; in % of the total number of linkages. Calculated for the KET-specific networks based on projects funded by the 7th Framework Programme (FP7), 2007-2013	EUPRO database
Human Resources	Regional population (25–64y) with tertiary education (ISCED 5—8) or employed in an S&T occupation, in % of the total regional population, mean, 2007–2013	Eurostat regional statistics
Business RD expenditures	Intramural R&D expenditures of the business enterprise sector (BES) in % of GRP, mean 2007-2013	Eurostat regional statistics
Employment in industry	Regional employment in the manufacturing sector, in % of total regional employment, mean 2008-2013	Eurostat regional statistics

Figure A3.1. *Spatial distribution of degree centrality in each KET field*



Notes: Regional centrality values are normalized between 0 and 1 for each field. Classification based on Jenks natural breaks

Regression results

Table A3.4. *Regression results for negative binomial models with interaction effect*

	Nanotechnology			*Microelectronics*			*Photonics*			*Advanced Materials*			*AMT*			*Industrial Biotechnology*		
	coeff.	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**	
Centrality	0.025	*0.006*	***	0.016	*0.005*	***	0.002	*0.001*	***	0.007	*0.002*	***	0.008	*0.002*	***	0.016	*0.002*	***
HR	0.072	*0.023*	***	0.046	*0.017*	***	0.069	*0.016*	***	0.078	*0.015*	***	0.086	*0.013*	***	0.088	*0.013*	***
Network x HR	0.000	*0.000*	***	0.000	*0.000*	**	0.000	*0.000*	**	0.000	*0.000*	***	0.000	*0.000*	***	0.000	*0.000*	***
RD exp.	0.242	*0.132*	*	0.543	*0.112*	***	0.287	*0.104*	***	0.409	*0.104*	***	0.251	*0.088*	***	0.347	*0.086*	***
Empl. in industry	-0.013	*0.025*		-0.016	*0.019*		0.044	*0.016*	***	0.046	*0.015*	***	0.062	*0.013*	***	0.007	*0.014*	
Inward links	-0.054	*0.041*		0.105	*0.030*	***	0.023	*0.018*	***	0.066	*0.029*	**	0.031	*0.022*		0.040	*0.020*	**
Spatial filters	yes			yes			yes			yes			yes			yes		
Dispersion	0.740	*0.113*	***	0.623	*0.072*	***	0.536	*0.054*	***	0.508	*0.050*	***	0.789	*0.117*	***	0.413	*0.042*	***
Model fit																		
Log likelihood	-521.788			-825.953			-1030.512			-1085.359			-1283.259			-526.140		
Mc Fadden's R2	0.182			0.203			0.175			0.172			0.175			0.175		
AIC*n	1175.575			1783.905			2193.024			2302.719			2698.519			2419.330		
BIC	122.776			-64.658			-82.718			-96.768			-138.287			-100.166		

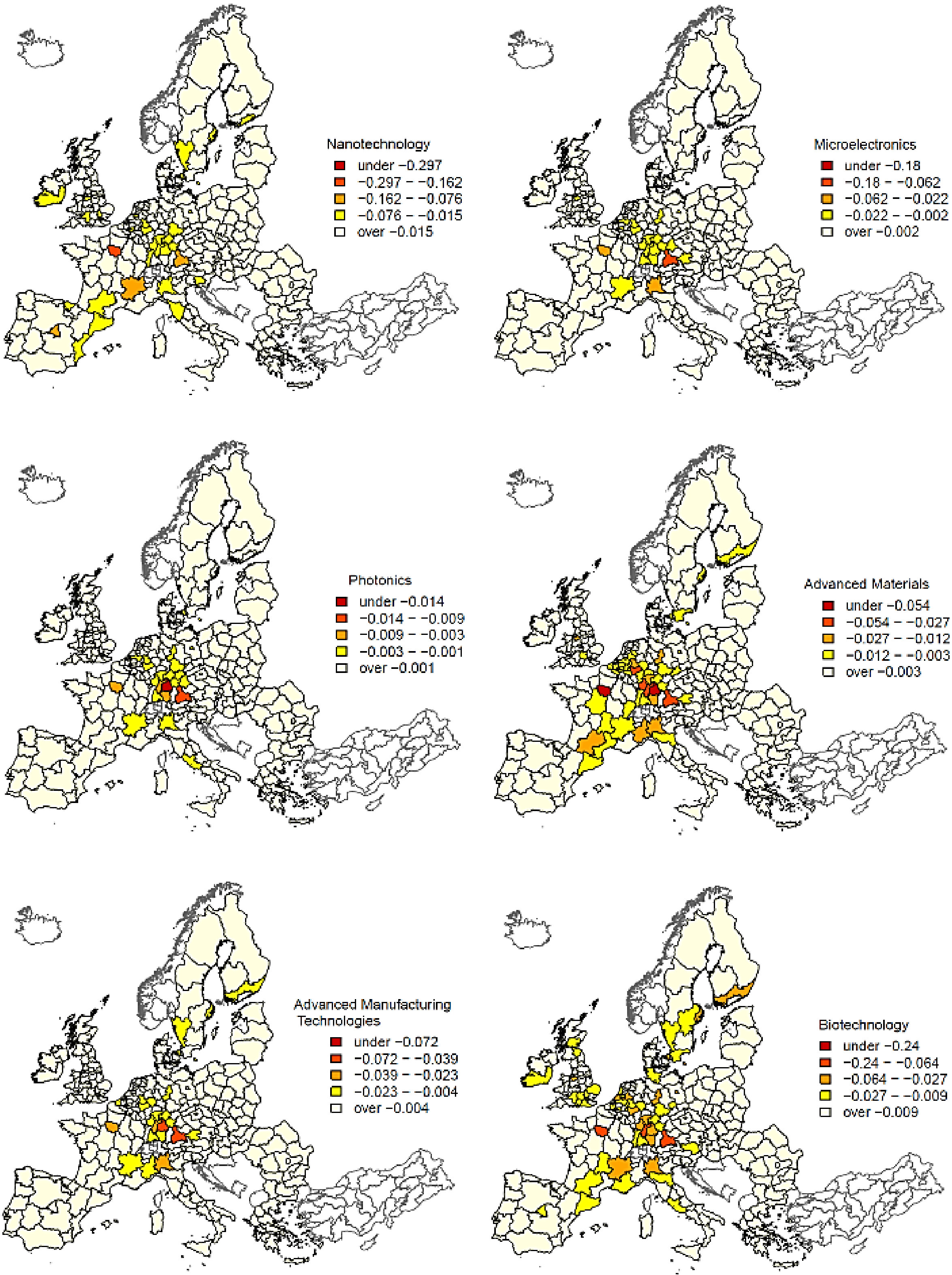
Notes: negative binomial regression incl. spatial filters; dependent variable is the number of regional patents; S.E. = Standard error; AMT = Advanced Manufacturing Technologies

Table A3.5. *Regression results for negative binomial models without interaction effect*

	Nanotechnology			*Microelectronics*			*Photonics*			*Advanced Materials*			*AMT*			*Industrial Biotechnology*		
	coeff.	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**		**coeff.**	**S.E.**	
Centrality	0.010	*0.002*	***	0.006	*0.002*	***	0.001	*0.000*	***	0.002	*0.001*	***	0.004	*0.001*	***	0.003	*0.001*	***
HR	0.049	*0.022*	**	0.036	*0.017*	**	0.056	*0.015*	**	0.061	*0.014*	***	0.074	*0.012*	***	0.071	*0.014*	***
RD exp.	0.332	*0.136*	**	0.570	*0.113*	***	0.326	*0.105*	***	0.439	*0.107*	***	0.268	*0.090*	***	0.436	*0.093*	***
Empl. in ind.	-0.008	*0.025*		-0.012	*0.019*		0.045	*0.016*	***	0.049	*0.016*	***	0.065	*0.013*	***	0.014	*0.015*	
Inward Links	-0.060	*0.041*		0.108	*0.031*	***	0.029	*0.018*		0.060	*0.030*	***	0.035	*0.023*		0.053	*0.022*	**
Spatial filters	yes			yes			yes			yes			yes			yes		
Dispersion	0.789	*0.117*	***	0.637	*0.073*	***	0.548	*0.055*	***	0.529	*0.052*	***	0.435	*0.041*	***	0.469	*0.046*	***
Model fit																		
Log likelihood	-526.140			-828.288			-1033.768			-1090.140			-1287.157			-1158.507		
Mc Fadden's R2	0.175			0.200			0.173			0.169			0.159			0.155		
AIC*n	1182.281			1786.577			2197.535			2310.280			2704.315			2447.015		
BIC	125.932			-65.536			-81.756			-92.756			-136.040			-76.031		
LR Test (models with vs. without interaction)	8.710	*0.003*		4.670	*0.031*		6.510	*0.011*		9.560	*0.002*		7.800	*0.005*		29.680	*0.000*	

Notes: negative binomial regression incl. spatial filters; dependent variable is the number of regional patents; S.E. = Standard error; AMT = Advanced Manufacturing Technologies

Figure A3.2. *Spatial distribution of the region-specific interaction effects (natural breaks)*



Notes: Classification based on Jenks natural breaks.

4 R&D networks and their effects on heterogeneous modes of knowledge creation (Article III)

This section is based on the study *"R&D networks and their effects on heterogeneous modes of knowledge creation: Evidence from a spatial econometric perspective"* (joint work with Thomas Scherngell, submitted)

Abstract: We argue that the effects of R&D networks on regional knowledge creation vary for different modes of knowledge creation – exploitative and explorative – as well as for the quantity and quality of knowledge created. To explore these differences across European regions, we estimate a set of spatial Durbin models (SDMs) with altering network indicators. The results show that EU funded networks are, in general, a significant driver for both modes of knowledge creation. While we find a higher positive impact of networks on explorative than on exploitative knowledge creation for the quantity of knowledge output, the opposite is true for knowledge quality.

Keywords: R&D networks, modes of knowledge creation, exploitation and exploration, spatial Durbin model

4.1 Introduction

The complex nature of the knowledge creation process demands access to diverse knowledge, its adaption, application and diffusion. This spurs the research interest in networks of research and development (R&D), as they serve as channels for transmitting knowledge – also over larger geographical distances (e.g. Autant-Bernard et al. 2007). By this, they are considered to help to overcome geographical barriers for accessing external knowledge (Neuländtner and Scherngell 2020), and to reduce regional disparities in the distribution of knowledge, by moderating in some way the strong geographical localisation of knowledge flows due to the sticky nature of knowledge (Asheim and Isaksen 2002). Especially, in a knowledge-intensive economy, such R&D networks can provide timely access to knowledge and resources that are otherwise unavailable but are necessary to stay competitive and up-to-date in a rapidly changing field (Powell et al. 1996). From a regional perspective, cross-regional collaborations can avoid a possible lock-in to an increasingly obsolete technological trajectory (Cantwell and Iammarino 2005).

While these considerations may be valid on average – based on empirical insights, e.g. from Wanzenböck and Piribauer (2018) – there are at the same time strong arguments that such network effects are not uniform and homogeneous across several dimensions. On the one hand, effects are considered to differ concerning the underlying type of knowledge creation. On the other hand, different network structural characteristics, e.g. a specific centrality in a network, are assumed to produce different effects on knowledge creation. In this study, we specifically argue that various network effects are at stake given the differing nature and extent of knowledge creation processes across regions since they rely on different resources and are shaped and coordinated by the region-specific institutional system of innovation (Tödtling and Trippl 2005).

These arguments are also underlined when taking a network science perspective, stressing that the creation of new knowledge results from collective actions of different actors that are connected by various linkages ranging from informal to formalised network relationships (Acs et al. 2002). Different knowledge bases and technological competences shape these interactions, research and collaboration strategies and rationales, and different spatial scales of knowledge creation (Tödtling et al. 2006). Scholars have conceptualised them as different *modes* or *regimes* of knowledge creation, referring to specific characteristics of the knowledge creation processes (e.g. March 1991, Gibbons et al. 1994, Nonaka 1994, Nowotny et al. 2003, Moodysson et al. 2008). In this study, we are inspired by the conception of March (1991) distinguishing between knowledge *exploitation* and *exploration*, where we understand exploitation-oriented knowledge creation as technological knowledge and application-oriented, mostly occurring in an industrial setting, and specify exploration-oriented knowledge creation as scientific knowledge, mostly occurring in an academic setting.

While there are a number of previous works exploring network effects on regional knowledge creation that acknowledge the critical role of networks in the process of knowledge creation (e.g. Sebestyén and Varga 2013, Wanzenböck and Piribauer 2018, Hazır et al. 2018, Wanzenböck et al. 2020), potential idiosyncrasies of network effects across different modes of knowledge creation are neglected so far. To fill this research gap, this study aims to analyse the effects of R&D networks on regional knowledge creation of different forms, namely knowledge *exploitation* and knowledge *exploration*. Moreover, we consider a distinction in the effect estimates on the *quantity* versus the *quality* of regional knowledge creation. The joint focus on different modes and outputs of knowledge creation at the same time allows for a comparison of network effects across technological and scientific output and the technological importance (quality) of knowledge created. Empirically, the R&D networks are constructed based on inter-regional R&D projects of the EU Framework Programmes (FPs) observed for the European NUTS 2 regions. To proxy exploitative and explorative knowledge creation, systematic and comprehensive information on regional patent applications and scientific publications, respectively, are used.

We employ a set of spatial Durbin models (SDM) in an empirically augmented Knowledge Production Function (KPF) framework to explore the differences between exploitative and explorative knowledge creation across European regions. The models differ with respect to the dependent variable – patent applications or scientific publications. Specifically, to analyse the impact of different network effects, we include a region's (i) degree centrality (the number of partner regions), and (ii) authority score (expresses how well-connected the region is to other regions that are themselves well-connected) as explanatory variables in the basic model. Using spatial econometric techniques allows us to explicitly model the spatial dependence structure of knowledge creation, leading to conclusions of potential agglomeration effects driven by spatial proximity. In doing so, we are able (i) to examine if the effects of R&D networks (degree and authority) vary across different modes of knowledge creation, but also (ii) to analyse the impact of neighbouring regions in terms of their network embeddedness and connectivity on a region's knowledge creation, and (iii) evaluate individual region-specific effects concerning their relative (direct and indirect, i.e. spillover) gains from R&D networks to identify regions with specific development potentials in this respect.

The remainder of the study is organised as follows. Initially, Subsection 4.2 is dedicated to the interrelation of R&D networks and knowledge creation. Subsection 4.3 shifts attention to the heterogeneity of different modes of knowledge creation, before Subsection 4.4 discusses heterogeneity in terms of knowledge outputs. In Subsection 4.5, the conceptualisation of the spatial Durbin model (SDM) is presented, followed by Subsection 4.6, which defines the study's empirical setting. In Subsection 4.7, the estimation results are presented and discussed, before Subsection 4.8 closes with some concluding remarks.

4.2 R&D networks and regional knowledge creation

While the ability of regions to tap into region-external knowledge sources as an important impetus for their knowledge creation capability has been stressed for about two decades (see, e.g. Lagendijk 2001), the crucial importance of cross-region networks to spur the transfer of such external knowledge has gained increasing attention more recently (see Bathelt et al. 2004), in particular in empirical terms (see Scherngell 2019 for an overview). Here, regional knowledge creation is characterised by an interplay between geographically localised knowledge flows within the region and globalised inter-regional knowledge flows most often transferred via R&D collaboration networks (Cooke 2001, Frenken et al. 2007, Asheim et al. 2011). Such R&D collaborations may enable active access to others' specific knowledge or capabilities, which are increasingly located further away in geographical and technological space, as evidenced by several recent empirical works (Scherngell 2013). By this, knowledge created or accessed through inter-regional relations may be incorporated into intra-regional knowledge diffusion mechanisms and regional knowledge bases (e.g. Bathelt et al. 2004). By tapping into region-external knowledge, regional disparities regarding the innovatory potential may be reduced, and possible technological lock-ins could be prevented by an increased technological diversification within the region (see, e.g. Boschma and Ter Wal 2007).

Although empirical studies investigating determinants of regional knowledge creation are manifold (e.g. Autant-Bernard 2001, Acs et al. 2002, Moreno et al. 2005; see Döring and Schnellenbach 2006 for an overview)[27], more recently, the percipience changed towards the increasingly important role of inter-organisational and inter-regional networks, underpinning the flow of knowledge within and across regions as key input of regional knowledge creation and diffusion, and even regional growth processes (Huggins and Thompson 2014)[28]. With the new focus on such networks, empirical studies have started to employ a social network perspective to understand and measure the structure and dynamics of R&D networks – also from a regional perspective (see, e.g. Wanzenböck et al. 2014).

On an organisational level, it is quite well acknowledged that different network structures and topologies – e.g. the position of single nodes and the network structure – greatly impact the creation of new knowledge and its diffusion (e.g. Ahuja 2000, Zaheer and Bell 2005, Giuliani 2007). However, taking a regional perspective, empirical studies on the role of R&D networks on regional knowledge creation are quite scarce; notable exceptions are the studies of Wanzenböck et al. (2020), Hazır et al. (2018), Wanzenböck and Piribauer (2018), Varga and Sebestyén (2017), Sebestyén and Varga (2013), Ponds et al. (2010), and Maggioni et al. (2007). These scholars have followed different modelling strategies – mostly in some kind of Knowledge Production Function (KPF) framework.

Sebestyén and Varga (2013), Varga and Sebestyén (2017), Wanzenböck and Piribauer (2018) and Wanzenböck et al. (2020) include network effects by means of an indicator as an explanatory variable in the model, while the other studies rely on weights matrices to account for spatial and network spillover effects. In particular, Hazır et al. (2018) investigate how R&D networks (co-publications, co-inventions and FP projects) impact regional innovation measured by patent activity. Employing a space-time model, Wanzenböck and Piribauer (2018) study the impact of embeddedness in R&D networks (FP projects) on regional knowledge production, while Wanzenböck et al. (2020) investigate differences in the effects

[27] One main literature stream investigates the impact of spatial proximity on regional innovation performance and economic performance, highlighting the geographically localised nature of knowledge creation and knowledge flows (e.g. Jaffe et al. 1993, Audretsch and Feldman 1996, Anselin et al. 1997). Moreover, in this vein, contributions on regional clusters (Porter 1998, Malmberg 2003, Wolfe and Gertler 2004), industrial districts (Markusen 1996, Harrison 2007), learning regions (Asheim 1996, Morgan 2007), and regional innovation systems (Cooke 2001, Doloreux and Parto 2005) are based on the idea of the facilitative role of spatial proximity for knowledge creation.

[28] Relating to the proximity debate (Boschma 2005), this shifts attention to relational proximity, that – implicitly – encapsulates a number of other types of proximities, such as institutional proximity.

of network centrality on knowledge creation in selected technologies. Sebestyén and Varga (2013) and Varga and Sebestyén (2017) study the effect of the quality (i.e. ego network quality comprising knowledge potential, local connectivity, and global embeddedness) of interregional R&D networks (FP projects) on research productivity, measured in terms of patent applications and scientific publications applying a spatial econometric approach[29].

In general, these studies find positive effects of R&D networks – be it embeddedness, network quality, or spillovers in co-publication and co-patent networks – on knowledge creation, though at the same time confirming the still important role of spatial proximity. However, all of them provide average estimates on network effects in neglecting potential differences across different modes and outputs of knowledge, as reflected in more detail in the following section.

4.3 Heterogeneity in modes of knowledge creation

While the importance of networks for regional knowledge creation has been stressed in the literature, recently also empirically, a finer-grained conceptual and empirical differentiation of these networks effects is relatively unexplored. However, such differentiation may be highly relevant, considering arguments from innovation research, on the one hand, and network science, on the other hand. For instance, it is agreed upon that collaborative knowledge creation is a non-linear and heterogeneous process, characterised by the complex interplay between actors and the knowledge that is created, transmitted and absorbed (Moodysson et al. 2008). Over the last decades, a vast amount of literature emerged (e.g. March 1991, Gibbons et al. 1994, Nonaka 1994, Nowotny et al. 2003, Asheim and Coenen 2005, Moodysson et al. 2008), providing frameworks to categorise the dimensions of the heterogeneity of the knowledge creation process, conceptualising different *modes* or *regimes* of knowledge creation.

In this study, we draw our particular attention to the notions of *exploration* of new possibilities and *exploitation* of old certainties, following the conception put forward by March (1991) and Levinthal and March (1993). While explorative knowledge creation refers

[29] Further studies in this vein are Ponds et al. (2010) and Maggioni et al. (2007). For the Netherlands, Ponds et al. (2010) estimate within a KPF framework the effect of university-industry collaboration (co-publications) on regional innovation as measured by patent applications; they find evidence for both, geographical and network proximity being at work. And finally, Maggioni et al. (2007) apply two separate models to assess the impact of knowledge flows from network partners (research collaborations within the 5th FP) and spatial neighbours on patent activity.

to the discovering of new knowledge dimensions by shifting away from existing rules, norms, routines and activities, exploitative knowledge creation is characterised as a process of routinisation, which adds to the existing knowledge base of industries without changing the nature of activities (March 1991, Gilsing et al. 2008). Hence, exploration broadens the existing knowledge base, whereas exploitation tends to deepen an organisation's or sector's core knowledge base (Guan and Liu 2016). Since the seminal work of March (1991), the idea of exploitative and explorative organisational learning has been adopted to many fields of applications, such as international learning and collaboration, knowledge management, and technology and innovation (see Wilden et al. 2018 for a review). Thus, over the last years, many different definitions and interpretations of the notion of exploitation and exploration have emerged (see Li et al. 2008 for a detailed overview), mostly due to different levels of analyses (e.g. individual level, firm-level, industry-level)[30].

Including a spatial perspective, Rosenkopf and Nerkar (2001) specify exploitation and exploration with respect to the combination of search across spatial and technological boundaries. Phene et al. (2006) define exploitation and exploration as the combination of local or distant search in technical knowledge, including a spatial dimension. This stresses the important role of collaboration networks for exploitative and explorative knowledge creation and innovation. On the one hand, engaging in R&D networks enables organisations to explore new, external knowledge; on the other hand, internal capabilities to understand and utilise such knowledge are essential to efficiently exploit own research potentials (Mowery et al. 1996, Powell et al. 1996).

Based on these existing definitions of exploitation and exploration and the theoretical concepts spanning the dimensions of the heterogeneity of knowledge in the previous

[30] Originally, March (1991) states that exploitation 'includes such things as refinement, choice, production, efficiency, selection, implementation, execution' (March 1991, p.71), and exploration 'includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation' (March 1991, p.71). In the context of innovation research, exploitation is described to be 'search market knowledge', 'technology search', and 'product development projects', whereas, exploration is interpreted as 'searching and recombining technology and science', 'science search', and 'research projects' (Garcia et al. 2003, Geiger and Makri 2006, Gilsing and Nooteboom 2006; in this order). Hence, the relation between exploration and exploitation can be understood as complementary strategies for creating new knowledge and innovations (Heller-Schuh et al. 2011).

subsection, we motivate our conception of exploitation and exploration underlying this study:

> *Exploitation-oriented knowledge creation:* we understand exploitation-oriented knowledge creation as technologically driven, application-oriented, mostly occurring in an industrial setting, targeted towards product development while applying market knowledge

> *Exploration-oriented knowledge creation*: we specify exploration-oriented knowledge creation as scientifically driven, mostly occurring in an academic setting, focusing on basic research activities and research projects reflected in scientific publications

Here, exploitation and exploration are described as two different forms of knowledge creation, driven by different motivations regarding interactive learning mechanisms and collaboration processes that shape the knowledge creation and innovation process. Evidently, knowledge creation processes feature elements of both modes and hence, explorative and exploitative knowledge creation activities are not unambiguously divisible. Nevertheless, whether exploitation or exploration is dominant depends on the rationale of knowledge creation.

On the one hand, exploitation-oriented knowledge creation is characterised by 'local' search for new information, i.e. an incremental knowledge creation process where research activities relate closely to prior research activities (Fleming and Sorenson 2004). Thus, in the case of exploitation, the focus lies on refining existing knowledge by accessing and creating new knowledge domains that deepen the knowledge base in that specific area (Rowley et al. 2000). In particular, this requires highly skilled research personnel active in mostly relatively narrow disciplinary fields, limiting the choice of potentially suitable collaboration partners. Generally, exploitative knowledge creation is largely based on existing core technologies and combinations thereof, aimed at advancing region-internal technology knowledge components. In their nature, such technologies are generally quite sensitive and crucial to strengthen regions' competitive advantages. In this respect, sharing and co-creating knowledge is expected to be driven by frequent and long-lasting collaboration arrangements (strong ties) that are known to have returns on knowledge creation due to enhanced social cohesion (Rowley et al. 2000, Capaldo 2007, Tiwana 2008)[31].

On the other hand, exploration-oriented knowledge creation – science-oriented driven by basic research – is identified as a non-linear process branching into different research fields striving for more radical new developments. Therefore, exploration demands a broad and, to some extent, highly specialised knowledge base, which – if not available within the region – attaches high importance to R&D networks in their function as channels for knowledge transmission and recombining knowledge components regarding knowledge creation. R&D collaborations enable (relatively fast) access to complementary knowledge leading to possibly more innovative research results. Hence, many and frequently changing inter-regional collaborations would be highly beneficial to develop new region-specific technological capabilities. Nevertheless, the often experimental nature of explorative knowledge creation requires – despite the sheer number of weak ties – strategic partners and trust-based and long-lasting collaborations, especially within an academic setting targeting comprehensive solutions for research endeavours, e.g. within research projects.

In analogy to insights on a firm-level (e.g. Gilsing et al. 2008), we argue that both modes of knowledge creation are essential for regions and can also be in a similar manner distinguished at the regional level of analysis. On the one hand, regions with exploitative knowledge creation may benefit from routinised processes with immediate positive returns. In this respect, a high degree of industry specialisation could be favourable to create economies of scope. However, on the other hand, only focusing on core technologies and capabilities possibly leads to a lock-in situation being unaware of new developments (March 1991). This makes explorative knowledge creation essential for a region to break with these existing patterns and strive to advance the existing knowledge stock.

[31] Nevertheless, the influence of the intensity of interactions is not straight forward and may also lead to lock-in resulting in reduced innovativeness and restricted access to new knowledge (Yli-Renko et al. 2001, Molina-Morales and Martínez-Fernández 2009). However, being well embedded in terms of the number of links enables fast and efficient access to new knowledge that can be immediately applied or adapted to specific requirements. This could also entice an argument towards the promoting effect of multiple network links to keep up with the state of the art and adapt rapidly to e.g. changes in demand, especially in an especially in a technology-driven and competitive environment close to market needs.

4.4 Heterogeneity in outputs of knowledge creation

While the previous section provides some basic arguments why network effects may differ across different modes of knowledge creation, also outputs may be affected by different network structural mechanisms. This study stresses a simple distinction of knowledge outputs, namely, according to its *quantity* and *quality*. These two contrary dimensions are widely discussed in the measurement of knowledge outputs (e.g. in terms of patent applications, scientific publications, research projects). The *quantity* of newly created knowledge is a measure of pure quantitative technological or scientific output (e.g. patent or scientific publication counts), and therefore usually used to understand inventive efforts by firms, universities and research organisations. The *quality* of research output – e.g. determined by citations – is assumed to allow for insights into the technological or sometimes even economic importance and/or value of knowledge created (Mowery and Ziedonis 2002, Acosta et al. 2012).

For a long time, the quantity of knowledge created has been used to measure, e.g. a firm's technological output or a university's scientific output; related to a lack of data and also conceptions for measuring the quality aspect. Studies have strongly shaped this research tradition following the Knowledge Production Function (KPF) framework, applied to empirically model the role of different R&D inputs and their influence on regional knowledge creation and innovation (Griliches 1979). In this respect, a vast amount of literature deals with the quantity perspective – i.e. measuring the knowledge created employing pure patent counts (e.g. Moreno et al. 2005, Rodríguez-Pose and Crescenzi 2008, Paci et al. 2014, Hazır et al. 2018), as well as the count of scientific publications (e.g. Coenen et al. 2004, Schwartz et al. 2012).

More recently, endeavours to capture the *quality* of knowledge outputs have come into play (e.g. in terms of value, impact, generality, originality, among others). For instance, patents and publications are heterogeneous in their value and impact, with many being of low market value, little impact, and/or poor quality (Jaffe and Trajtenberg 2002). In academia, we find at the same time the increasing attention towards quality measures derived from characteristics of scientific publications, in particular, their impact often measured by different kinds of citation measures (see Waltman 2016 for a review). This turn towards the measurement of economic, as well as technological and scientific importance is accelerated by the emergence of indicators dedicated to measuring knowledge quality (e.g. Jaffe and

Trajtenberg 2002, Hagedoorn and Cloodt 2003, Ejermo 2009), as well as the increasing availability of ready-to-use data sets, such as data on citations in scientific publications and patents[32].

We claim that in creating knowledge of high quality in contrast to focusing on quantity, R&D network characteristics are at work in different ways. In general, on an organisational level, a central position of an inventor contributes to patents of higher quality (Beaudry and Schiffauerova 2011); we argue that this also holds true for the case of inventive regions. Knowledge output resulting from integrating knowledge from different regions is recognised to be of higher quality since it arises from diverse sources within heterogeneous knowledge networks (Singh 2008). Hence, we suggest connectivity – as in the number of network links – to be conducive for the quality of knowledge created.

In contrast, a high number of reoccurring collaborations may have a hampering effect on the quality of knowledge output, as shown for persistent pairs of patent inventors by Beaudry and Schiffauerova (2011). However, we argue that such established collaborations are usually long-term relationships that enable relatively fast and efficient access to new knowledge, accelerating the creation of new knowledge, especially quantity-oriented knowledge output.

4.5 The model

Turning to our empirical focus, i.e. estimating the role of networks for regional knowledge creation, we need to consider the arguments discussed in Subsection 4.3 and Subsection 4.4 in our modelling approach. To analyse differences in networks effects, we employ a spatial econometric perspective. We specify a spatial Durbin model (SDM) that allows us to account for spatial dependence among the regions, i.e. values observed in one region depend on the values of neighbouring regions; this violates the assumption of independent observations in

[32] To measure quality of knowledge output, in particular, citation-based measures are widely used, both for patents (Jaffe et al. 1993, Maurseth and Verspagen 2002, Singh 2005, Morescalchi et al. 2015), and scientific publications (Frenken et al. 2009, Cowan and Zinovyeva 2013, van Raan 2017). Moreover, driven by bibliometric analysis, a huge variety of literature is dedicated to measure the quality (e.g. impact) of publications by means of bibliometric performance indicators (Zitt et al. 2003, Durieux and Gevenois 2010, Tang and Shapira 2012, Heimeriks and Boschma 2014). For patents, apart from citation-based measures of quality, other indicators such as the number claims, size of patent family, and composite indices thereof are widely-spread quality measures (Lanjouw and Schankerman 2004, Nagaoka et al. 2010, Beaudry and Schiffauerova 2011).

a classical regression framework (Fischer and Wang 2011). Specifically, the SDM accounts for the interconnectivity structure among the regions by augmenting the standard linear regression model by a spatially lagged dependent variable, as well as by spatially lagged explanatory variables. By this, spatial dependence is incorporated by applying a spatial lag operator to the dependent and independent variables (i.e. variable values of the neighbouring regions), allowing for the estimation of local and global spatial externalities (see LeSage and Pace 2009).

Accounting for neighbouring observations and respective spatial spillovers reflects – to some extent – spatial proximity and its role for knowledge creation. With our model, we intend to systematically compare the effects of R&D networks on knowledge creation of different forms: (i) exploitation and exploration, as well as (ii) quantity and quality. Hence, we estimate four separate sets of models that differ regarding their dependent variable – each proxying one of the dimensions of interest. To assess the role of R&D networks also from different angles, we use two distinct network indicators (see Subsection 4.6) that enter the models separately to determine their impact unaffected by other network influences. The incorporation of network measures as explanatory variables in a spatial regression framework allows us to assess spatial and relational proximity, reflecting two perspectives of inter-regional knowledge creation. Particularly, it enables us to infer the role of spatial spillovers in terms of network effects.

Thus, we propose an SDM to estimate regional knowledge creation for a multi-regional system with $i = 1, \dots, N$ regions taking the form

$$y_p = \rho W y_p + X\beta + z_l\gamma + WX\theta + \lambda W z_l + \varepsilon \qquad (4.1)$$

$$\text{with} \qquad \varepsilon \sim N(0, \sigma^2 I_N)$$

where y_p is a N-by-1 vector of knowledge output with p indicating the type of knowledge output, W is a N-by-N row-standardised k-nearest neighbours spatial weights matrix ($k = 5$)[33], the scalar ρ is the coefficient associated to the spatial lag of the dependent variable, X

[33] The number of neighbours ($k = 5$) reflects the median value of the region's neighbours in the sample; robustness checks against alternative numbers of neighbours show similar results. We opt for a k-nearest neighbours specification, since we wish to ensure a certain comparability of region-specific spatial spillover effects from neighbouring regions; other specifications, such as inverse distance or distance band would bias the results towards the number of a region's neighbours. For the case of k-nearest neighbours matrix,

is a N-by-m matrix of knowledge creation input variables (with m indicating the number of respective variables) with β being a m-by-1 vector of corresponding coefficients and θ a m-by-1 vector of coefficients associated to their spatial lag, z_l representing a N-by-1 vector of a network variable of interest, where l specifies the type of network indicator, γ the respective scalar parameter associated with the network indicator l and respective scalar λ representing its spatial lag, and finally, ε is a N-by-1vector of *i.i.d.* error terms.

We construct the model using observations on two periods (2013-2015 for the dependent variables and 2007-2009 for the independent variables) with a time lag of three years between the dependent and independent variables to reduce possible endogeneity (Paci et al. 2014). By this, the current network structure cannot determine past knowledge output. However, an endogeneity problem may still exist if the current network structure is correlated with past network structure (see Bellemare et al. 2017 for a discussion)[34]. To control the influence of outlier regions, we augment the model with a dummy variable that indicates the top and lowest 5% regions regarding the respective dependent variable.

Specific to the SDM is the dependence structure between regions and the additional information from neighbouring regions. Hence, a change in a specific region associated with any given explanatory variable affects the region itself (direct impact) and potentially affects all other regions indirectly (indirect impact) through the spatial multiplier effect, making a straightforward interpretation of parameter estimates difficult (Fischer and Wang 2011). Following LeSage and Pace (2009), we define the direct, indirect and total impacts for the SDM as

$$\bar{M}(m)_{direct} = N^{-1}\mathrm{tr}(S_m(W)) \tag{4.2}$$

$$\bar{M}(m)_{total} = N^{-1}\iota_m'(S_m(W))\iota_m \tag{4.3}$$

$$\bar{M}(m)_{indirect} = \bar{M}(m)_{total} - \bar{M}(m)_{direct} \tag{4.4}$$

row-normalisation equals the alternative of maximum-eigenvalue standardisation, since the maximum eigenvalue is equal to the number of neighbours specified.

[34] Moreover, a second source of endogeneity arises by definition of the SDM specification by including a spatially lagged dependent variable. Rearranging the model reassures exogeneity of variables but induces heteroskedasticity having an error term that is no longer randomly distributed (see Fischer and Wang 2011 for details). Hence, the estimation by means of Ordinary Least Squares (OLS) would lead to inconsistent estimation of parameters as well as standard errors, which is why we estimate the parameters of the model by means of Maximum Likelihood (ML) estimation.

$$\text{where} \qquad S_m(W) = (I_N - \rho W)^{-1}(I_N(\beta_m + \gamma) + W(\theta_m + \lambda)). \qquad (4.5)$$

The average direct impact $\bar{M}(m)_{direct}$ given by equation (4.2) is specified as the average changes in the ith observation of the mth explanatory variable on y_i. Specifically, the expression $S_m(W)_{ii}$ representing the diagonal elements of the matrix denotes the region-specific direct effects (i.e. including impacts passing through neighbouring regions and back to the region itself). The average total impact $\bar{M}(m)_{total}$ given by equation (4.3) is composed of (i) an average total impact *to* an observation $N^{-1}\iota'_N c_m$, i.e. sum across the ith row of $S_m(W)$ representing the total impact on individual observation y_i resulting from changing the mth explanatory variable, and (ii) an average impact *from* an observation $N^{-1}r_m\iota_N$, i.e. sum of the jth columns of that yields the total impact over all y_i from changing the mth explanatory variable by an amount in the jth observation. Finally, whereas the average indirect impact $\bar{M}(m)_{indirect}$ can be determined by subtracting the average direct from the average total impact, region-specific indirect effects (i.e. spillover effects from neighbouring regions) are represented by the off-diagonal matrix elements of $S_m(W)_{ij}$. To draw inference regarding the statistical significance of the impact measures, we follow LeSage and Pace (2009) and simulate the distribution of the effects using the variance-covariance matrix implied by the Maximum Likelihood (ML) estimates.

4.6 Data and variables

Implementing the empirical model as described in the previous section, we describe the operationalisation of the variables and the data sources in what follows. The analysis covers 270 NUTS 2 regions of current 27 EU member states[35]and the United Kingdom and Norway.

Dependent variables – proxying exploitation/exploration and quantity/quality

We employ four different dependent variables in relation to our theoretical framework – each representing one dimension of exploitative and explorative knowledge creation, as well as the aspect of the quantity and quality of the knowledge output (see Table 1). All dependent variables are averages over the period of 2013-2015; this is required to reduce the effect of yearly variations (see Table A4.1 in the appendix of this section for descriptive statistics).

[35] We excluded the regions Åland (FI20) and Notio Egeo (EL42) from the sample due to missing data.

Table 4.1. *Dependent variables*

	Exploitation	**Exploration**
Quantity	No. of patent applications	No. of scientific publications
Quality	Patent quality index	Mean Normalised Citation Score (MNCS)

To proxy exploitative and explorative knowledge creation, we follow previous literature and employ information on regional patent applications and scientific publications, respectively. Many scholars employ these indicators to measure *applied* and *basic* research, respectively – as two fundamental counterpoints of knowledge creation. Whereas basic research is generally understood as experimental or theoretical, applied research is primarily directed towards a specific practical aim or objective (OECD 2002). In a similar vein, the distinction between *analytical* and *synthetic* modes of knowledge creation (Asheim and Coenen 2005, Moodysson et al. 2008) reflects the idea of scientific knowledge being dominant versus focusing on application and recombination of existing knowledge. Clearly, the ideas and intentions of these frameworks partly overlap with the processes of exploitative and explorative knowledge creation; with exploitation generally having a commercial application goal as the main motivation (therefore proxied by patent applications), and exploration is often being related to advances the existing scientific body of knowledge (therefore proxied by scientific publications). Although the two modes of knowledge creation are strongly interwoven, either of the two modes of knowledge creation is to a greater or lesser extent dominant in either science-driven or applied research.

To proxy the *quantity* of exploitation- and exploration-oriented knowledge creation, we use the number of patent applications and the number of scientific publications, respectively[36]. To assess the *quality* of exploitative knowledge creation, we apply a composite index measuring patent quality proposed by Squicciarini et al. (2013). The indicator comprises patents forward citations, the number of claims, patent family size, and the patent generality

[36] For both measures, full counting is applied; i.e. if one patent or publication has more than one inventor, or author respectively, it is fully counted for each region (not divided equally among them).

index[37]. It has several benefits: (i) it covers different dimensions of the quality aspect comprising a set of widely-applied measures, (ii) all individual components are normalised according to patent cohorts stratified by year and technological field (iii) it is predefined and publicly available, and (iv) it is tested for robustness over years, technological fields and countries (Squicciarini et al. 2013). To measure the *quality* of explorative knowledge creation, we use the so-called Mean Normalised Citation Score (MNCS), which is specified as the average number of citations per publications normalised by year and field. The MNCS index is one of the most commonly used state-of-the-art scientific impact indicators (Waltman 2016; see Figure A4.1 in appendix of this section for the dependent variables' spatial distribution).

Data on scientific publications, citations thereof as well as the MNCS indicator are derived from the CWTS Publication Database or are provided by the CWTS directly, whereas data on EPO patent applications at the regional level is extracted from the OECD REGPAT database (both datasets can be accessed as cleaned versions via the RISIS infrastructure, rcf.risis2.eu). The composite index measuring patent quality is constructed by Squicciarini et al. (2013); the indicators are calculated on patent applications filed to the EPO (see Squicciarini et al. 2013 for documentation). In total, the variables are derived from 81,662 patents and 840,149 publications (averaged over the years 2013-2015) with at least one inventor/author located in the study area[38].

Explanatory variables – estimating the role of R&D networks

To capture R&D networks, we use collaborative R&D projects of the European Framework Programme (FP). The FP subsumes temporary, successive research funding programmes with changing focus topics initiated by the European Commission to strive for an integrated European Research Area (ERA). Research projects implemented within the FPs have the character of collaborative R&D projects where consortia consisting of various project partners across Europe follow a common research goal. Hence, the specific nature of the

[37] The *number of forward citations* indicates the technological importance of the patent for the development of subsequent technologies, and hence also reflects somehow the economic value of inventions, the *patent family size* measures the value of patents by means of their geographical scope of patent protection, the *number of claims* mirrors the technological breadth of a patent, and the *patent generality index* measures if the considered invention is relevant for a number of later inventions in other than its own technology class.

[38] Due to full counting of patent applications and scientific publications, double counting may occur.

European FPs constitutes a suitable proxy for pan-European R&D collaborations. Data on FP projects are retrieved from the EUPRO database (also available via RISIS, rcf.risis2.eu), which comprises systematic information on EU funded collaborative research projects of FP1-FP7 and H2020. It includes information on participating organisations, such as their name, type, and geographical location (see Heller-Schuh et al. 2019 for details).

To assess the role of R&D networks, we derive from our theoretical discussion two distinct relevant network indicators covering different dimensions of a region's centrality: (i) *degree centrality*, and (ii) *authority score*. All network indicators are computed as three-year averages of the years 2007-2009, based on a N-by-N network adjacency matrix A that contains our $i = 1, \ldots, N$ regions in the rows and columns, and in the cells the respective cross-region FP collaboration intensities (see, e.g. Scherngell and Barber 2009; see Table A4.1 in the appendix of this section for descriptive statistics and Table A4.2 for correlations).

(i) The *degree centrality* (see Wasserman and Faust 1994 for a formal derivation) of a region is defined as its number of collaboration partners in the network, indicating how well a region is connected to other regions. A region with many partner regions is thought to be diverse in terms of R&D collaborations, which allows accessing sources of knowledge spread across broader sets of actors, spurring a region's knowledge creation and innovation performance.

(ii) The *authority score* (Kleinberg 1999) of a network node is a measure of the amount of valuable information that this node holds; it is defined as the principal eigenvector of $t(A) * A$, where A denotes the adjacency matrix of the graph and $t(A)$ its transpose. In the regional context with an undirected network, the authoritative centrality score of a region expresses how well-connected a region is to other regions that are themselves well-connected (i.e. have a high number of links).

In contrast to the simple degree centrality, showing a high authority score implies some kind of 'multiplier effect' to access new sources of knowledge. However, this only holds if the knowledge is actually accessible through the intermediate (first-order) network neighbour, which in that case is in a gatekeeper position. Intermediate entities may also have some sort of knowledge filtering effect, not diffusing all kinds of knowledge elements; this is especially true for tacit knowledge components that require face-to-face interaction and are hence facilitated by direct knowledge interactions. Modelling the two types of network measures as explanatory variables reflects knowledge network relations between regions. This allows us to analyse direct and indirect links since access to new knowledge and the knowledge received from direct versus indirect partners differs considerably. Moreover, similarly to geographical distance, an increasing network distance reduces the knowledge exchange between partners (Neuländtner and Scherngell 2020). Still, having indirect partners (and hence, the embeddedness of the region in the system of linkages) is of crucial importance to get access to potentially new research communities and clusters.

Control variables – knowledge creation input variables

To isolate network effects from other intervening factors, we need to control for general aspects of knowledge creation. Accordingly, we include the usual KPF inspired factors: (i) *R&D intensity* as a measure of total financial resources devoted to R&D in a region, a measure of (ii) *human resources* proxying a region's potential to generate new knowledge, as well integrate external knowledge by means of the highly skilled labour force, (iii) *population* as a measure of agglomeration to control for a region's size, (iv) a measure of *specialisation* that represents a region's industrial structure regarding its diversity or specialisation tendencies, and (v) a measure of *relational capacity*, reflecting a region's success in acquiring international R&D projects presented by the number of FP projects related to the number of organisations involved in FPs. Data for the variables (i)-(iv) are drawn from the Eurostat regional database and are averaged over 2007-2009. Precisely, R&D intensity is specified as a region's R&D expenditures as a percentage share of GRP, human resources are proxied by the share of persons with tertiary education and/or employed in science and technology, the population is defined as the number of inhabitants, and specialisation is specified as the Index of Specialisation[39] based on employment in NACE classes (2-digit level). Data on FP projects are again retrieved from the EUPRO database. Specifically, in constructing variable (v), the stock of FP projects over the years 2007-2009 is divided by the number of universities and industry organisations involved in FPs in these years.

[39] The Index of Specialisation assesses the degree of specialisation of each region (relatively to the other regions); it is defined as $S_i = 1/2 \sum_{k=1}^{m} |y_{ik} - \bar{y}_k|$ where $y_{ik} = x_{ik} / \sum_{k=1}^{m} x_{ik}$ and $\bar{y}_k = \sum_{i=1}^{n} x_{ik} / \sum_{i=1}^{n} \sum_{k=1}^{m} x_{ik}$ and x indicates the number of employees, and i and k refer to the region $i = 1, \dots, n$ and NACE sector $k = 1, \dots, m$ respectively. The index ranges from 0 to 1, where 1 indicates full specialisation and 0 implies diversification.

4.7 Empirical results

Table 4.2 and Table 4.3 present the impact estimates of the estimated SDMs as defined by equations (4.2) - (4.5). Two models are estimated for knowledge exploitation and knowledge exploration, discriminating for both modes between the quantity and quality of the knowledge created, respectively. In the discussion of the results, we shift attention to the comparison of the network effects on *exploitative* versus *explorative* knowledge creation, as well as the *quantity* versus *quality* of knowledge created[40].

Initially, we briefly reflect on some estimates of the control variables which are generally not surprising but bear some interesting side results. In general, we find R&D intensity to significantly positively affect both modes of knowledge creation, while the effects are higher for exploitative knowledge creation. Similarly, we also find human resources to be positive and significantly, except for the quality aspect of explorative knowledge creation, where we find negative direct impacts. This finding may be owed to the choice of the quality indicator since the MNCS index is a size-independent indicator strongly reflecting the existence of star scientists and research specialists, which is not necessarily related to the general presence of highly educated people[41].

Moreover, we find a positive and significant direct impact of population in all models estimated; hereby, population acts as a control variable for size effects. Turning to the role of industrial specialisation on the modes of knowledge creation under investigation, strikingly, only the quality of explorative knowledge creation is significantly determined. This points to positive returns from specialisation stemming from R&D intensive regional clusters, driven by star scientist and research specialists' basic and experimental research activities. Furthermore, the relational capacity, interestingly, reveals two different directions. Whereas exploitative knowledge creation is clearly negatively affected, the effect is positive for explorative knowledge creation. In line with the above arguments, this underlines that

[40] Due to the absence of spatial dependence for exploration, OLS leads to similar results (see Table A4.3 in the appendix of this section; and see Table A4.4 for coefficients of the SDM).

[41] A high share of highly educated people is typically found in capital city regions with extensive research infrastructure, however, this does not necessarily coincide with the presence of highly cited scientists and researchers.

dominantly large organisations drive explorative rather than exploitative knowledge creation with high capacities to engage in collaborative R&D projects.

Table 4.2. *Model estimates on knowledge production (quantity)*

	(1)	(2)	(3)	(4)
	Exploitation	*Exploration*	*Exploitation*	*Exploration*
Direct effects				
Degree (log)	0.684***	0.711**	-	-
Authority (log)	-	-	0.381***	0.458***
R&D intensity	0.281***	0.203*	0.235***	0.140
Human resources	0.035***	0.044**	0.020**	0.029
Population	0.036***	0.031***	0.024***	0.023**
Specialisation	0.860	0.170	0.821	0.208
Relational capacity	-0.015***	0.040***	-0.024***	0.027*
Indirect effects				
Degree (log)	0.817	1.869***	-	-
Authority (log)	-	-	0.290	0.675**
R&D intensity	0.037	-0.677***	0.058	-0.644**
Human resources	0.051***	-0.001	0.044**	-0.011
Population	-0.025	-0.009	-0.027	-0.006**
Specialisation	-1.173	0.125	-1.587	-0.106
Relational capacity	-0.023	-0.056**	-0.031	-0.065**
Total effects				
Degree (log)	1.501***	2.581***	-	-
Authority (log)	-	-	0.671*	1.133***
R&D intensity	0.318	-0.474*	0.293	-0.504*
Human resources	0.086***	0.043*	0.065***	0.018*
Population	0.011	0.023	0.001	0.017
Specialisation	-0.312	0.294	-0.766	0.102
Relational capacity	-0.038	-0.016	-0.054**	-0.038
ρ (spatial parameter) [a]	0.523***	0.007	0.529***	0.001
Moran's I residuals	-0.029	0.013	-0.029	0.016
σ^2	0.350	1.699	0.332	1.676
LR test	-48.81***	-0.01	-50.73***	-0.00

Notes: [a] ρ denotes the coefficient (not impact estimate); dummy variable indicating top and lowest 5% regions w.r.t. resp ective dependent variable included in all models; dependent variables in logged form; population in 100,000; spatial weig hts matrix constructed using k-nearest neighbours with $k = 5$; average direct, indirect and total impacts determined accor ding to equations (4.2) to (4.5); statistical significance if impact measures based on 1,000 simulation runs (see LeSage an d Pace 2009); the number of observations is 270; the likelihood ratio test compares the spatial autoregressive (SAR) mod el (H0) against the SDM; *** indicates significance at the 0.001 level, ** indicates significance at the 0.01 level, and *in dicates significance at the 0.05 level

The role of R&D networks

Turning to the core of our research questions, i.e. the estimates for network effects on different modes of knowledge creation and differing types of outputs, some highly interesting results appear. With respect to exploitation-oriented and exploration-oriented knowledge creation, we find that EU funded networks are, in general, a significant driver for both of these modes of knowledge creation. Hence, in the sense of being closely connected via R&D collaboration links, relational proximity is statistically beneficial for knowledge creation; both, degree centrality and authority score point towards the same direction in terms of significance and also size.

Directly opposing the two types of network link structures, represented by degree centrality and authority score, the difference between direct and indirect collaboration partners has to be made explicit. Whereas degree centrality reflects direct network links, the authority score also includes the impact of indirect links via intermediate regions, potentially acting as 'knowledge filters' or 'knowledge accelerators'. Although both network effects investigated support the generally benefiting role on knowledge creation, comparing the magnitude of the effects identifies the region's centrality in terms of their number of collaboration partners as a dominating effect. This highlights the necessity to integrate knowledge from many diverse regional knowledge bases in view of creating new knowledge of either kind, exploitation- and exploration-oriented. In particular, this becomes eminent for the case of knowledge quantity.

Hence, the relatively high impact of degree centrality on knowledge creation attributes particular importance to direct network linkages as channels to access and create new knowledge and, therefore, underlines the importance of having the possibility to easily branch into different knowledge domains by means of a wide range of inter-regional collaborations. Moreover, the additionally relatively high impact of authority on knowledge creation attributes also particular importance to indirect network linkages as channels to access and create new knowledge. This is especially the case when creating patents of high quality; these research activities particularly seem to rely on access to a diverse and specialised sets of knowledge components via many direct and indirect channels.

To gain more fine-grained insights from a spatial perspective, we shift attention to region-specific total impact effects, especially at the magnitude of the total effects and the actual knowledge outputs of regions. Although we find that the different network effects are highly correlated and hence, are quite similar in terms of their spatial distribution (see Figure A4.2

and Figure A4.3 in the appendix of this section), some interesting insights appear[42]. When comparing the spatial distribution of the effects on knowledge exploration to exploitation, it can be seen that the distribution for exploitation and knowledge quantity differs the most from the other patterns; showing a smaller number of regions that belong to the group benefiting the most from R&D networks, while at the same time also the number of close to zero profiteers is much higher. It seems that for exploiting new knowledge, quite specific and relevant framework conditions of the regional innovation system must be in place, e.g. effective knowledge transmission mechanisms between the science and the industry sector.

Moreover, particular regions may – relatively to the other regions – benefit more from networks due to their neighbour's network centrality and neighbourhood structure, for instance, because they are spatially close to economic hubs such as surrounding regions of Brussels and Rome. However, interestingly, this is not limited to economic hubs in central Europe but is also observed for neighbouring regions of Bratislava and Bucharest. This reveals economic, technological or scientific potentials, especially for less developed regions, and suggests – in line with similar findings by Varga and Sebestyén (2017) – that strengthening cross-regional networking activities via EU funded joint projects could be a promoting initiative to increase regional knowledge creation. This finding is particularly interesting, given the quite large differences in the ranges of the region-specific effect sizes, since these differences across regions possibly determine the speed of a potential catching-up process of strongly benefiting but generally lagging regions.

In terms of the region-specific impact of R&D networks, the UK takes an outstanding position. Strikingly, the UK regions are among the regions that benefit the most from EU funded networks in all models estimated (with about ten regions under the top-20 profiteers in terms of network effects). Apart from Greater London, especially regions in North-Western England (Greater Manchester, Lancashire, Cheshire, and Merseyside) are amongst the regions with the highest benefits. These results point towards a quite pessimistic conclusion regarding the UK's exit from the EU ('Brexit') and its consequences on the UK innovation system.

[42] This is due to the fact that the differences in the effects only occur from differences in the respective coefficients γ and λ as well as the spatial parameter ρ, whereas the neighbourhood structure given by W remains the same (see equation (A.4) for specification of $S_m(W)$).

Table 4.3. *Model estimates on knowledge production (quality)*

	(1)	(2)	(3)	(4)
	Exploitation	*Exploration*	*Exploitation*	*Exploration*
Direct effects				
Degree (log)	0.887***	0.027	-	-
Authority (log)	-	-	0.417***	0.125
R&D intensity	0.298***	0.179	0.263***	0.144*
Human resources	0.033**	-0.032***	0.018	-0.037**
Population	0.036***	0.017***	0.029***	0.013*
Specialisation	0.669	2.635**	0.761	2.519**
Relational capacity	-0.023***	-0.000	-0.031***	-0.005
Indirect effects				
Degree (log)	0.041**	0.345	-	-
Authority (log)	-	-	0.252	0.199
R&D intensity	0.105	-0.216	0.159	-0.263
Human resources	0.040	0.031	0.035*	0.029
Population	0.016	-0.010	-0.020	-0.012
Specialisation	-0.859	-3.402*	-1.390	-3.325*
Relational capacity	-0.023	0.005	-0.002	-0.001
Total effects				
Degree (log)	1.715***	0.371	-	-
Authority (log)	-	-	0.669**	0.324
R&D intensity	0.404*	-0.036	0.422*	-0.119
Human resources	0.073***	-0.001	0.052***	-0.008
Population	0.016	0.007	0.010	0.001
Specialisation	-0.191	-0.768	-0.627	-0.806
Relational capacity	-0.021	0.005	-0.032	-0.006
ρ (spatial parameter) [a]	0.182***	0.146	0.202*	0.140
Moran's I residuals	-0.016	-0.009	-0.018	-0.009
σ^2	0.882	0.976	0.887	0.968
LR test	-4.23*	-2.25	-5.39*	-2.10

Notes: [a] ρ denotes the coefficient (not impact estimate); dummy variable indicating top and lowest 5% regions w.r.t. resp ective dependent variable included in all models; dependent variables in logged form; population in 100,000; spatial weig hts matrix constructed using k-nearest neighbours with $k = 5$; average direct, indirect and total impacts determined accor ding to equations (4.2) to (4.5); statistical significance if impact measures based on 1,000 simulation runs (see LeSage an d Pace 2009); the number of observations is 270; the likelihood ratio test compares the spatial autoregressive (SAR) mod el (H0) against the SDM; *** indicates significance at the 0.001 level, ** indicates significance at the 0.01 level, and *in dicates significance at the 0.05 level

The role of spatial spillovers

The essence of the SDM is, on the one hand, the additional parameter ρ indicating spatial spillovers from neighbouring regions of the dependent variables (i.e. the effect of the neighbouring regions' patent or publication activity), and on the other hand, the spatially lagged explanatory variables indicating spatial spillovers reflecting the indirect impact of neighbouring regions with respect to the knowledge creation input, and network variables in the model. The coefficient indicating the spillover effects stemming from the dependent variables is only significant for exploitative knowledge creation (both for knowledge quantity and quality), which speaks only for the existence of spatial dependence for patent applications, not for scientific publications. Hence, a region's patenting activity is positively determined by its neighbouring regions' patenting behaviour, highlighting the role of spatial proximity; this is also confirmed by previous works (see, e.g. Wanzenböck and Piribauer 2018). However, the average spatial agglomeration effect seems much higher for the quantity of knowledge creation rather than its quality.

Turning to the second source of spatial spillovers in the SDM, the indirect impacts stemming from the explanatory variables, we find evidence of quite considerable positive indirect effects of the included network measures for the quantity aspect of exploration-oriented knowledge creation, as well as a smaller positive effect of degree centrality alone on exploitation in the case of knowledge quality. Thus, especially for the quantity of explorative knowledge creation, positive externalities from being located close to strongly connected regions arise. Again, in the light of catching-up processes of lagging regions, being co-located to regions that are well embedded in collaborative R&D projects may serve as a stepping stone in becoming involved and becoming beneficiaries of inter-regional knowledge networks. Nevertheless, with a generally positive direct impact of R&D networks on knowledge creation, it is still essential for each region to be embedded in such networks; this is especially the case for knowledge quality. In this respect, missing regional spillover of network effects express the importance of being central and authoritative to create trust-based and long-term partner structures to create knowledge of high quality.

Moreover, we find that exploitation-driven knowledge quantity is positively influenced by the level of human resources of neighbouring regions. Apart from this positive indirect effect, we find several negative indirect effects on knowledge exploration: first, for quantity, the R&D intensity and the relational capacity exhibit negative signs, and second, for quality, the specialisation has a negative indirect effect (while the direct effect is strongly positive). These findings seem kind of counterintuitive to the idea of spatial spillovers. However, with respect to the spatial distribution of exploration-oriented knowledge output, this can be explained by the presence of spatial outliers rather than spatial clusters (see Figure A4.1 in the appendix of this section). That is, single regions with high levels of exploration-oriented knowledge creation, high R&D intensity, high relational capacity, as well as a high degree of specialisation are generally surrounded by regions with low levels of the respective explanatory variables, causing a negative coefficient.

4.8 Concluding remarks

Over the past years, the role of R&D networks for regional knowledge creation has been widely discussed in the literature, generally recognising the importance of such networks for creating and accessing new knowledge. The interactions within R&D networks are by no means homogeneous but are shaped by different research and collaboration strategies and rationales, knowledge bases, and spatial scales (Tödtling et al. 2006). These specific characteristics of the knowledge creation process have been categorised by means of different modes of knowledge creation; well-known the notions of *exploitation* and *exploration* put forward by March (1991). Driven by different motivations regarding interactive learning mechanisms and collaboration processes, we see exploitation and exploration as two forms of knowledge creation. Whereas we understand exploitation as technology-driven and applications-oriented knowledge creation, we specify exploration as scientifically based knowledge creation within a mostly academic setting. Further, heterogeneity in the process of knowledge creation occurs from the distinction between the quantity and quality of knowledge created, constituting two different ways of measuring knowledge output: sheer quantity versus quality of knowledge created in terms of its technological or even economic impacts.

In this study, we aim at investigating differences in the role of R&D networks accounting for these heterogeneities in knowledge creation processes and their output. Empirically, we construct the R&D networks based on publicly funded collaborative R&D projects within the EU FPs observed for 270 European NUTS 2 regions. To proxy exploitative and explorative knowledge creation, patent applications and scientific publications, respectively, are used. Precisely, knowledge quantity is measured in terms of the number of either patent application or publications, and the quality perspective is quantified by a composite patent quality index and the MNCS. The models differ with respect to the dependent variable – patent applications or scientific publications. Specifically, to analyse the impact of different network effects derived from a region’s centrality in the cross-region FP collaboration network, we include the (i) degree centrality and (ii) authority score as additional explanatory variables in the basic regression model.

The results of the model estimations show clearly that EU funded networks are a significant driver for both exploitative and explorative knowledge creation. Specifically, the results point to a higher positive impact of networks on explorative than on exploitative knowledge creation for the quantity of knowledge output; the opposite is true for knowledge quality. This indicates higher importance of cross-region network embeddedness and connectivity within a scientific and academic setting in terms of increasing knowledge quantity and emphasises the role of networks in an application and technology-oriented environment for enhancing the quality of knowledge creation. In other words, cross-regional interactions – thus, integrating knowledge from various other regions – are more important to create a high publication output rather than patenting output. Nevertheless, creating knowledge of economic and technological importance (high-quality patents) indeed benefits from a large and diverse set of network partners, even to a more considerable extent than for advancing scientific excellence. Additional positive indirect network effects for exploration- and quantity-driven knowledge creation highlight the importance of being co-located to central and authoritative regions to benefit from respective regional spillovers. Thus, being geographically close to strongly connected regions, may especially accelerate the regions' output in terms of explorative knowledge creation.

Moreover, examining the region-specific effect estimates reveals economic, technological or scientific regional potentials, especially for less developed regions that are amongst the most benefiting regions regarding direct and spillover network effects from neighbouring regions (e.g. surrounding regions of Bratislava and Bucharest). Strikingly, we find a large number of UK regions exhibiting relatively high positive impact from EU funded networks, pointing to a pessimistic finding in the context of 'Brexit' and its consequences for the UK innovation system.

Policy-wise, these results allow deriving a set of conclusions regarding the role of EU funded joint projects as drivers for regional knowledge creation: fostering the access to such R&D networks may be beneficial to (i) accelerate knowledge quantity, especially in terms of the region's explorative knowledge creation, and (ii) enhance a region's exploitation-oriented and quality-driven knowledge output, i.e. represented by the high-quality patents. However, exploration-oriented knowledge quality, high-quality and high-impact scientific publications (as proxied by the MNCS) is not affected by network connectedness and hence requires alternative policy measures, such as providing fruitful research environments to attract star-scientists and high-level research experts. Emphasising the role of spatial spillover effects of R&D networks, funding of EU funded joint projects may facilitate a catch-up process by less developed regions regardless of their initial region-internal knowledge-stock and network capabilities. In particular, the results show strong positive spatial externalities regarding the regions' output in terms of explorative knowledge creation, thus possibly enabling lagging regions to also take part in more explorative and scientific orientated knowledge communities and debates.

Restrictively, the findings in this study rest on the choice of the R&D network, which in this case are publicly funded collaborative projects within the EU FP that follow certain rules and political intentions. Hence, notably, the interpretation of the impacts is limited to this kind of R&D networks; we see this as an entry point for future research establishing generalisability of the results obtained. Moreover, the aspects of knowledge quality could be highlighted in much more detail, considering, e.g. a comparison of multiple measures of knowledge quality. Furthermore, we see envisioning a dynamic perspective on the role of R&D networks as particularly fruitful in enhancing the future scientific discussion on modes of knowledge creation.

4.9 Appendix to Article III

Table A4.1. *Descriptive statistics of regression variables*

	Mean	Standard Deviation	Minimum	25% Percentile	75% Percentile	Maximum
Number of patent applications	302.43	499.84	0.33	36.08	377.92	3,916.33
Number of scientific publications	3,111.66	5,072.75	0.00	461.75	3,968.75	58,363.00
Patent quality index	23.345	39.966	0.000	2.153	29.933	264.739
MNCS index	3.125	2.235	0.000	1.403	4.281	13.024
R&D intensity	1.468	1.188	0.070	0.672	1.900	6.940
Human resources	36.56	8.56	15.90	31.43	41.60	73.97
Population (in 100,000)	18.54	14.91	1.26	9.86	22.85	116.62
Specialisation	0.199	0.081	0.000	0.200	0.246	0.481
Relational capacity	19.62	11.40	0.000	11.23	25.12	62.63
Degree centrality	201.890	64.998	10.667	159.667	254.000	302.667
Authority score	0.095	0.131	0.001	0.016	0.112	1.000

Note: variables as specified in subsection 4.5

Table A4.2. *Correlations of dependent and independent variables*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Patent applic-ations	Scient. public-ations	Patent quality index	MNCS index	R&D intensity	Human resources	Popula-tion	Special-isation	Relat-ional capacity	Degree centrality	Authority score
(1)	1.000	0.602	0.983	0.426	0.505	0.409	0.546	0.066	0.215	0.424	0.616
(2)	0.602	1.000	0.513	0.476	0.330	0.452	0.630	0.135	0.485	0.507	0.852
(3)	0.983	0.513	1.000	0.396	0.472	0.357	0.526	0.053	0.160	0.392	0.539
(4)	0.426	0.476	0.396	1.000	0.234	0.214	0.505	0.068	0.267	0.430	0.398
(5)	0.505	0.330	0.472	0.234	1.000	0.532	0.091	-0.082	0.342	0.528	0.447
(6)	0.409	0.452	0.357	0.214	0.532	1.000	0.055	0.064	0.317	0.454	0.527
(7)	0.546	0.630	0.526	0.505	0.091	0.055	1.000	0.176	0.198	0.484	0.586
(8)	0.066	0.135	0.053	0.068	-0.082	0.064	0.176	1.000	-0.030	0.098	0.111
(9)	0.215	0.485	0.160	0.267	0.342	0.317	0.198	-0.030	1.000	0.570	0.608
(10)	0.424	0.507	0.392	0.430	0.528	0.454	0.484	0.098	0.570	1.000	0.656
(11)	0.616	0.852	0.539	0.398	0.447	0.527	0.586	0.111	0.608	0.656	1.000

Note: variables as specified in subsection 4.5

Table A4.3. *OLS estimates*

	Quantity			
	Exploitation	*Exploration*	*Exploitation*	*Exploration*
Degree (log)	0.706***	0.587*	-	-
	(0.150)	(0.241)	-	-
Authority (log)	-	-	0.363***	0.434***
	-	-	(0.072)	(0.118)
R&D intensity	0.402***	0.186*	0.365***	0.114
	(0.059)	(0.093)	(0.061)	(0.096)
Human resources	0.068***	0.035**	0.056***	0.020
	(0.008)	(0.012)	(0.008)	(0.013)
Population	0.029***	0.039***	0.023***	0.029***
	(0.004)	(0.007)	(0.005)	(0.008)
Specialisation	-0.110	1.185	-0.122	1.204
	(0.691)	(1.115)	(0.687)	(1.096)
Relational capacity	-0.027***	0.038***	-0.035***	0.026*
	(0.005)	(0.009)	(0.006)	(0.010)
Constant	-2.070**	0.890	3.553***	6.377***
	(0.754)	(1.143)	(0.558)	(0.924)
	Quality			
	Exploitation	*Exploration*	*Exploitation*	*Exploration*
Degree (log)	0.966***	0.015	-	-
	(0.209)	(0.177)	-	-
Authority (log)	-	-	0.399***	0.126
	-	-	(0.104)	(0.087)
R&D intensity	0.412***	0.158*	0.395***	0.119
	(0.082)	(0.068)	(0.087)	(0.071)
Human resources	0.072***	-0.016	0.059***	-0.021*
	(0.011)	(0.009)	(0.011)	(0.010)
Population	0.027***	0.012*	0.022**	0.008
	(0.006)	(0.005)	(0.007)	(0.006)
Specialisation	-0.227	1.410	-0.197	1.329
	(1.003)	(0.814)	(1.015)	(0.809)
Relational capacity	-0.041***	0.002	-0.047***	-0.003
	(0.008)	(0.007)	(0.009)	(0.007)
Constant	-5.942***	0.640	1.096	1.548*
	(1.052)	(0.852)	(0.801)	(0.668)

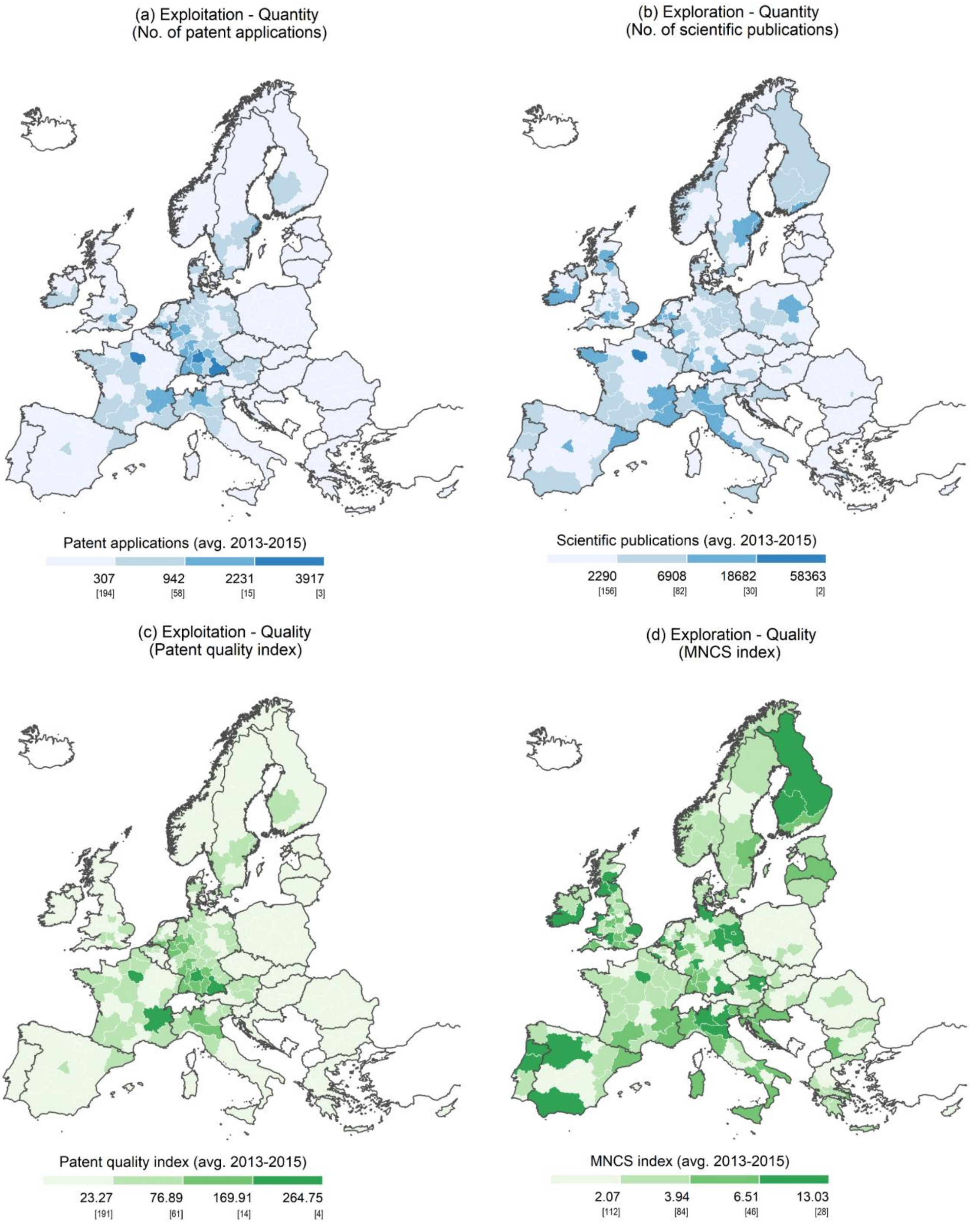
Notes: variables as specified in subsection 4.5; standard OLS model estimated; dummy variable indicating top and lowest 5% regions w.r.t. respective dependent variable included in all models

Table A4.4. *Coefficients of the SDMs*

	Exploitation & Quantity		**Exploration & Quantity**		**Exploitation & Quality**		**Exploration & Quality**	
Human resources	0.032*** (0.007)	0.018* (0.007)	0.044** (0.015)	0.029 (0.016)	0.032** (0.011)	0.017 (0.011)	-0.033** (0.011)	-0.038** (0.012)
R&D intensity	0.279*** (0.043)	0.231*** (0.043)	0.204* (0.095)	0.140 (0.097)	0.296*** (0.067)	0.258*** (0.070)	0.184** (0.071)	0.150* (0.073)
Population	0.037*** (0.003)	0.030*** (0.003)	0.031*** (0.007)	0.023** (0.008)	0.037*** (0.005)	0.030*** (0.006)	0.017*** (0.005)	0.013* (0.006)
Specialisation	0.930 (0.545)	0.916 (0.530)	0.169 (1.211)	0.208 (1.199)	0.691 (0.877)	0.801 (0.878)	2.706** (0.903)	2.586** (0.897)
Relational capacity	-0.014*** (0.004)	-0.022*** (0.004)	0.040*** (0.009)	0.027** (0.010)	-0.023*** (0.006)	-0.031*** (0.070)	-0.001 (0.007)	-0.005 (0.007)
Degree (log)	0.635*** (0.105)	- -	0.709** (0.229)	- -	0.865*** (0.166)	- -	0.019 (0.173)	- -
Authority (log)	- -	0.363*** (0.050)	- -	0.458*** (0.113)	- -	0.410*** (0.082)	- -	0.121 (0.085)
Constant	-1.786 (1.211)	3.227** (1.182)	1.728 (2.613)	6.750 (2.674)	-6.395*** (1.919)	-0.362 (1.944)	1.243 (1.988)	2.121 (2.011)
Lag Human resources	0.010 (0.010)	0.013 (0.010)	-0.001* (0.021)	-0.011 (0.021)	0.029 (0.016)	0.025 (0.015)	0.033* (0.016)	0.031 (0.016)
Lag R&D intensity	-0.127 (0.093)	-0.093 (0.092)	-0.679*** (0.184)	-0.648*** (0.190)	0.035 (0.145)	0.079 (0.147)	-0.216 (0.143)	-0.253 (0.147)
Lag Population	-0.032*** (0.006)	-0.029*** (0.007)	-0.009 (0.016)	-0.006 (0.016)	-0.024* (0.010)	-0.023* (0.011)	-0.011 (0.010)	-0.012 (0.011)
Lag Specialisation	-1.084 (0.883)	-1.284 (0.854)	0.123 (1.886)	-0.107 (1.872)	-0.851 (1.462)	-1.309 (1.453)	-3.381* (1.446)	-3.298* (1.439)
Lag Relational capacity	-0.005 (0.008)	-0.004 (0.010)	-0.056** (0.019)	-0.066** (0.021)	0.006 (0.014)	0.005 (0.016)	0.005 (0.014)	0.000 (0.016)
Lag Degree (log)	0.085 (0.255)	- -	1.864*** (0.521)	- -	0.541 (0.408)	- -	0.300 (0.386)	- -
Lag Authority (log)	- -	-0.046 (0.114)	- -	0.678** (0.247)	- -	0.125 (0.187)	- -	0.159 (0.183)
Lag Constant	-1.252 (1.764)	-0.842 (1.372)	-9.229* (3.665)	4.411 (3.108)	-1.967 (2.831)	2.028 (2.198)	-2.220 (2.789)	0.089 (2.272)
Rho [ρ]	0.523*** (0.064)	0.529*** (0.064)	0.007 (0.099)	0.001 (0.100)	0.182* (0.087)	0.202* (0.086)	0.146 (0.095)	0.140 (0.095)
Observations	270	270	270	270	270	270	270	270
Log Likelihood	-248.52	-241.55	-454.71	-452.82	-366.91	-367.89	-380.35	-379.18
σ^2	0.350	0.332	1.699	1.676	0.882	0.887	0.976	0.968

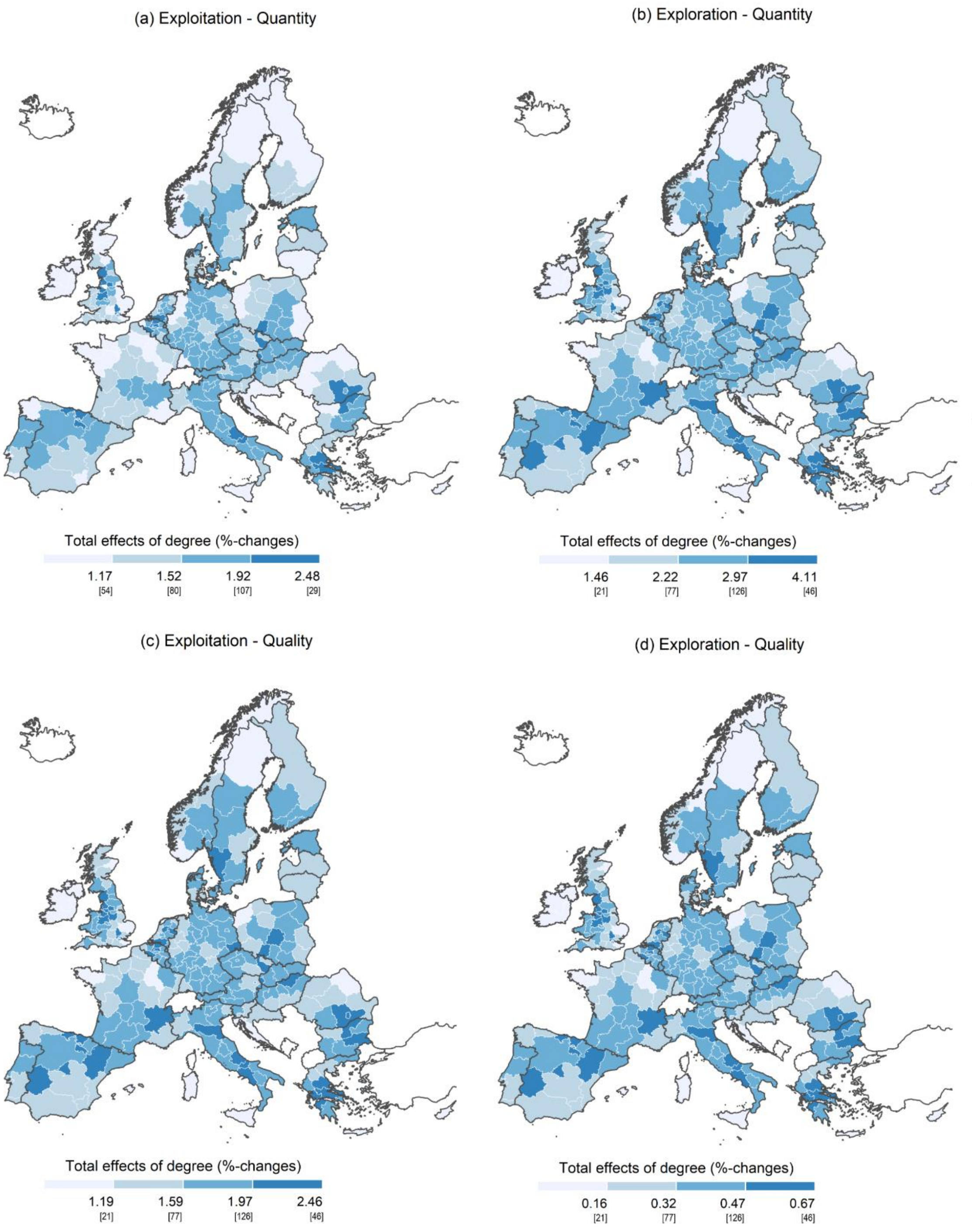
Note: Models as specified in subsection 4.4; dummy variable indicating top and lowest 5% regions w.r.t. respective dependent variable included in all models; dependent variables in logged form; standard errors in parenthesis; *** indicates significance at the 0.001 level, ** indicates significance at the 0.01 level, *indicates significance at the 0.05 level

Figure A4.1. *Spatial distribution of dependent variables*



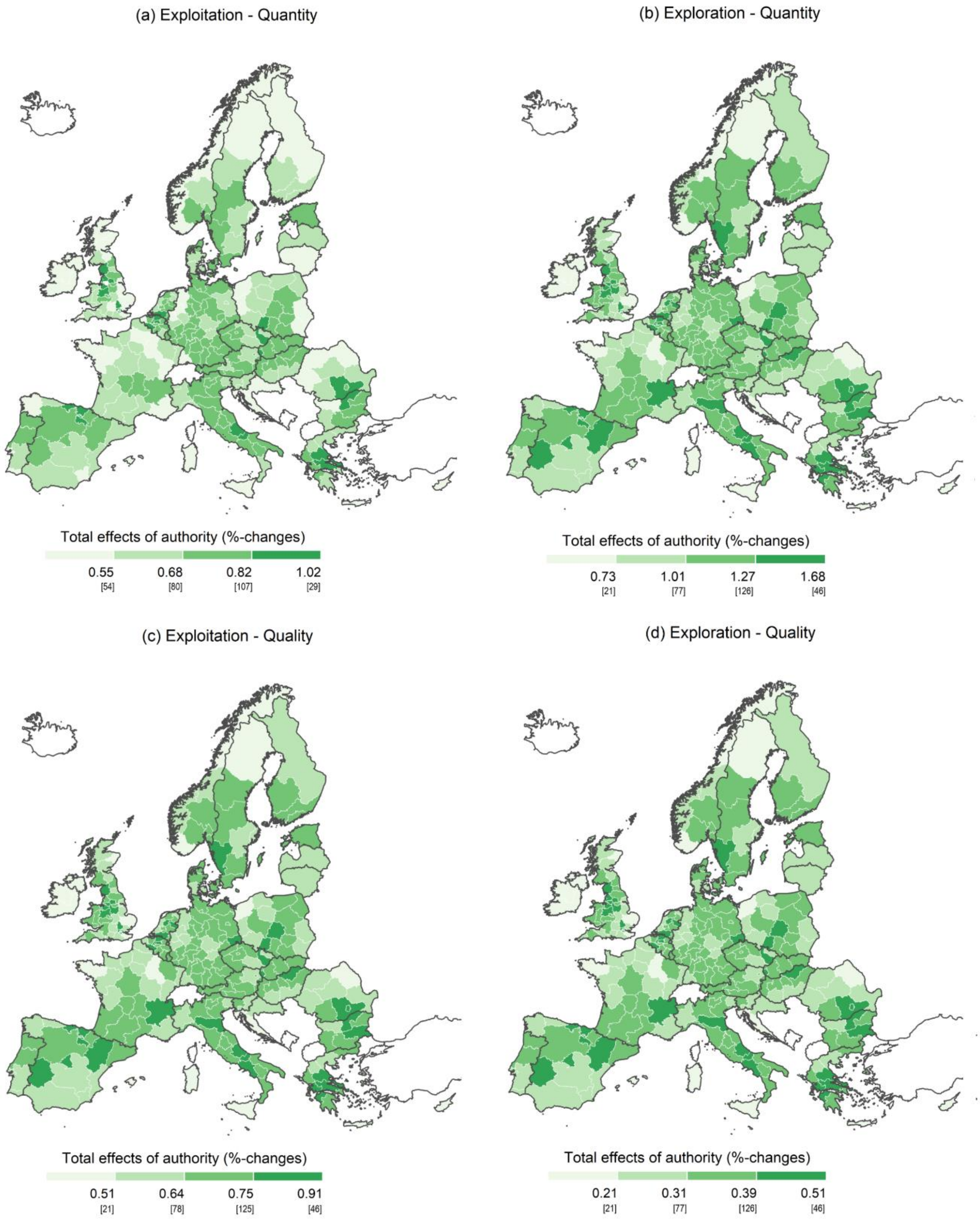
Notes: variables as specified in subsection 4.5; grouping of variables by means of natural breaks

Figure A4.2. *Spatial distribution of region-specific total impact estimates – degree centrality*



Notes: effects calculated as described in subsection 4.5; grouping of variables by means of natural breaks; number of regions per group in parentheses; ρ (spatial parameter) not significant for exploration

Figure A4.3. *Spatial distribution of region-specific total impact estimates – authority score*



Notes: effects calculated as described in subsection 4.5; grouping of variables by means of natural breaks; number of regions per group in parentheses; ρ (spatial parameter) is not significant for exploration

5 An empirical agent-based model for regional knowledge creation in Europe (Article IV)

This section is based on the study *"An empirical agent-based model for regional knowledge creation in Europe"* (published in *ISPRS International Journal of Geo-Information*, 2020, 9(8))

Abstract: Modelling the complex nature of regional knowledge creation is high on the research agenda. It deals with identifying drivers for regional knowledge creation of different kinds, among them inter-regional networks and agglomeration factors, as well as their interplay, i.e. in which way they influence regional knowledge creation and, accordingly, innovation capabilities – in the short- and long-term. Complementing a long line of tradition – establishing a link between regional knowledge input indicators and knowledge output in a regression framework – we propose an empirically founded agent-based simulation model that intends to approximate the complex nature of the multi-regional knowledge creation process for European regions. Specifically, we account for region-internal characteristics and a specific embedding in the system of region-internal and region-external R&D collaboration linkages. With first exemplary applications, we demonstrate the model's potential in terms of its robustness and empirical closeness. The model enables the replication of phenomena and current scientific issues of interest in the field of geography of innovation and hence, shows its potential to advance the scientific debate in this field in the future.

Keywords: regional knowledge creation; geography of innovation; collaboration networks; agent-based modelling; spatial simulation; Europe

5.1 Introduction

Understanding and explaining the complexities of regional knowledge creation constitutes an ongoing challenge for empirical scholars in regional science. Specifically, the literature has long been concerned with the spatial distribution of knowledge creation and innovation, concluding that these kinds of activities are not equally distributed in space but rather tend to be spatially clustered (Malmberg et al. 1996, Audretsch and Feldman 2004). With knowledge being not easily accessible at every point in space, the location of knowledge creation, as well as the processes of knowledge diffusion become a crucial point in understanding regional development and growth (Acs et al. 2002, Tödtling and Trippl 2005). In this respect, attention has been shifted to the investigation and modelling of regional knowledge creation processes as an interplay between (i) geographically localised knowledge interactions within the region, and (ii) the embedding of the region in inter-regional R&D collaborations (see, e.g. Wanzenböck et al. 2014), in particular by means of region-internal and region-external knowledge interactions in the form of R&D collaborations (see Scherngell 2013 for an overview).

Modelling regional knowledge creation follows a long line of research tradition, often applying the Knowledge Production Function (KPF) framework to model determinants of regional knowledge creation and innovation (Fischer and Varga 2003, Rodríguez-Pose and Crescenzi 2008, Neves and Sequeira 2018). These studies typically attempt to establish a direct link between some kind of regional knowledge input, such as industrial and university R&D, and knowledge outputs measured in terms of patents, innovation or publication counts (see, e.g. Jaffe 1989, Fischer and Varga 2003, Marrocu et al. 2013, Paci et al. 2014)[43]. However, all these studies are done at an aggregate regional level and accordingly do not account for the regions' underlying microstructure, for instance, by assuming that each regional organisation benefits in the same way from inter-regional knowledge spillovers. However, a better approximation and understanding of the real-world complexity of regional knowledge creation processes requires models accounting for the heterogeneity of the agent population, the non-linearity of the interactions between agents, and the complexity of the environment. Considering these aspects allows observing emergent phenomena such as

[43] In this context, the role of knowledge spillovers (Ó hUallacháin and Leslie 2007, Rodríguez-Pose and Crescenzi 2008, Ponds et al. 2009), spatial proximity (Breschi and Lissoni 2001, Greunz 2003, Moreno et al. 2005), and nonspatial forms of proximity (Maggioni et al. 2007, Breschi and Lenzi 2012, Miguélez and Moreno 2013) on regional knowledge creation and innovation are widely studied.

specialisation and concentration tendencies in regional knowledge creation driven by the structure of R&D collaborations.

Recent contributions to the discussion on knowledge creation have been made by adding a dynamic perspective using computer simulation techniques, especially agent-based modelling (ABM[44]). However, they are mainly implemented at an abstract, theoretical level: for instance, from a spatial perspective, theoretical contributions by Batty (2012) and Crooks et al. (2008) propose general spatial modelling frameworks and pose key challenges in geo-spatial modelling (see Heppenstall et al. 2011 for an overview), whereas, Ausloos et al. (2015) discuss simulating spatial interactions in ABM. From a conceptional viewpoint, Dawid (2006) and Gilbert et al. (2001) target innovation and technological change and knowledge dynamics in innovation networks, respectively. Moreover, Vermeulen and Pyka (2018) address the spatial distribution of knowledge in the setting of regional innovation policy.

Whereas theoretical models are built as tools for theory-building, very recently, few empirical models of regional knowledge creation aim at analysing real-world scenarios, Wang et al. (2014) use an agent-based model for analysing the diffusion of technologies across Chinese regions, while Beckenbach et al. (2007) present an agent-based simulation of regional innovation dynamics including agents with explicit and implicit knowledge endowment, Paier et al. (2017) focus on the evolution of a single region's technological profile in a context of policy analysis (in the Viennese biotechnology sector). However, so far, the limited simulation studies following this research path are either of purely theoretical and conceptual nature lacking empirical foundation and hence, real-world applicability, e.g. in a (regional) innovation policy context (e.g. März et al. 2006, Vermeulen and Pyka 2018), have only a limited geographical and/or sectoral scope (e.g. Paier et al. 2017, Pyka et al. 2019) neglecting the theoretical fundamentals of regional knowledge creation, and/or deal in a quite narrow way with network formation (e.g. Savin and Egbetokun 2016, Sebestyén and Varga 2019), as well as knowledge transfer and diffusion (Thiriot and Kant 2008, Wang et al. 2014, Mueller et al. 2017).

Hence, we propose an empirically founded agent-based simulation model for regional knowledge creation in Europe. It intends to better approximate the complex nature of the

[44] Abbreviation ABM is used for 'agent-based models' and 'agent-based modelling' consecutively.

multi-faceted regional knowledge creation processes by explicitly accounting for (i) region-internal characteristics, (ii) agent heterogeneity in the knowledge creation process, and (iii) a specific embedding in the system of region-internal and region-external interdependencies in the form of R&D collaboration linkages. By this, we particularly reflect the idea of geographical and relational aspects of the knowledge creation process, which is driven by the debate on ‘local buzz’ and ‘global pipelines’ as two forms of interactive knowledge creation (Bathelt et al. 2004). This allows us to model local dynamics, such as learning and knowledge transfer, and structural evolution in the form of inter-regional network formation and transformation on a global level. The strong theoretical and explicitly empirical foundation enables us to apply the model to real-world contexts, such as simulation experiments referring to Research, Technology, and Innovation (RTI) policy measures at the European, the national as well as regional levels (e.g. smart specialisation, mission-oriented public funding). In this study, we present a comprehensive model overview, providing details on the model elements and processes and technical specifications and robustness checks. The model’s potential is demonstrated by small example applications on currently debated research issues in the geography of innovation literature, namely regional concentration and specialisation patterns, as well as the role of networks as drivers for regional knowledge creation.

The remainder of this study is organised as follows. In Subsection 5.2, we shortly outline the agent-based modelling approach and give a detailed presentation of the proposed simulation model of multi-regional knowledge creation, subsuming a description of the model elements and processes. Subsection 5.3 is dedicated to the model’s empirical foundation, including the agent initialisation, calibration and output evaluation. In Subsection 5.4, we demonstrate the potential of the simulation model by small example applications to current scientific debates. In Subsection 5.5, we conclude with a discussion of the model results and a critical assessment of the functionality of the model. Moreover, the model’s future development steps are outlined, and ideas for further fields of application are presented.

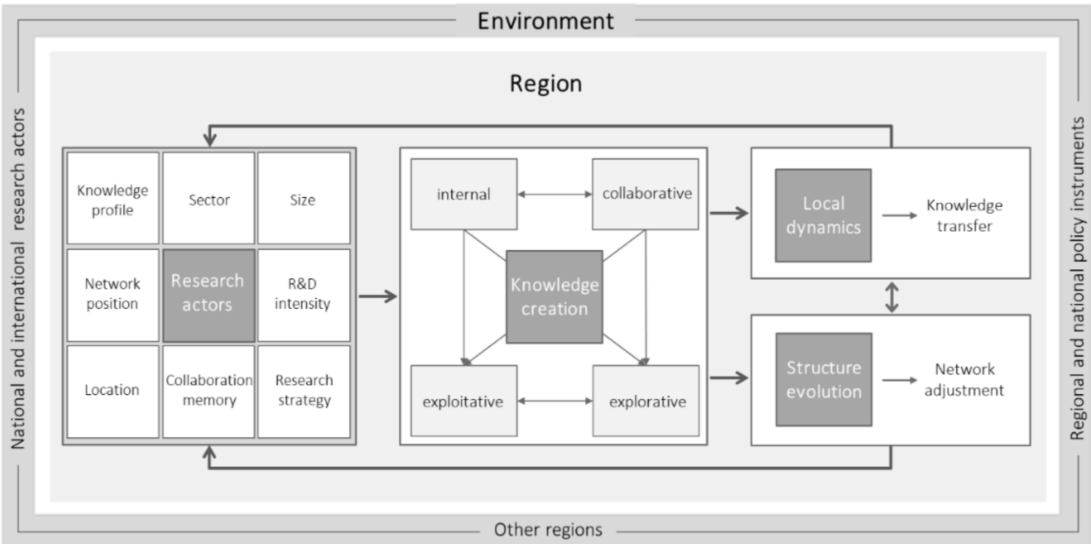
5.2 Model description

This section is dedicated to the description of the multi-region ABM of knowledge creation. The model description is deliberately kept brief; details on model elements and processes are given in the appendix of this section (Subsection 5.6).

Generally, ABMs are developed to discover emergent properties from a bottom-up perspective and – in an attempt to replicate real-world concepts, actions, relations, or mechanisms, they are used to anticipate future developments and outcomes (Nikolic et al. 2013). In this respect, ABM is particularly suited to examine the complex and adaptive nature of regional innovation systems. It provides a framework to model and simulate the behaviour of heterogeneous agents and investigate the complex dynamics of system-wide interactions amongst them. Hence, the aim of this simulation model is to investigate inter-regional knowledge creation across European regions. In doing so, we adopt an empirically driven ABM approach, utilising large-scale data sets on regional knowledge creation and research collaboration activities.

The model conception closely follows state-of-the-art theoretical and conceptual contributions, as well as empirical findings in the fields of regional science, economic geography, and the geography of innovation literature (in particular Feldman 1994, Cooke 2001, Ponds et al. 2007, Boschma and Frenken 2011). Moreover, we integrate ABM and advanced methodological tools from social network analysis (SNA) and econometrics. The simulation model is implemented in Java, drawing on elements of the MASON (Multi-Agent Simulator Of Neighborhoods) multi-agent simulation environment. The proposed model comprises three key characteristics (i) a set of interacting agents, their attributes and behaviours, (ii) a set of relationships and methods of interaction, situated within (iii) a model environment (Macal and North 2010) that serve as cornerstones for the development of the model, visually illustrated by Figure 5.1.

Figure 5.1. *Overview of model elements and processes – region perspective*



In this conception, **agents** are modelled as research actors characterised by organisation-level empirically-based attributes. Each agent is equipped with a *knowledge profile* representing the agent's knowledge endowment, indicative of the technology classes the agent is active in, as well as the expertise in the respective class. Hence, the knowledge endowment of agent i can be defined as a vector of length k (with k being the number of technology classes included in the model)

$$\kappa_i = \{\kappa_{i1}, \kappa_{i2}, \dots, \kappa_{ik}\} \tag{5.1}$$

where $i = 1,2,\dots,N$ with N being the total number of agents in the model, and the value of κ_{ik} determining the expertise in the respective technology class k (level of knowledge).

The agents' *location* is specified by the European NUTS 2 region (EC 2015c) – in total 283 regions in the EU-27 countries plus the United Kingdom and Norway are covered. Since there are no large-scale firm-level data on location, industry association, size and R&D intensity available systematically for European regions, such data are constructed based on region and industry characteristics. Hence, in the process of empirically assigning the industry firm agents to regions, their industry *sector*, *size* (number of employees) and *R&D intensity* are specified as well. The primary objective for the agents in the model is to be representative of each region concerning its characteristics (see Subsection 5.3 for details on agent initialisation). In addition to empirical attributes, agents are also characterised by model-inherent attributes: *research strategy*, *collaboration memory*, and *network position* indicating external networking capability.

The agents' **relationships and interactions** are defined within the *knowledge creation process*, subsuming the total of agents' actions in creating new knowledge. This process is designed as a learning process along a specified research path that each agent decides upon individually according to its *mode of knowledge creation* – exploitative or explorative (following the concept brought forward by March 1991) – and with respect to its current knowledge and research target. Knowledge creation is based on the concept of *technology space*, which is defined as a network comprising a set of technology classes (TCs), with weighted links indicating the technological proximity between these classes (see Figure A5.3 in the appendix of this section for an illustrative example). Formally, the technology space S can be defined as a symmetric matrix

$$S = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}_{k\mathrm{x}k} \tag{5.2}$$

where $p_{..}$ denote the Jaccard coefficients as a measure of proximity (empirically-based), and k is the number of technology classes considered in the model.

The technology space S serves as a framework for the agents to gain new knowledge as they move along their ***research paths.*** Each research path comprises selected technology classes, indicating the way and direction of learning. Generally, a research path P is defined as a subset of all technology classes in the technology space

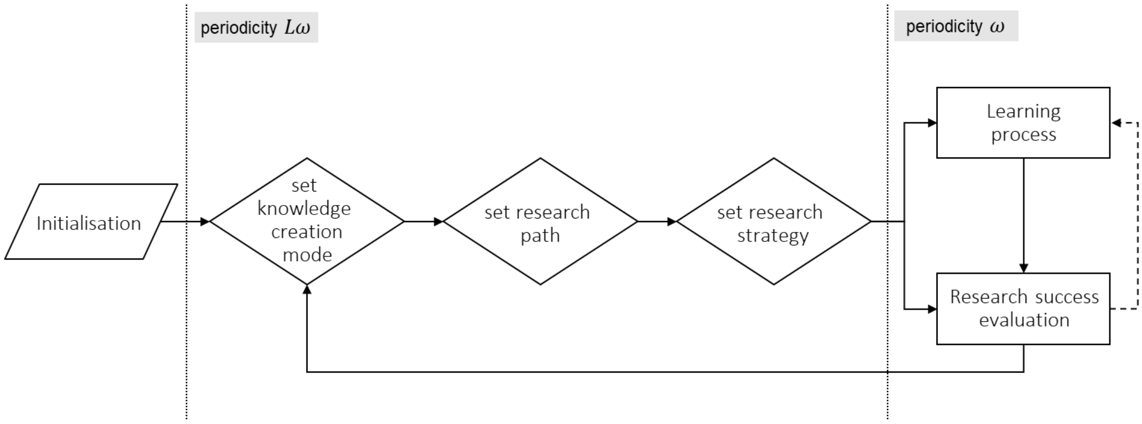
$$P = \{\tau_1, \tau_2, \dots, \tau_L\} \in T = \{\tau_1, \tau_2, \dots, \tau_k\} \tag{5.3}$$

where $l = 1,2,\dots,L$ with L being the length of the research path and T the set of all technology classes.

The agent's individual knowledge creation process follows a predefined sequence of actions (see Figure 5.2 for simplified flow chart); however, it still implies many degrees of freedom, allowing for heterogeneous interactions and processes that result in varying outcomes. In total, the agent's knowledge creation process includes selecting a *mode of knowledge creation*, setting a *research path*, selecting a *research strategy*, and a *learning* and *research success evaluation* process.

The single processes follow two different periodicities: whereas, the subprocesses in the learning process and output evaluation are carried out each model step (periodicity ω), setting a research agenda and research strategy occurs every other step after completing the learning process (periodicity $L\omega$); where ω denotes one simulation step and L an integer value indicating the length of the agent's research path. To initiate the process at the beginning of the simulation, a one-time-only initiation of a starting technology class for the research process takes place; i.e. a random technology with non-zero expertise of the agent's knowledge endowment.

Figure 5.2. *The agent's knowledge creation process (simplified)*



Setting the research path depends on the agent's ***mode of knowledge creation***: *exploitative* or *explorative*; where exploitative knowledge creation reflects a direct and targeted, commercially oriented way of performing research, exploration expresses a non-targeted and more indirect path selection for knowledge creation. In a next step, the agent decides on the ***research strategy*** – i.e. whether to follow this path by means of *internal research*, i.e. perform in-house research, or look for a suitable research partner to perform *collaborative research*. Hence, a core element of the collaborative research process is the agent's choice of a suitable collaboration partner. The partner choice relies on collaboration probabilities resulting from the estimation of a Spatial Interaction Model (SIM) considering the geographical distance between the regions and variables indicating a neighbouring region and country[45]. These probabilities are complemented by collaboration shares based on statistical data.

A final criterion for a suitable partner is cognitive proximity, as in the presence of a certain overlap of knowledge endowments. In collaborative research, a further distinction is made between two *modes of collaboration*, (i) research-mode and (ii) service-mode. Whereas the first mode is aimed at representing basic and applied research projects, primarily focused on the creation and deepening of knowledge, the second mode exemplifies commission projects, usually characterised by an efficient and straight-forward research agenda.

[45] Spatial distance as well as country and region borders are generally acknowledged to be among the most important determinants to explain inter-regional R&D collaborations (e.g. (Scherngell and Barber 2009, Hoekman et al. 2010)). Despite increasing globalisation and new information and communication technologies, spatial proximity is (still) a crucial factor in establishing and maintaining R&D network links (Rallet and Torre 1998, Storper and Venables 2004). Especially, more complex knowledge requires the exchange of tacit knowledge elements via face-to-face interaction.

Knowledge creation is defined as a ***learning process*** along the specified research path, i.e. performing research along trajectories of technology classes. To determine whether a new step (i.e. new technology class) on the research path is reached, the ***research success*** is ***evaluated***. This is a necessary intermediate step for every transition from one to another technology class. The research success is a scaled composite indicator (interpretable as success probability) depending on (i) agent-specific characteristics (overall expertise, R&D quota, internal and external capability) and (ii) technological proximity between the involved technology classes. In the case of successful research, the level of expertise is updated in the respective technology class. If the research is evaluated as not successful, the agent either chooses a new but similar research path or stays on the original path. For collaborative research, the process of research success evaluation is dependent on the collaboration partners' expertise.

Moreover, possible knowledge transfer in the collaborative knowledge creation process is based on the cognitive distance (knowledge distance) between the two collaborating agents. The actual knowledge gain depends on the absorptive capacity – representing the trade-off between novelty value and understandability of new knowledge. Specifically, it is assumed that the amount of knowledge gain corresponds to an inverted u-shaped relationship, i.e. both, low and high knowledge complementary results in low knowledge gains, indicating the presence of an 'optimal distance' entailing a trade-off between 'learning something new' and 'mutual understanding' (e.g. Cohen and Levinthal 1990, Nooteboom 1999).

The **environment** defines the space in which the agents operate; accordingly, it contains all the information external to the agents used in the decision-making process and provides a structure or space for agent interaction (Nikolic et al. 2013). From the perspective of a regional innovation system, we see national and international research actors and other regions as external elements of the specific region. Especially, studies in the vein of regional innovation systems (RIS) stress the importance of such external factors on the knowledge creation of individual research actors located within a specific region, e.g. universities and public research organisations that conduct basic and applied research and regional policy institutions that implement regional innovation policies (Cooke 2001).

External to the whole system of regions and interrelated agents, national and European policy interventions may also affect the region-specific knowledge creation processes. Evidently, these external factors are by no means isolated from the region-internal processes and dynamics but rather strongly interrelate with agent-specific capabilities. In an ABM, this is reflected in the relationship between agent behaviour and its environment comprising external factors – steered by the modeller through exogenous parameters.

5.3 Empirical foundations

This section is dedicated to outline the specification of the model's empirical foundation, including the agent initialisation, calibration and output evaluation. While theoretical models need to be less concerned with methods for initialising the simulation with empirical data, practical applications and policy analyses do require such methods (Kiesling et al. 2012). The empirical foundation is one of the crucial aspects in which the proposed simulation model differs from purely theoretical and conceptual models of regional knowledge creation. A thorough empirical foundation is essential for representing real-world processes, practical applications, and policy analyses since it increases their integrative strength and liability. The empirical foundations of the model complement the conceptual model, as presented in the previous section. In particular, in this model, three central elements are driven by empirical data: agent initialisation, calibration of model parameters, and output evaluation. In addition, throughout the model, agents' decision-making processes are empirically driven by means of statistical figures.

Agent initialisation using spatial microsimulation. The empirical agent initialisation focuses on the generation of a representative agent population for each region. Since detailed micro-level data on an organisational level is not available for European regions comprehensively, model agents are created based on region-level empirical data. Agent-level data has been constructed in an elaborate process of drawing samples from empirical distributions of industry sectors, R&D intensity and the number of employees (determined from the Eurostat Structural Business Statistics for an initialisation period corresponding to the years 2012 to 2014), while considering feasible combinations of characteristics for each agent based on the characteristic's empirical correlations using Cholesky decomposition[46].

[46] Cholesky decomposition can be used to create correlations among random variables (Golub and Van Loan 2013). With Σ being a correlation matrix of empirically observed correlations between agent characteristics, Σ can be uniquely factored into a product $\Sigma = LL^T$ where L is a lower triangular matrix. For instance, given two independent random variables X and Y, the matrix L can be used to create new variables Z and W with $ZW = LXY$ such that the correlation matrix between Z and W equals Σ.

To generate a representative agent population for each region, we employ spatial microsimulation techniques. Spatial microsimulation is a method to allocate individuals (organisations) to zones (regions) by combining individual (organisation-level) and geographically aggregated data (Lovelace and Ballas 2012). Here, we opt for Iterative Proportional Fitting (IPF) as a statistical technique for combining individual and geographical data to allocate the primarily specified agents to European NUTS 2 regions using reweighting algorithms resulting in maximum likelihood values for each zone-individual combination represented in a weight matrix (see, e.g. Lovelace and Ballas 2012 for details on spatial microsimulation and IPF).

Since the overall aim of this model is the simulation of knowledge creation, a special focus lies on the initialisation of the agents' knowledge profiles. Each agent is endowed with a unique set of technological fields – empirically represented by patent classes – representing their knowledge profile. The patenting records are extracted from the PATSTAT database[47]; patent classes are used on a three-digit subclass level (e.g. A61K) as specified by the International Patent Classification (IPC) and are assigned to the agents based on their industry sector (NACE classification) using the table of concordance proposed by Dorner and Harhoff (2018). In total, over 21,000 agents are included in the model, which is a fraction of 1,000 of the actual number of local firm entities located in the NUTS 2 regions of interest (based on the Eurostat Structural Business Statistics).

Calibration of model parameters. The calibration process aims at finding values for the input parameters that make the model reproduce patterns observed in reality sufficiently well (Thiele et al. 2014). Parameter fitting must span the entire set of parameters, which rapidly increases the number of possible parameter combinations to be tested. To reduce the dimension of parameter combinations, we employ Latin hypercube sampling, which is a technique that considers the entire set of parameters to get the most representative subset of the space in a relatively efficient (and computationally saving) manner by means of uniform sampling of the scenario space given a particular parameter space and with a limit of a specified number of experiments (McKay et al. 2000).

[47] PATSTAT is the Worldwide Patent Statistical Database by the European Patent Office and is the most important data source for scientific research on patent activities and patent data.

The core of the empirical calibration is the fitting of model parameters so that the resulting output variables fit best, here, lie within a range of the selected empirical measures. Thiele et al. (2014) point out two different strategies for fitting model parameters to observational data: (i) best-fit and (ii) categorical calibration. Whereas best-fit calibration aims at finding the parameter combination the best fits the observational data (i.e. there exists one exact value as a quality measure to evaluate the fit of the parameter values), using categorical calibration, not a single value is obtained, but a range of plausible values is defined for each calibration criterion.

As proposed by Thiele et al. (2014) a hybrid approach by transforming the categorical criteria to a best-fit criterion is followed here. This is done by means of conditional equations and the specification of a cost function, evaluating the cost for a parameter value of not being in the acceptable value range (which is defined externally)

$$criterium_r\ (x_r) = \left\{ \begin{matrix} 0 & if\ x_{min}\ \leq x_r \leq x_{max} \\ \left(\dfrac{mean(x_{min}, x_{max}) - x_r}{mean(x_{min}, x_{max})}\right) & else \end{matrix} \right\} \tag{5.4}$$

$$cost(x_r) = \sum_{r=1}^{R} criterium_r\ (x_r), \qquad r = 1, \ldots, R \tag{5.5}$$

where x_r are corresponding simulation results of criterium r, x_{min} and x_{max} denote the respective minimum and maximum value, and R is the total number of calibration criteria included. For each selected empirical measure, an acceptable value range is defined. If the simulated value lies within this interval, no costs incur. If this is not the case, a cost factor based on the squared relative deviation to the mean value of the acceptable range is assigned. The final cost function is the sum of the individual costs of each criterion. Finally, the parameter combination with the lowest cost is chosen to fit the real-world system best. Applying this cost function approach enables combining multiple calibration criteria to one single decision criterion (Thiele et al. 2014).

Empirically, four measures are chosen as criteria for the cost function: (i) the total number of patents in the agent population, (ii) the patenting profile across regions, (iii) the patenting profile across technological fields (as defined by Schmoch 2008), and (iv) the regions degree centralities, i.e. the number of collaboration partners in the collaboration network. The empirical reference datasets are the patent data on European regions, as well as – for the centrality measure – data on collaborative research projects in EU Framework Programmes that are widely used to proxy inter-regional R&D collaboration in Europe (see, e.g. Maggioni and Uberti 2009, Scherngell and Lata 2013). The empirical measures are calculated as aggregate values over the years 2014 to 2018; the calibration is performed with the simulated output after 60 time steps (12 time steps representing one year; see the appendix to this section for calibrated system parameters). The calibrated parameter set defines the so-called *baseline scenario*, as a reference for the simulations presented in the results section.

Output evaluation. Successful research efforts by agents result in knowledge gains (see the appendix to this section for details), which furthermore can result in patents. In the simulation model, we use the number of patents as a proxy variable to capture knowledge outputs and establish a link between a rather generic knowledge gain – a pure result of learning processes in the model – and patents. Patents are considered a suitable indicator to measure the ability to create commercially relevant new knowledge, specifically, as an output of industrial innovation efforts in firms (see, e.g. Griliches 1990, Jaffe and Trajtenberg 2002). This allows for an interpretation closer to empirical observations. Whether or not a patent emerges from a knowledge gain follows an independent evaluation criterium. To that, we implement an empirical output filter through econometrically estimated coefficients that determine the patenting propensity of an individual agent based on a region-specific probability that depends on regional characteristics (human resources, GRP per capita, R&D per capita, degree centrality). Due to the lack of organisation-level data, the regression model is estimated on the regional level. We estimate a Poisson regression model to account for the true integer nature and the distributional assumptions of the number of patents as the dependent variable (see the appendix to this section for details).

5.4 Results

In this section, we demonstrate the potential of the simulation model by means of three small example applications derived from current scientific debates. We evaluate the knowledge creation exclusively from an aggregate regional perspective since the model itself is designed to be representative on a regional level (e.g. representativity of agent population). Nevertheless, agent-level processes are reflected in the regional knowledge creation output via local and structural dynamics. On a local level, knowledge is transferred between agents through collaborative knowledge creation processes, which subsequently results in an update of the agents' knowledge profiles. On a network structural level, changes in the agents' knowledge status, as well as knowledge creation performance, affect their network embeddedness and the global network structure as a whole.

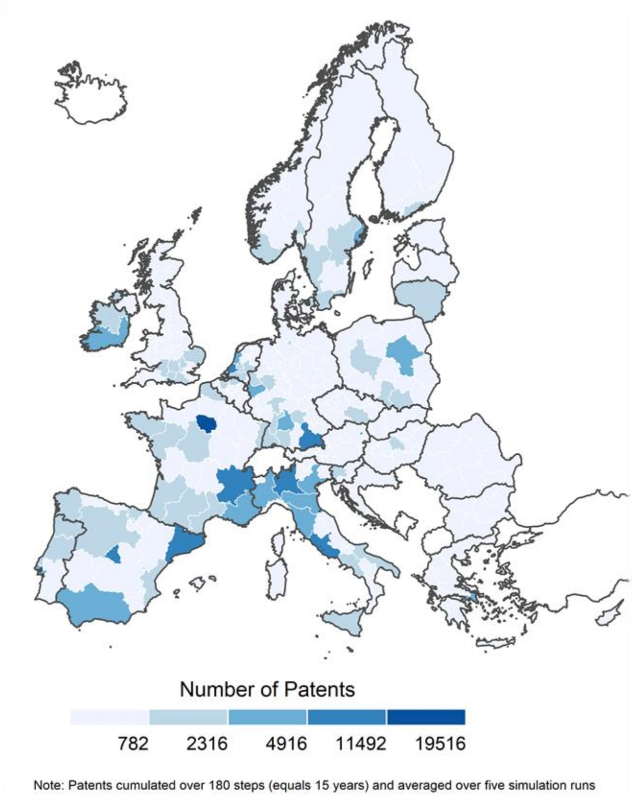
To ensure empirical interpretability and allow for empirical calibration, a link between the model's knowledge gains and patenting, as an empirical knowledge output is established (see Subsection 5.3 on output evaluation). Note that the simulation results presented in this section are averages over five model runs to ensure robust findings and hence, limit the possibility of artefacts occurring by variability in the results. The three small example applications for demonstrating the model's potential are on regional concentration patterns, regional specialisation dynamics, and networks as drivers for regional knowledge creation, all of them intensively debated in the geography of innovation literature.

(i) Spatial distribution and concentration of knowledge creation

The first application aims at the phenomenon of spatial concentration of knowledge creation. To a large extent, knowledge creation is driven by geographically localised knowledge flows, especially in learning processes driven by tacit and region-specific knowledge elements. This highlights the facilitative role of spatial proximity for knowledge creation (Paci and Usai 2000, Acs et al. 2002, Gertler 2003, Moulaert and Sekia 2003). Although, recent findings also suggest a decreasing effect of distance (Glaeser and Kohlhase 2004, Scherngell and Lata 2013). As described in the previous section, we use patents as an empirically-based measure of the model's knowledge output. Figure 5.3 illustrates the spatial distribution of the simulated patents as main model outputs resulting from the agents' individual knowledge creation processes, aggregated to a regional level.

It can be seen that some typical regions, such as, e.g. Île-de-France (FR), Madrid (ES), Catalunya (ES), Oberbayern (DE), Rhône-Alpes (FR), and Northern regions of Italy such as Lombardia clearly stand out in term of their patent outcome, whereas, the majority of regions exhibit only a fair number of patents; although we cannot observe (almost) no distinct spatial clusters of multiple regions showing high patenting activity (except Northern Italy and South-East France).

Figure 5.3. *Spatial distribution of patents (natural breaks)*



In terms of demonstrating the potential of the model, these results are quite promising. Clearly, the model and the implemented processes are able to approximate the empirically observed spatial distribution of knowledge (see Paci and Usai 2000).

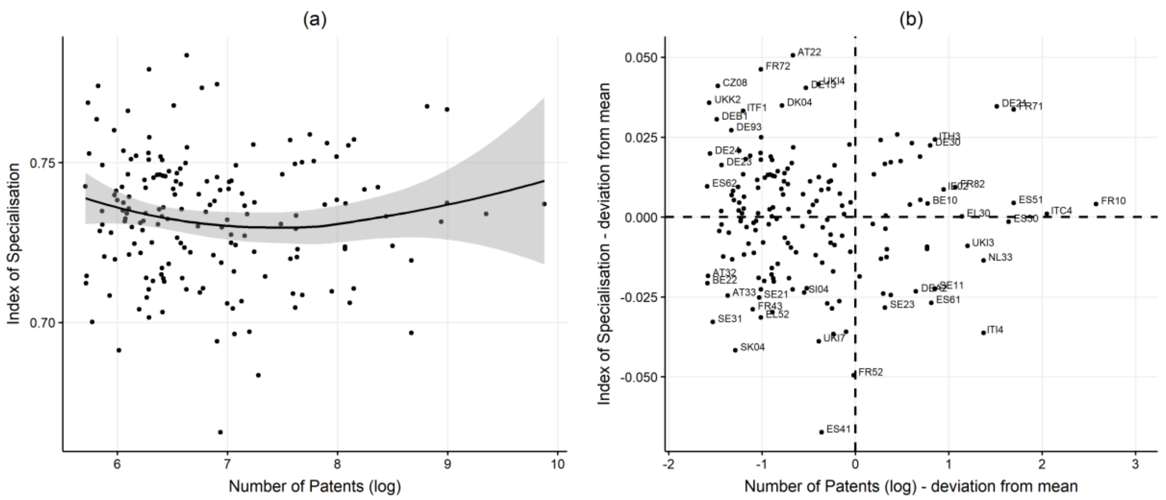
(ii) Specialisation of regional knowledge creation

The second application focuses on the debate of the relative importance of sectoral specialisation versus diversification for a region's knowledge created (Van der Panne 2004, Beaudry and Schiffauerova 2011). This dichotomy is rooted in the two concepts of localisation and diversity economies, as initially put forward by Marshall (1890) and Jacobs (1969), respectively. In this example, we use not only the total counts of simulated patents by region as in the previous example, but also their technological field to calculate the degree of technological specialisation of regions based on the simulated patents[48]. The spatial

[48] We use the Index of Specialisation to assess the degree of specialisation of each region (relatively to the other regions); however, the index does not indicate in which sectors the regions are specialised. The Index of Specialisation is defined as $S_i = 1/2 \sum_{k=1}^{m} |y_{ik} - \bar{y}_k|$ where $y_{ik} = x_{ik} / \sum_{k=1}^{m} x_{ik}$ and $\bar{y}_k = \sum_{i=1}^{n} x_{ik} / \sum_{i=1}^{n} \sum_{k=1}^{m} x_{ik}$ and x indicates patents, and i and k refer to the region $i = 1, \dots, n$ and sector $k = 1, \dots, m$ respectively. The index ranges from 0 to 1, where 1 indicates full specialisation and 0 implies diversification.

distribution of regional technological specialisation is given in Figure A5.8 (see Subsection 5.6), while the relation between technological specialisation simulated knowledge output is illustrated in Figure 5.4. The relationship between the degree of specialisation and the number of patents – as shown in Figure 5.4 (a) – shows no significant correlation. Hence, there is no direct link between the regions' sectoral specialisation and their respective knowledge output in the model.

Figure 5.4. *Relationship between specialisation and knowledge output (patents)*



Note: Only regions with >300 patents are displayed; number of patents used for calculation of index cumulated over 180 steps (i.e. 15 years) and averaged over 5 simulation runs

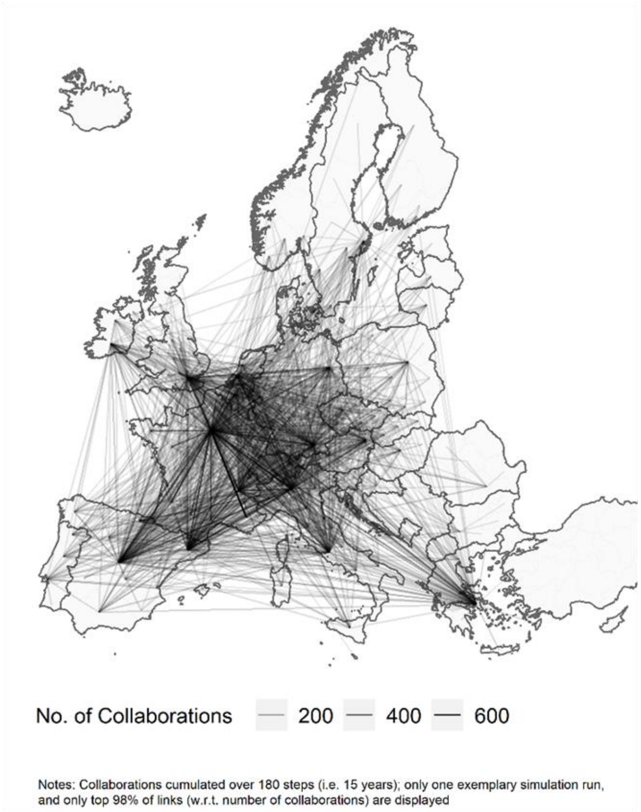
Note: See Table C1 in Appendix for list of region codes; only regions with >300 patents are displayed and number of patents used for calculation of index cumulated over 180 steps (i.e. 15 years) and averaged over 5 simulation runs

On the one hand, according to the concept of localisation economies, a high degree of sectoral specialisation of regions points towards considerable advantages of these regions due to economies of scope when making use of local and specialised R&D infrastructure and local and dense R&D networks that facilitate the exchange of knowledge at relatively low costs. Looking at Figure 5.4 (b), which displays the centralised number of patents and specialisation indices (i.e. deviation from the respective mean values), this applies in the model simulation to the regions Rhône-Alpes (FR71) and Oberbayern (DE21) that exhibit relatively high technological specialisation and knowledge output. This may signal the importance of sectoral specialisation to gain a higher output – a finding in the vein of Marshall (Marshall 1890, 1920). On the other hand, e.g. the regions Lazio (ITI4) and Andalucía (ES61) suggest that a relatively low degree of specialisation (i.e. diversification) and knowledge output are positively related – supporting diversity economies as put forward by Jacobs (Jacobs 1969). Hence, although there is no clear relationship between knowledge output and degree of technological specialisation of regions, we find specialist regions with industrial districts or sector-specific spillovers and generalists that benefit from industry diversification among the leading regions in terms of knowledge creation.

(iii) Networks as drivers for regional knowledge creation

In the third application, attention is shifted towards networks – viewed as inter-organisational arrangements in R&D – that are widely considered essential for increasing a region's knowledge creation capability (e.g. Bathelt et al. 2004, Rodríguez-Pose and Crescenzi 2008, Neves and Sequeira 2018, Wanzenböck and Piribauer 2018). In this study, the network under consideration is defined as a regional knowledge network comprising a set of regions as nodes, inter-linked via edges representing the knowledge flows resulting from collaborative R&D efforts (Scherngell and Barber 2009, Sebestyén and Varga 2013). Figure 5.5 displays the simulated regional knowledge network showing the collaboration links between the agents (when following the collaborative mode of knowledge creation) aggregated to a regional level.

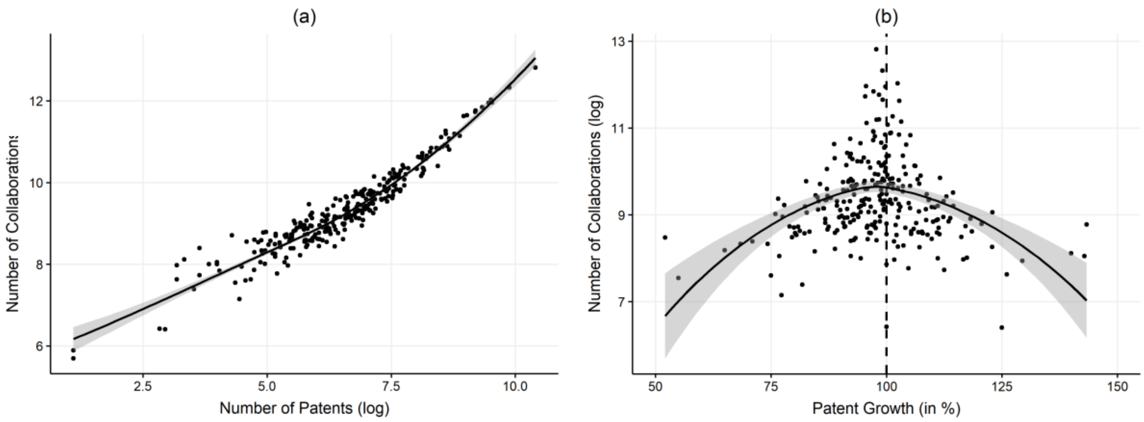
Figure 5.5. *European inter-regional R&D collaboration network*



As for the spatial distribution in the first example, the model is able to re-create observed spatial network patterns in the literature using project collaboration with the EU Framework Programmes (see, e.g. Scherngell 2019). The European regions seem quite strongly engaged in research collaborations, however, only a few network hubs (in terms of their number of network partners) stand out: first, Île-de-France (FR), leading to the characteristic star-shaped network formation that is also known from empirical studies (Scherngell and Barber 2009), followed by Oberbayern (DE), Madrid (ES), Lombardia (IT), and Rhône-Alpes (FR). These regions also exhibit the highest knowledge output in the model, suggesting a positive relationship between a region's network connectivity and knowledge output (as measured by patents).

In Figure 5.6, we further reflect on the relation between the regions' numbers of simulated collaborations and knowledge output. Figure 5.6 (a) shows a positive and slightly exponential relationship between the number of collaborations and quantity of knowledge output (as measured by the number of patents), implying that the more collaborations the agents – located within the regions – have, the higher are the regions' knowledge outputs. Note that these findings only reflect the quantity of network links (number of collaborations of regions), not their quality that would identify certain regions as hubs with authoritative positions in the collaboration network.

Figure 5.6. *Relationship between number of collaborations and knowledge output (patents)*



Notes: Patents cumulated over 180 steps (i.e. 15 years) and averaged over five simulation runs; average number of collaborations over five simulation runs and 180 steps

Notes: Patent growth (in %) determined as $(p_t - p_{t-1})^* \, 100/p_t$ where p_t denotes cumulated patents over steps 36-108, and p_{t-1} denotes cumulated patents over steps 36-180; average number of collaborations over five simulation runs and 180 steps

However, looking at the relationship between the number of collaborations and patent growth in Figure 5.6 (b), we cannot observe that a high number of collaborations also coincides with high patent growth. Hence, a high number of international collaborations is not a driving force for high patent growth (i.e. growth in knowledge output) in the model. Evidently, regions starting from a relatively high level of patent output exhibit lower rates of patent growth.

5.5 Discussion and concluding remarks

This study introduces an empirical agent-based model of multi-regional knowledge creation and demonstrates its potential for applications to current research issues intensively debated in the geography of innovation literature. By employing an agent-based simulation approach, we intend to complement the prevailing research tradition of econometrically modelling regional knowledge creation, focusing on regional characteristics and knowledge creation and diffusion determinants. ABM offers several benefits compared to these conventional modelling techniques, which allow for new perspectives and insights in the process of regional knowledge creation. Especially, agent heterogeneity, underlying micro-structures of the regions, and network dynamics as an interplay between region-internal and region-external interdependencies in the form of R&D collaboration linkages can be explicitly considered. Moreover, the possibility to conduct simulation scenarios allows comparing system behaviour within a controlled environment directly.

However, the ABM approach is – up to now rarely used in the context of the geography of innovation. In our understanding, this is to a large extent due to the lack of credibility and lack of empirical closeness and hence, lack of applicability to real-world questions. We react to this by drawing on large-scale data sets and applying state-of-the-art methods to initialise and calibrate the simulation model empirically. Moreover, we support the ABM by employing well-established econometric tools and concepts of network science, using 'the best of each world', which we believe changes the perception of methodological criticisms regarding simulation models being a black-box.

The example applications of the model show quite promising results in terms of robustness and empirical approximation, speaking for the model's representativity. We can show the spatial concentration of knowledge creation, illustrate the mechanisms of the ambiguity of the effect of sectoral specialisation versus diversification on knowledge created, and confirm the driving role of networks for regional knowledge creation. The replication of real-world phenomena, supported by empirical findings in related studies, is an essential step of the model validation. Hence, we conclude that the proposed simulation model indeed shows the potential to advance the scientific debate in the field of geography of innovation in future applications making use of simulation experiments.

Additionally to contribute to the scientific debate on regional knowledge creation, the proposed simulation model is also of high relevance in the field of research, technology and innovation (RTI) policy. Considering technological, institutional and geographical aspects of knowledge creation in the model, it allows for simulation experiments referring to RTI policy measures at the European, national, and regional levels. For instance, such policy interventions may refer to regional specialisation policies, the coordination of regional policies, increased incentives to engage in R&D collaborations or to mission-oriented public funding of specific thematic areas. One particular field of application to this respect is using the ABM for ex-ante impact assessment of policy interventions, such as public R&D programs. Especially, the evolutionary and forward-looking perspective of ABMs considers the openness of socio-technical development, and the micro-perspective on agent systems may help to understand the complexity of public policy interventions.

For both, scientific research issues as well as policy aspects, the current model is sufficiently flexible to be easily tailored to new research issues of interest while relying on the robustness of the core model elements and processes. Admittedly, the proposed model also has limitations that the modeller has to be aware of: *First*, the knowledge creation process is tailored to Europe as a geographical entity. On the one hand, this is the case in terms of the empirical initialisation and calibration data. On the other hand, this also applies implicitly with respect to the model elements and processes since the model conception is driven by the European spirit of performing R&D within the EU Framework Programmes that connect regions all over Europe via collaborative, publicly funded research projects. It remains to be examined if the model is also suited for other regional, national and supra-national innovation systems, such as China or the US. *Second*, the model aims to simulate regional knowledge creation in sufficient detail. We deliberately exclude any considerations on the valuation of the newly created knowledge and its measurement. Hence, one has to be aware that respective statements cannot be made.

Nevertheless, to have some approximation, we include the distinction between simple knowledge gains and patents resulting from a knowledge gain in the model (based on econometrically estimated probabilities using empirical information on region characteristics). *Third*, regarding the representativity of the model, it is explicitly adapted to a regional level. Although the agents are modelled on an organisational level and hence, also their knowledge creation and learning processes, the model's initialisation and calibration are targeted at model results that are representative for regions. This entails that we need to refrain from any analyses on the agent level, such as observing a single agent's behaviour, to ensure the credibility of the results.

In this study, we demonstrated the potential of the proposed simulation model with first application examples. However, many possibilities and model features have not been exploited so far (e.g. knowledge gain, learning processes), allowing for many future applications and simulation experiments. Specifically, characterising the influence of inter-regional R&D collaboration on regional knowledge creation, disentangling local effects from global network effects on regional knowledge creation, or the analysis of technological specialisation and geographical concentration tendencies come to mind. In particular, scenario analyses referring to specific RTI policy measures are of interest to shed light on the mechanisms of policy interventions on the European, the national and the regional level. Moreover, the application of the model to other geographical areas, such as China, is of great interest to gain a unique comparative perspective of regional knowledge creation in innovation systems showing different development paths, different overall socio-economic characteristics and conditions, different approaches in policymaking and societal systems as a whole.

5.6 Appendix to Article IV

Technical Appendix – Glossary of model elements and processes

In alphabetical order

Collaboration memory. The collaboration memory is specified as a vector of length s (steered by external model parameter) containing entry pairs of the last s former collaboration partners a_i with a respective probability value γ_s representing how successful the past collaboration has been

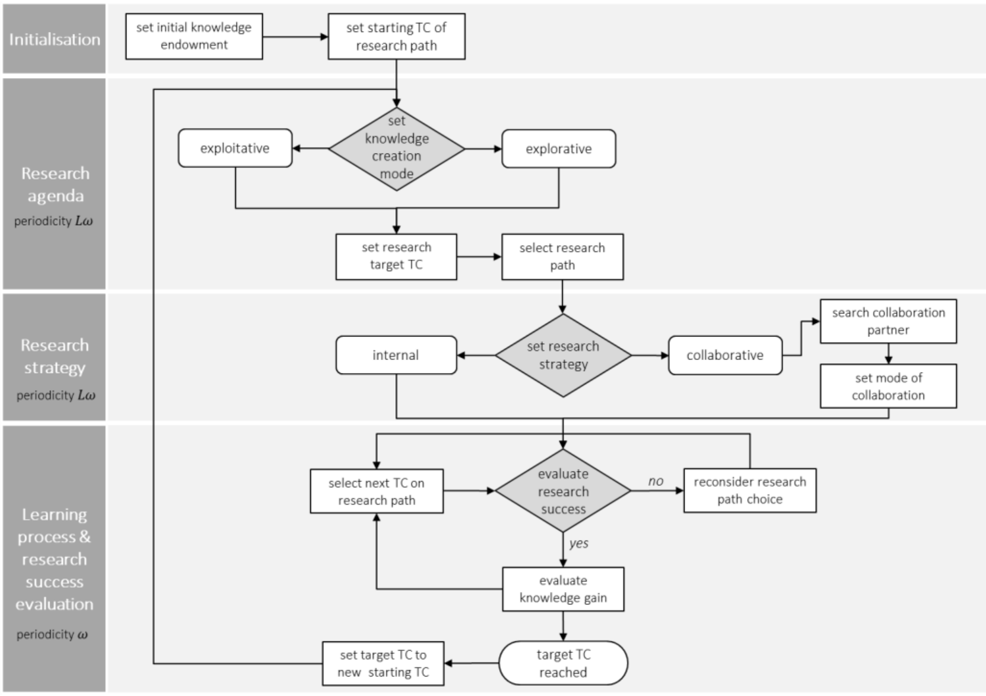
$$\Gamma = \left\{ \binom{a_i}{\gamma_1}, \dots, \binom{a_i}{\gamma_s} \right\} \qquad \text{for } i \in N \tag{A5.1}$$

where N is the set of total agents. The collaboration memory vector is renewed in a way that new collaboration partners are ranked first in the vector, while the partner ranked last (i.e. the one longest in the collaboration memory) drops out.

To determine the degree of success of a collaboration, the share of the actual collaborative knowledge gain over the whole research path with respect to the maximum possible knowledge gain is evaluated; this share is interpreted as probability for a repeated collaboration. Within the partner choice process, a random entry pair is selected, and the respective probability is evaluated. Whereas a positive return leads to a repeated collaboration between the two agents, a negative return initiates the remaining process of partner choice.

Knowledge creation process. The knowledge creation process subsumes the total of the agents' actions in creating new knowledge, including the setting of *knowledge creation mode*, the setting of *research path*, selection of *research strategy*, *learning process*, and *research success evaluation* (see also Figure 5.2 in the main text for simplified process diagram). In Figure A5.1, the agent's knowledge creation process is illustrated in more detail.

Figure A5.1. *Detailed illustration of an agent's knowledge creation process*



Knowledge endowment/profile. The knowledge endowment of agent i represents the knowledge profile of the agent, i.e. indicates the technology classes the agent is active in, as well as the expertise in the respective class. Hence, the knowledge endowment can be defined as a vector of length k (with k being the number of technology classes in the *technology space*)

$$\kappa_i = \{\kappa_{i1}, \kappa_{i2}, \dots, \kappa_{ik}\} \tag{A5.2}$$

where the value of κ_{ik} determines the expertise in the respective technology class k (level of knowledge). Combining all vectors of the agents' knowledge endowments results in the *knowledge space*.

Knowledge gain. In the case of successful research, the agent's knowledge gain is evaluated, i.e. the level of expertise in the subsequent technology class (TC) is updated. In the case of *internal research*, the learning outcome ψ_t in time step t can be written as

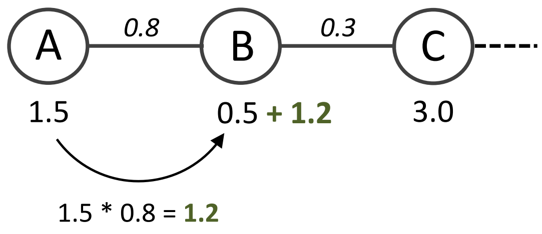
$$\psi_t = \kappa_{il,t} * (1 - p_{\tau_l,\tau_{l+1}}) \tag{A5.3}$$

where $p_{\tau_l,\tau_{l+1}}$ denotes the proximity between two consecutive TCs on the research path in the technology space (indicating the similarity of the two TCs); this increases the expertise level κ_{ik+1} of

$$\kappa_{il+1,t+1} = \kappa_{il+1,t} + \psi_t \tag{A5.4}$$

An example of an update in the agent's expertise level is given in Figure A5.2. Departing from technology class A with current expertise of 1.5, the knowledge gain with a transition to TC B on the learning path – bridging a distance of 0.8 ($1 - p_{A,B}$) between the two TCs amounts to 1.2.

Figure A5.2. *Example of expertise level update*



In the case of *collaborative research*, there is an additional ***knowledge transfer*** between the collaborating agents. Knowledge transfer in the collaborative knowledge creation process is based on the cognitive distance (knowledge distance) d_{ijk} between the two collaborating agents

$$d_{ij} = \sqrt{\sum (\kappa_{ik} - \kappa_{jk})^2} \tag{A5.5}$$

To determine the increase of expertise for the technology class k of interest, the distance of the levels between the collaborating agents is determined; however, the agent only gains from the partner's knowledge if the partner's expertise is higher than its own. This is represented by

$$d_{ijk} = \max[(\kappa_{jk} - \kappa_{ik}), 0] \tag{A5.6}$$

The actual amount learnt depends on the absorptive capacity – representing the trade-off between novelty value and understandability of new knowledge. Hence, it is assumed the amount of knowledge gain corresponds to an inverted u-shaped relationship. Specifically, this relationship is used to scale d_{ijk}

$$\epsilon_k = -\frac{1}{\delta}(d_{ijk} - \delta)^2 + \delta \tag{A5.7}$$

where δ denotes the optimal learning distance (which is specified by an external model parameter). As in the case of knowledge gain in internal research, this scaled expertise level distance is set in relation to the technological distance that is overcome from one to the next technology class on the research path

$$\varphi_t = \epsilon_k * \left(1 - p_{\tau_l,\tau_{l+1}}\right) \tag{A5.8}$$

In total, the knowledge gain in each technology class when performing collaborative research is specified as an additive function comprising knowledge gains from the *internal research* ψ_t the agent is performing regardless of the collaboration partner and *collaborative research*. This can be written as

$$\Delta \mathrm{E}_t = \psi_t + \varphi_t \tag{A5.9}$$

Knowledge space. Combining all vectors of the agents' knowledge endowments results in the knowledge space. With N being the total number of agents in the model, that can be interpreted as the total portfolio and amount of knowledge available in the model

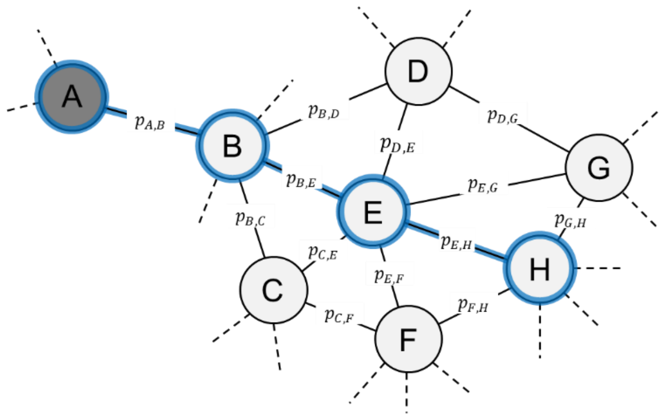
$$\mathrm{K} = \begin{pmatrix} \vec{\kappa}_1 \\ \vdots \\ \vec{\kappa}_N \end{pmatrix} \tag{A5.10}$$

Learning process. The core of the knowledge creation process is the learning mechanism. It is designed as a sequenced process along a specified *research path* that each agent decides upon individually according to its *mode of knowledge creation* (exploitative or explorative) and with respect to its current knowledge and research target. The basis for knowledge creation along a research path is the concept of *technology space*. It serves as a framework for the agents to gain new knowledge as they move along their research paths comprising *technology classes*. Each research path P is a subset of the technology space, indicating the way and direction of learning (following the concept of a 'path' known from social network analysis).

$$P = \{\tau_1, \tau_2, \dots, \tau_L\} \in T = \{\tau_1, \tau_2, \dots, \tau_k\} \tag{A5.11}$$

where $l = 1,2, \dots, L$ with L being the length of the research path. Assuming technology class A being the present knowledge that is built upon, the agent moves along its research path (e.g. $\{A, B, E, H\}$ as illustrated in Figure A5.3).

Figure A5.3. *Example of research path in a subset of the technology space*



Between each transition to the next technology class, the research is evaluated to be successful or not. Only in the case of successful research, the new technology is acquired; otherwise, the agent tries again or eventually chooses an alternate path. How this path is chosen depends on the mode of knowledge creation.

Mode of collaboration. There are two different modes of collaboration (the ratio is steered by an external parameter): (i) *research-mode* and (ii) *service-mode*. Technically, following the research-mode, both partners follow the research path as determined by the partner actively looking for a collaboration partner, assuming he is the consortium leader and hence specifies the research direction. However, since the partner has (by definition of partner search) expertise on at least one technology class on that research path, the agent can learn from the partner in these classes; in a sense that he receives a certain amount of the partner's expertise – additionally to his own research results (the partner's level of expertise is not reduced by this). How much the level of expertise is increased (how much is learnt in each technology class) depends on the cognitive distance between the two collaborators and the absorptive capacity (see *'knowledge gain'*).

Alternatively, following the service-mode, it is evaluated if the collaboration partner has higher expertise in one of the technology classes on the selected research path; if so, the starting TC of the research path to reach the research target is changed to the closest TC to the target TC (but not the target TC itself) where the partner has higher expertise. Hence, the research target is probably reached faster; however, there is potentially less knowledge gain since there are fewer possibilities to gain knowledge in the particular technology classes. In the case the partner has no higher expertise level in any of the technology classes on the research path, its collaborative strategy is shifted to the research-mode. Although the agent is still not able to learn directly from the partner (less expertise in all relevant TCs), the success of the research is (most likely) positively influenced by the collaboration partner (see *'research success evaluation'*).

Mode of knowledge creation. In selecting their research strategy, the agents can choose between *exploitative* and *explorative* knowledge creation, where exploitative knowledge creation reflects a direct way, and exploration a more indirect way of creating knowledge. Dependent on the research strategy, the agents' set their research path P. In the case of exploitation, having set the starting point of the research path in the technology space, the agent chooses a research target (a target TC). The selection of the target TC occurs according to a decaying probability function based on the technological distance between the classes in the technology space, such that closer TCs exhibit a higher probability of being chosen as a target. Hence, the distance between the starting TC and target TC indicates the degree of radicality λ of the agent's research endeavour.

Next, the agent identifies the set of shortest paths and can either select the shortest weighted path or one of the shortest paths with respect to the number of TCs on the path. In the case of exploration, the research path is determined by subsequently choosing the next most proximate TC (originating from the agent's current TC), where the length of the path is determined randomly, representing the equivalent of a researching period between one and five years (set by external parameter). The last TC of the research path is specified as the designated target TC.

Output evaluation. Whether or not a patent emerges from a knowledge gain is determined by employing econometrically estimated coefficients influencing the patenting propensity of an individual agent through a region-specific probability determined by regional characteristics, i.e. human resources, GRP per capita, R&D expenditures per capita, and degree centrality. Note that only fully accomplished research paths are subject to the evaluation for a patent, where each technology class on the path represents a patent class (analogously to patent documents issued by, e.g. the European Patent Office). We estimate a standard Poisson regression model (see Long and Freese 2006) to account for the true integer nature and the distributional assumptions of the number of patents as the dependent variable. The estimated parameters are used to compute the region's predicted empirical probabilities (to receive at least one patent) by means of

$$\widehat{\Pr}(y > 0|\mathrm{x}) = 1 - \widehat{\Pr}(y < 1|\mathrm{x}) = 1 - \frac{e^{-\mathrm{x}\hat{\beta}}\left(\mathrm{x}\hat{\beta}\right)^{m}}{m!} \quad \text{with} \quad m = 0 \tag{A5.12}$$

Parameters. The model comprises external system parameters that are not empirically based. The first set of parameters (Table A5.1) is fixed to values determined by the calibration process and specifies a baseline scenario.

Table A5.1. *Calibrated external system parameters*

Parameter	Description	Type	Calibrated value
BasePatentProb	Scaling parameter for patent probability (originally estimated econometrically)	$[0,1] \in \mathbb{R}_0$	1.0
CollabInternalProb	Share of agents performing collaborative research (vs. internal research)	$[0,1] \in \mathbb{R}_0$	0.5
CollabMemorySize	Length of collaboration memory vector (determines number of former collaboration partners being remembered)	$[1,10] \in \mathbb{N}$	9
CollabModeProb	Share of agents with service-oriented mode of collaboration (vs. research-mode)	$[0,1] \in \mathbb{R}_0$	0.8
ResearchStrategyProb	Share of agents with exploitative mode of knowledge creation (vs. explorative)	$[0,1] \in \mathbb{R}_0$	0.3

Additionally, the second set of external system parameters (Table A5.2) is purely specified by means of user input (i.e. they are not calibrated and not empirically based).

Table A5.2. *Non-calibrated external system parameters*

Parameter	Description	Type	Initialisation value
ExplorativePathLength	Indicates the maximum length of research project for explorative mode of knowledge creation (1 step = 1 month)	$\mathrm{unif}(0, x)$, $x \in \mathbb{R}_+$	5
Delta [δ]	Determines how much is learnt from the partner in collaborative research (optimal learning distance)	$\mathbb{R}_+$	1
Lambda [λ]	Determines the degree of radicality in search for a research target technology class (in exploitative research)	$[0,1] \in \mathbb{R}_0$	1.0

Partner choice. A core element of the collaborative research process is the agent's choice of a suitable collaboration partner. Ahead of the general process of partner choice, a collaboration memory (see *'collaboration memory'* for details) serves to account for re-occurring collaborations with partners of previously successful joint research projects. The general partner choice is organised along three main steps covering the spatial, sectoral, and cognitive dimension to guarantee a most possible real-world collaboration behaviour of agents. For the *spatial dimension*, i.e. the choice of a probable empirical region to look for a suitable partner agent, we draw upon estimated probabilities for collaboration between two agents located in two regions. Therefore, we estimate a Spatial Interaction Model (SIM; see Fischer and Wang 2011 for details) using data from the EUPRO database[49] comprising systematic information on collaborative research projects in EU Framework Programmes and explicitly take the geographical distance between the regions and variables indicating a neighbouring region and country into account in the model. Thus, we receive individual collaboration probabilities for all combinations of regions with respect to their geographical relations. Concerning the *sectoral dimension*, to determine the sector where the partner search is carried on, the proximity between the sectoral classes (based on co-occurrences of IPC patent classes attributed to each NACE class in Dorner and Harhoff 2018) is used to identify a suitable sector; closer sectors exhibit higher probabilities to be chosen. Once, empirically, a suitable region and sector for the partner choice have been identified, the agent looking for a cognitively proximate collaboration partner, i.e. a partner having expertise in one of the technology classes being on its research path.

Research path. A research path P is a subset of the technology space, indicating the way and direction of learning (following the concept of a 'path' known from Social Network Analysis).

$$P = \{\tau_1, \tau_2, \dots, \tau_L\} \in T = \{\tau_1, \tau_2, \dots, \tau_k\} \tag{A5.13}$$

where $l = 1,2, \dots, L$ with L being the length of the research path and T the set of all technology classes (see *'learning process'* and *'technology space'* for details).

Research strategy. There are two different research strategies: *internal* and *collaborative* research. Dependent on the research strategy, different mechanisms are in place regarding the *learning process, knowledge gain* and *research success evaluation* (see respective items in this appendix for details).

Research success evaluation. The evaluation of the research success is a necessary intermediate step for every transition from one technology class to another, i.e., determining whether a new TC on the path is reached. In the case of *internal research*, the research success is determined through a scaled composite indicator (interpretable as success

[49] The EUPRO database comprises systematic information on EU funded collaborative research projects of FP1-FP7 and H2020. It includes information on participating organisations, such as their name, type, and geographical location (see risis2.eu for details and access).

probability) depending on (i) agent-specific characteristics c_{im} (overall expertise and R&D quota) and (ii) technological proximity between the involved technology classes $p_{\tau_l,\tau_{l+1}}$. This can be formalised as

$$P(success) = \gamma A + (1-\gamma) p_{\tau_l,\tau_{l+1}} \qquad \text{(A5.14)}$$

with $A = \sum \frac{1}{m} \mathbf{1}(c_{im} > \text{median}(c_{\cdot m}))$ and $p_{\tau_l,\tau_{l+1}} \in [0,1]$, where m indicates the number of agent-specific characteristics included. For *collaborative research*, evaluating the research success is similar to the case of internal research, however, differs regarding the agent-specific characteristics; such that the maximum value of either agent in the collaboration partnership is used for the evaluation of the research success, and hence, increasing the probability for a success. Recalling the formula of the research success $P(success) = \gamma A + (1-\gamma) p_{\tau_l,\tau_{l+1}}$, A is now defined as

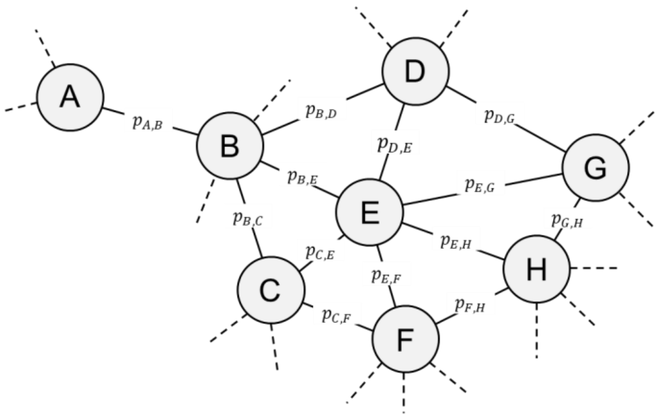
$$A = \sum \frac{1}{m} \mathbf{1}(\max(c_{im}, c_{jm}) > \text{median}(c_{\cdot m})) \qquad \text{(A5.15)}$$

Technology space and technology class (TC). The technology space is defined as a network comprising a set of technology classes, with the links indicating the technological proximity between these classes (see Figure A5.3 for an illustrative example). The network is constructed by extracting patent data from the European Patent Office for the EU-27 countries, including United Kingdom and Norway, from 2012 to 2016 and determining the co-occurrences of IPC patent classes (3-digit) on patent documents. Formally, the technology space S can be defined as a symmetric matrix

$$S = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{pmatrix}_{k\text{x}k} \qquad \text{(A5.16)}$$

where $p_{\cdot\cdot}$ denote the Jaccard coefficient as a measure of proximity (derived from the co-occurrences of IPC patent classes), and k is the number of technology classes considered. In Figure A5.4 an example of an exemplary subset of the technology space is illustrated. The nodes indicate technology classes, and links represent connectivity between these nodes.

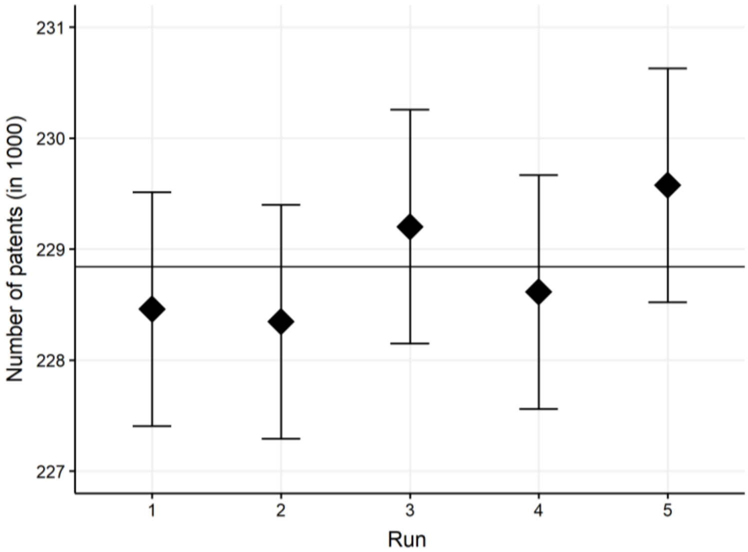
Figure A5.4. *Exemplary subset of the technology space*



Robustness checks

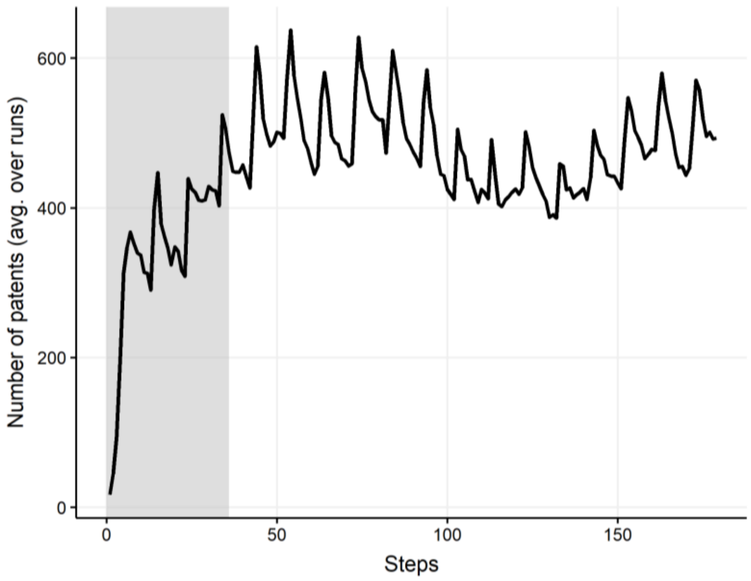
The robustness checks present basic model behaviour over time and model runs to demonstrate the dynamics and robustness of the simulation model with respect to the variability of model outcomes. The results of the robustness checks presented cover 180 steps (equals 15 years) and five simulation runs.

Figure A5.5. *Distribution of patents over runs*



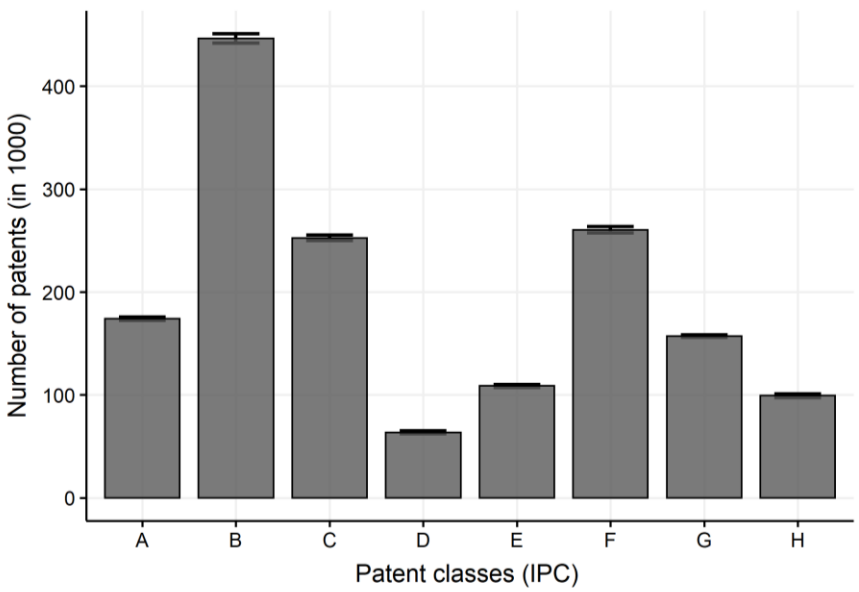
Notes: Error bars indicate an interval of twice the standard deviation; horizontal line denotes the mean value of the runs

Figure A5.6. *Number of patents per step (average over runs)*



Note: The grey area indicates an adjustment phase of 36 steps (equals 3 years)

Figure A5.7. *Sectoral distribution of patents (average over runs)*



Notes: Error bars indicate an interval of twice the standard deviation; patent classes following the International Patent Classification (IPC): A…Human Necessities, B…Performing Operations and Transporting, C…Chemistry and Metallurgy, D…Textiles & Paper, E…Fixed Constructions, F…Mechanical Engineering, Lighting, Heating, Weapons and Blasting, G…Physics, H…Electricity.

Supplementary material

Figure A5.8. *Sectoral specialisation of regions (natural breaks)*

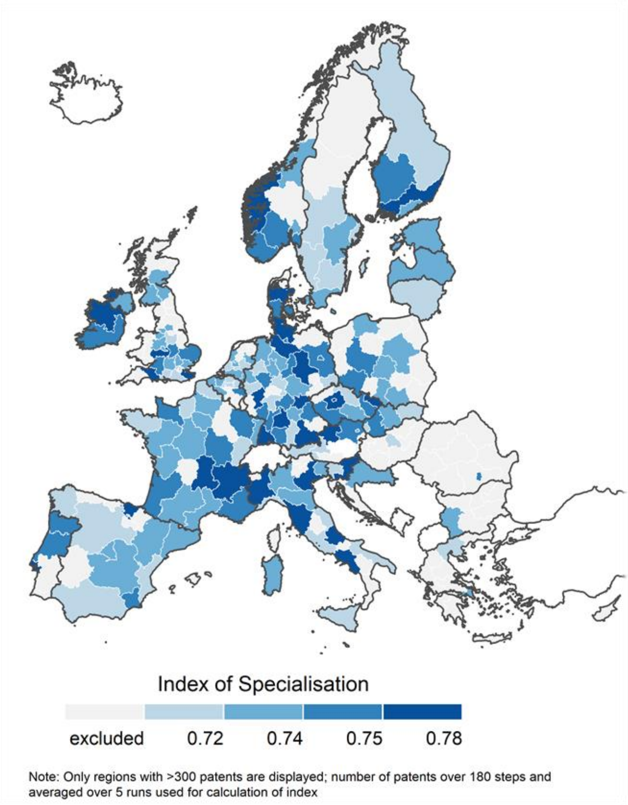


Table A5.3. *Region codes (selected)*

NUTS 2 Code	Region name	NUTS 2 Code	Region name
AT32	Salzburg	FR43	Franche-Comté
AT33	Tirol	FR52	Bretagne
BE10	Région de Bruxelles-Capitale	FR71	Rhône-Alpes
BE22	Prov. Limburg	FR72	Auvergne
CZ08	Moravskoslezsko	FR82	Provence-Alpes-Côte d'Azur
DE13	Freiburg	IE02	Southern and Eastern Ireland
DE21	Oberbayern	ITC4	Lombardia
DE23	Oberpfalz	ITF1	Abruzzo
DE24	Oberfranken	ITH3	Veneto
DE30	Berlin	ITI4	Friuli-Venezia Giulia
DE93	Lüneburg	NL33	Zuid-Holland
DEA2	Köln	SE11	Stockholm
DEB1	Koblenz	SE21	Småland med öarna
DK04	Midtjylland	SE23	Västsverige
EL30	Αττική	SE31	Norra Mellansverige
EL52	Κεντρική Μακεδονία	SI04	Zahodna Slovenija
ES30	Comunidad de Madrid	SK04	Východné Slovensko
ES41	Castilla y León	UKI3	Inner London - West
ES51	Cataluña	UKI4	Inner London - East
ES61	Andalucía	UKI7	Outer London - West and North West
ES62	Región de Murcia	UKK2	Dorset and Somerset
FR10	Île de France		

6 Conclusions

In the recent past, the perception of regional knowledge creation has evolved from being strongly localised and unequally dispersed in space to an increasingly interlinked and inter-regional phenomenon. Accordingly, attention has been shifted to the investigation and modelling of regional knowledge creation processes by means or their region-internal and region-external knowledge interactions in the form of R&D collaborations. In this context, recent research endeavours have provided initial evidence on the complex relationships between knowledge creation and inter-regional R&D collaboration networks (see, e.g. Varga et al. 2014, Wanzenböck and Piribauer 2018), as well as on potential barriers and drivers affecting the establishment and dynamics of such networks (see, e.g. Scherngell and Barber 2009, Bergé 2017). In the vein of this research, this dissertation takes on new perspectives to advance the understanding of the interplay between regional knowledge creation and R&D collaboration networks. Specifically, it offers new theoretical and empirical perspectives by systematically shifting the focus to different sources of heterogeneity in the knowledge creation process and emphasises a diversified picture of the mechanisms being in place by adopting and combining different methodological approaches.

In particular, this dissertation fathoms new evidence-based insights on how R&D collaboration networks drive regional knowledge creation, accounting for different kinds of heterogeneity, by

(i) investigating the role of geographical and relational effects driving the constitution of such networks while accounting for technological idiosyncrasies,
(ii) identifying the impact of network embeddedness on regional knowledge creation for distinct technological fields,
(iii) analysing the effects of R&D collaboration networks on different modes and outputs of regional knowledge creation and the impact of neighbouring regions in terms of these effects, and
(iv) proposing a simulation perspective on actor-specific heterogenous knowledge creation mechanisms employing a multi-region agent-based model.

The dissertation is strongly empirically driven in using novel, large-scale data on R&D activities in its attempt to provide robust evidence against actual theoretical debates of the field. Most importantly, data on collaborative R&D projects of the EU FPs are mobilised to proxy pan-European inter-regional R&D collaboration networks, along with quite detailed data on regional patent and publication activities that comprise the primary empirical basis for capturing regional knowledge creation. Fostering inter-regional knowledge flows all over Europe via collaborative, publicly funded research projects, the EU FPs constitute an integral element of the European research landscape. This rationale is also strongly reflected by the simulation model developed (as part of Article IV of the cumulative dissertation).

With the conceptual and theoretical embedding in the state-of-the-art literature and the targeted advancement of different methodological approaches, the dissertation significantly contributes to, and advances the current scientific debate, both from a theoretical, but in particular from a methodological and empirical perspective. Most significantly, the strong empirical focus of all three research articles allows for meaningful conclusions in *scientific debates* regarding the real-world complexity of regional knowledge creation processes and hence, enables establishing a link to ongoing *policy debates*. Subsequently, the main findings for both debates are emphasised.

With respect to the ***scientific debate***, particularly, the main contribution can be summarised by the following points:

First, the dissertation **reinforces the theoretical debate on regional innovation systems and their dynamics** by means of fine-grained and original empirical insights obtained from the four articles; especially, regarding the role of spatial versus relational proximity in the conditioning and evolution of regional innovation systems, or the role of spatial spillovers and the dynamics of regional knowledge creation processes. In this context, the dissertation also promotes establishing and further developing new concepts based on a more diverse and critical understanding of knowledge creation and innovation (such as e.g. Tödtling et al. 2021).

Second, the dissertation provides **novel empirical evidence on the interplay between spatial versus network structural effects on the constitution and dynamics of R&D collaboration networks.** In general, geographical barriers (i.e. geographical distance and country borders) constitute a significant hurdle for establishing network links across regions in the technologies investigated (Key Enabling Technologies). However, the results show that network structural components positively affect the creation of R&D collaboration links, and hence, can partly compensate for geographical barriers – regardless of the technology investigated. Specifically, disaggregated by relevant technological fields – in terms of geographical effects – in *Nanotechnology* and *Photonics*, R&D collaborations are more localised, while *Microelectronics* and *Advanced Manufacturing Technologies* are relatively global. Interestingly, *Photonics* holds a unique position with a hub-and-spoke structured network rather than an authority- and hub-structured network being the general case. This finding suggests a different mechanism of creating R&D collaboration links being in place.

Third, the dissertation provides **evidence for the versatile impacts of R&D collaboration networks for regional knowledge creation**. Specifically, disentangles differences in such network effects regarding (i) technology-specific knowledge creation as well as concerning different (ii) modes of knowledge creation (exploitation- and exploration-driven) and different types of knowledge output (quantity- and quality-oriented). The former shows that inter-regional network embeddedness is particularly important for knowledge creation in science-based technological fields (also acting as a substitute for lower levels of region-internal resources). At the same time, it is lower in more industrial and engineering-based fields where intra-regional knowledge interaction seem to matter more. In terms of different modes of knowledge creation, the dissertation finds a higher positive impact of networks on exploration-oriented (i.e. scientifically driven and dominantly basic research) than on exploitation-oriented (i.e. technologically driven and application-oriented) knowledge creation, when just the quantity of knowledge output is considered. In contrast, the effects from R&D collaboration networks tend to be higher for exploitative knowledge creation in terms of a higher quality of the knowledge produced.

Fourth, the dissertation produces original **evidence on the characteristics of spatial spillovers of the observed network effects**. Especially for the quantity of explorative knowledge creation, positive externalities from being located close to strongly connected regions arise. To a smaller extent, this effect is also observed for knowledge exploitation in the case of quality. Hence, in light of the catching-up processes of lagging regions, being co-located to regions that are well embedded in collaborative R&D project networks may serve as a stepping stone in becoming beneficiaries of inter-regional knowledge networks. However, despite the positive externalities, it still is important for each region itself to be embedded in R&D networks. Especially, to create knowledge of high quality, regions need to be themselves central and authoritative in R&D networks.

Fifth, the dissertation **contributes methodologically by developing an empirical ABM of knowledge creation in Europe** that strikes new paths for investigating regional knowledge creation mechanisms. In particular, the innovative application of various statistical methods and large-scale data sets to ensure a sound empirical foundation of the simulation model are considered substantial methodological contributions to current social simulation standards; specifically, the use of spatial microsimulation for agent initialisation, path-driven learning processes based on a network of technology classes, regression-based collaboration probabilities, sophisticated empirical calibration. Exemplary applications of the developed empirically founded agent-based simulation model demonstrate the robustness and empirical closeness in replicating phenomena and current scientific issues of interest. By showing the plausibility of the developed model by particularly establishing a strong empirical link, the dissertation takes a step towards gaining credibility of such simulation-type approaches – so far only rarely used in the context of regional science; especially as an additional way of modelling regional knowledge creation – apart from within a traditional Knowledge Production Framework.

Sixth, the dissertation contributes to **investigating the dynamic processes of inter-regional knowledge creation across Europe by taking advantage of the ABM approach**. In particular, simulation results show the geographically localised nature of knowledge creation, supporting a strongly right-skewed distribution of patent counts; indicating that most regions only exhibit relatively low levels of knowledge output, while a few regions have many patents. Moreover, findings allow identifying groups of specialist regions with industrial districts and sector-specific spillovers (also referred to a as Marshall externalities) and generalists that benefit from industry diversification (referred to as Jacobs externalities).

From a ***policy perspective*** – based on these main conclusions presented – tailored implications and recommendations for region-level and European policy- and decision-makers can be derived:

First, empirical evidence suggests the necessity to strengthen the regional network capability to participate in cross-regional R&D collaborations to help overcome geographical barriers. Findings show that regions that lack the ability to establish and maintain network links (e.g. due to missing research infrastructure and/or lack of scientific excellence) are handicapped in finding access to network hubs and hence are hampered in overcoming given geographical barriers. Making regions 'fit' to participate in pan-European R&D collaboration networks may be one of the most promising policy instruments to reduce such barriers.

Second, looking at the region-specific network effects, the results show that selected regions are particularly positively affected by R&D networks, in the form of collaborative EU FP projects. Cutting these knowledge flows would have a quite alarming effect regarding the potential consequences for the specific regional innovation system. This becomes apparent when looking at UK regions, which are found to be among the regions that benefit the most from EU funded networks, regardless of the mode of knowledge creation and type of knowledge output. This result points to a quite pessimistic conclusion in the context of the exit of the UK from the EU ('Brexit') and could be precedent for other regions.

Third, the empirical closeness of the presented agent-based model – as demonstrated in this dissertation – is the first step to assure also credibility in a policy context. The specific focus on technological, institutional and geographical aspects of knowledge creation in the model constitutes the proposed model to be a suitable tool to (ex-ante) design and shape research, technology and innovation (RTI) policy measures by conducting simulation experiments (e.g. w.r.t. regional specialisation policies).

Given specific dissertations' limitations, both the empirical findings and the dissertation's policy conclusions raise focal points for a future research agenda. *First* and foremost, the results presented in this dissertation rest on the choice of the R&D networks, which are, in the case of all four articles, publicly funded collaborative projects within the EU Framework Programme. Evidently, these projects follow specific rules, rationales and policy intentions that are reflected in the results, which limits their interpretation of this kind of R&D networks. Hence, ensuring robustness and establishing generalisability of the results by employing alternative collaboration-based networks (such as co-patent and co-publication networks) is a first entry point for future research.

Second, by applying data from the European FPs, the geographical focus is restricted to EU-27 regions (plus the UK and Norway). Nevertheless, the investigation of other geographical areas, such as China, would be of great interest to gain a comparative perspective of regional knowledge creation in innovation systems showing different development paths, different overall socio-economic characteristics and conditions, different policymaking approaches, and societal structures as a whole.

Third, emphasising the differences in network effects between knowledge quantity and quality in the dissertation, the quality dimension of knowledge creation remains mostly unexplored but has proven to be an interesting focus that bears great potential for future research. Especially, different aspects of knowledge quality could be highlighted in much more detail, e.g. taking a comparative perspective analysing multiple measures of knowledge quality, both for scientific as well as technology-driven knowledge creation.

Fourth, the results presented in first three articles are static, i.e. they constitute aggregates over a certain period and hence, represent the status-quo at a certain point in time. Taking a dynamic perspective on R&D networks is considered particularly fruitful in enhancing the future scientific discussion on regional knowledge creation. In this respect, spatial panel regression models are widely used, however, particularly an agent-based model, as developed in this dissertation, offers promising characteristics for investigating space-time relationships.

An additional future research agenda targets the further advancement and application of the developed empirical ABM of regional knowledge creation. The dissertation shows the suitability of simulation modelling in analysing spatial dynamics of inter-regional knowledge creation processes. Hence, on the one hand, future research is set to employ simulation techniques on currently discussed research issues in the fields of economic geography (e.g. disentangling local from global effects on regional knowledge creation, technological specialisation and geographical concentration tendencies among and within regions, regional path development). On the other hand, developing a quasi-standard or best-practice guideline for spatial simulation models integrating empirical data is highly desirable to strengthen the position of simulation modelling and more traditional methods investigating regional knowledge creation.

References

Acosta, M., D. Coronado and M. Á. Martínez (2012). Spatial differences in the quality of university patenting: Do regions matter? *Research Policy* 41(4): 692-703.

Acs, Z. J., L. Anselin and A. Varga (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy* 31(7): 1069-1085.

Ahuja, G. (2000) Collaboration networks, structural holes, and innovation: A longitudinal study" *Administrative Science Quarterly* 45(3): 425-455.

Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics letters* 80: 123–129.

Anderson P. and M. L. Tushman (1990). Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. *Administrative Science Quarterly* 35: 604–633.

Anselin, L., A. Varga and Z. Acs (1997). Local geographic spillovers between university research and high technology innovations. *Journal of urban economics* 42(3): 422-448.

Antonelli, C. (2011). The economic complexity of technological change: interactions, knowledge and path dependence. In Antonelli (ed) *Handbook on the economic complexity of technological change*, pp. 3–60.

Aschhoff, B., D. Crass, C. Grimpe, et al (2010). European Competitiveness in Key Enabling Technologies. Final Report. ZEW May 28th 2010.

Asheim, B. T. (1996). Industrial districts as 'learning regions': a condition for prosperity. *European planning studies* 4(4): 379-400.

Asheim, B. T. (2007). Differentiated knowledge bases and varieties of regional innovation systems. *Innovation: The European Journal of Social Science Research* 20: 223–241.

Asheim, B. T. and L. Coenen (2005). Knowledge bases and regional innovation systems: Comparing Nordic clusters. *Research Policy* 34(8): 1173-1190.

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

Asheim, B. T., H. L. Smith and C. Oughton (2011). Regional innovation systems: theory, empirics and policy. *Regional Studies* 45(7): 875-891.

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

Ausloos, M., H. Dawid and U. Merlone (2015). *Spatial interactions in agent-based modeling*. Complexity and Geographical Economics, Springer: 353-377.

Autant-Bernard, C. (2001). Science and knowledge flows: evidence from the French case. *Research Policy* 30(7): 1069-1078.

Autant-Bernard, C., J. Mairesse and N. Massard (2007a). Spatial knowledge diffusion through collaborative networks. *Papers in Regional Science* 86(3): 341-350.

Autant-Bernard, C., P. Billand, D. Frachisse and N. Massard (2007b). Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science* 86: 495–519.

Balland P-A, Boschma R, Frenken K (2015). Proximity and innovation: from statics to dynamics. *Regional Studies* 49: 907–920.

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

Barber, M. J. and T. Scherngell (2013). Is the European R&D network homogeneous? Distinguishing relevant network communities using graph theoretic and spatial interaction modelling approaches. *Regional Studies* 47(8): 1283-1298.

Barthélemy, M. (2011). Spatial networks. *Physics Reports* 499(1-3): 1-101.

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

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

Batty, M. (2012). *A generic framework for computational spatial modelling*. Agent-based models of geographical systems, Springer: 19-50.

Beaudry, C. and A. Schiffauerova (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *European Management Journal* 29(5): 362-376.

Beckenbach, F., R. Briegel and M. Daskalakis (2007). Behavioral foundation and agent based simulation of regional innovation dynamics, Citeseer.

Bell, G. G. and A. Zaheer (2007). Geography, networks, and knowledge flow. *Organization Science* 18(6): 955-972.

Bergé, L. R. (2017). Network proximity in the geography of research collaboration. *Papers in Regional Science* 96(4): 785-815.

Bergé, L. R., I. Wanzenböck and T. Scherngell (2017). Centrality of regions in R&D networks: A new measurement approach using the concept of bridging paths. *Regional Studies* 51(8): 1165-1178.

Blei, D.M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4): 77-84.

Boschma, R. (2004). Competitiveness of regions from an evolutionary perspective. *Regional Studies* 38(9): 1001-1014.

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

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

Boschma, R. and K. Frenken (2011). The emerging empirics of evolutionary economic geography. *Journal of economic geography* 11(2): 295-307.

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

Bozeman, B., P. Laredo and V. Mangematin (2007). Understanding the emergence and deployment of "nano" S&T. *Research Policy* 36: 807–812.

Breschi, S. and L. Cusmano (2004). Unveiling the texture of a European Research Area: Emergence of oligarchic networks under EU Framework Programmes. *International Journal of Technology Management* 27(8): 747-772.

Breschi, S. and C. Lenzi (2012). Net city: how co-invention networks shape inventive productivity in U.S. cities. Citeseer.

Breschi, S. and F. Lissoni (2001). Knowledge spillovers and local innovation systems: a critical survey. *Industrial and corporate change* 10(4): 975-1005.

Breschi, S., F. Malerba and L. Orsenigo (2000). Technological regimes and Schumpeterian patterns of innovation. *The Economic Journal* 110(463): 388-410.

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

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

Broekel, T. and H. Graf (2012). Public research intensity and the structure of German R&D networks: a comparison of 10 technologies. *Economics of Innovation and New Technology* 21(4): 345-372.

Broekel, T. and M. Hartog (2013). Explaining the structure of inter-organizational networks using exponential random graph models. *Industry and Innovation* 20(3): 277-295.

Cainelli, G., M. A. Maggioni, T. E. Uberti and A. De Felice (2015). The strength of strong ties: How co-authorship affect productivity of academic economists? *Scientometrics* 102(1): 673-699.

Cameron, C. and P. Trivedi (1998). *Models for count data*, Cambridge University Press, Cambridge.

Cantwell, J. and S. Iammarino (2005). Multinational corporations and European regional systems of innovation. Routledge, London, New York.

Capaldo, A. (2007). Network structure and innovation: The leveraging of a dual network as a distinctive relational capability. *Strategic Management Journal* 28(6): 585-608.

Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of geographical systems* 10(4): 317-344.

Coenen, L., J. Moodysson and B. T. Asheim (2004). Nodes, networks and proximities: on the knowledge dynamics of the Medicon Valley biotech cluster. *European Planning Studies* 12(7): 1003-1018.

Cohen, W. M. and D. A. Levinthal (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*: 128-152.

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and corporate change* 10(4): 945-974.

Cowan, R. and N. Zinovyeva (2013). University effects on regional innovation. *Research Policy* 42(3): 788-800.

Crooks, A., C. Castle and M. Batty (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems* 32(6): 417-430.

Dawid, H. (2006). Agent-based models of innovation and technological change. Handbook of computational economics 2: 1235-1272.

Dijkstra, L. (2009). Metropolitan regions in the EU. Regional Focus. Brussels, European Union - Regional Policy, 01/2009.

Doloreux, D. and S. Parto (2005). Regional innovation systems: Current discourse and unresolved issues. *Technology in Society* 27(2): 133-153.

Döring, T. and J. Schnellenbach (2006). What do we know about geographical knowledge spillovers and regional growth?: a survey of the literature. *Regional Studies* 40(03): 375-395.

Dorner, M. and D. Harhoff (2018). A novel technology-industry concordance table based on linked inventor-establishment data. *Research Policy* 47(4): 768-781.

Durieux, V. and P. A. Gevenois (2010). Bibliometric indicators: quality measurements of scientific publication. *Radiology* 255(2): 342-351.

EC (2008). Green Paper on Territorial Cohesion Turning Territorial Diverstiy into Strength. Brussels, European Commission. COM(2008) 616 final.

EC (2009). Preparing for our future: Developing a common strategy for key enabling technologies in the EU. Brussels, European Commission.

EC (2012). A European strategy for Key Enabling Technologies – A bridge to growth and jobs. Brussels, European Commission.

EC (2015a). KETs: Time to act. Final Report of the High-level Expert Group on Key Enabling Technologies. Brussels, European Commission.

EC (2015b) Key Enabling Technologies (KETs) Observatory. First annual report. DG Growth. European Union

EC (2015c). Regions in the European Union. Nomenclature of territorial units for statistics NUTS 2013/EU-28, Publication Office of the European Union Luxembourg.

Ejermo, O. (2009). Regional Innovation Measured by Patent Data—Does Quality Matter? Research Paper. *Industry and Innovation* 16(2): 141-165.

Evangelista, R., V. Meliciani and A. Vezzani (2018). Specialisation in key enabling technologies and regional growth in Europe. *Economics of Innovation and New Technology* 27: 273–289.

Feldman, M. P. (1994). *The geography of innovation*, Springer Science & Business Media.

Fischer, M. M. (2001). "Innovation, knowledge creation and systems of innovation." The Annals of Regional Science 35(2): 199-216.

Fischer, M. M. and D. Griffith (2008). Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the European Union. Journal of Regional Science 48: 969–989.

Fischer, M. M. and A. Varga (2003). Spatial knowledge spillovers and university research: Evidence from Austria. *The annals of regional science* 37(2): 303-322.

Fischer, M. M. and J. Wang (2011). *Spatial data analysis: models, methods and techniques*. Springer Science & Business Media, Berlin.

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

Foray, D., P. A. David and B. Hall (2009). Smart Specialisation – The Concept. *Knowledge Economists Policy Brief* 9(85): 100.

Fornahl, D., Broekel T, Boschma R (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: 395 – 418.

Freeman, C. (1982). The economics of industrial innovation. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.

Frenken, K., S. Hardeman and J. Hoekman (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics* 3(3): 222-232.

Frenken, K., F. Van Oort and T. Verburg (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies* 41(5): 685-697.

Garcia, R., R. Calantone and R. Levine (2003). The role of knowledge in resource allocation to exploration versus exploitation in technologically oriented organizations. *Decision Sciences* 34(2): 323-349.

Geiger, S. W. and M. Makri (2006). Exploration and exploitation innovation processes: The role of organizational slack in R & D intensive firms. *The Journal of High Technology Management Research* 17(1): 97-108.

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.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott and M. Trow (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London, Sage.

Gilbert, N., A. Pyka and P. Ahrweiler (2001). Innovation networks-a simulation approach. *Journal of Artificial Societies and Social Simulation* 4(3): 1-13.

Gilsing, V. and B. Nooteboom (2006). Exploration and exploitation in innovation systems: The case of pharmaceutical biotechnology. *Research policy* 35(1): 1-23.

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

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

Glaeser, E. L. and J. E. Kohlhase (2004). Cities, regions and the decline of transport costs. *Fifty Years of Regional Science,* Springer: 197-228.

Glückler, J. (2006). A relational assessment of international market entry in management consulting. *Journal of Economic Geography* 6(3): 369-393.

Glückler, J., E. Lazega and I. Hammer (2017). *Exploring the interaction of space and networks in the creation of knowledge: An introduction.* Knowledge and networks, Springer, Cham: 1-21.

Golub, G. H. and C. F. Van Loan (2013). *Matrix computations*, 4th edition. Johns Hopkins, Baltimore, MD.

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters* 107: 291–296.

Greunz, L. (2003). Geographically and technologically mediated knowledge spillovers between European regions. *The Annals of Regional Science* 37(4): 657-680.

Griffith, D. (1996). Spatial autocorrelation and eigenfunctions of the geographic weight matrix accompanying geo-referenced data. *The Canadian Geographer* 40: 341-367.

Griffith, D. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*, Springer, Berlin Heidelberg.

Griffith, D. and Y. Chun (2014). *Spatial autocorrelation and spatial filtering*. In: Fischer, M. M. and P. Nijkamp (eds) *Handbook of Regional Science Springer,* Berlin Heidelberg: 1477-1507.

Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *The bell journal of economics*: 92-116.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey. No. w3301, National Bureau of Economic Research.

Grossman, G. M. and E. Helpman (1994). Endogenous innovation in the theory of growth. *Journal of Economic Perspectives* 8(1): 23-44.

Guan, J. and N. Liu (2016). Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological field of nano-energy. *Research Policy* 45(1): 97-112.

Hagedoorn, J. and M. Cloodt (2003). Measuring innovative performance: is there an advantage in using multiple indicators? *Research Policy* 32(8): 1365-1379.

Hansen, M. T. and B. Løvås (2004). How do multinational companies leverage technological competencies? Moving from single to interdependent explanations. *Strategic Management Journal* 25(8-9): 801-822.

Harrison, B. (2007). Industrial Districts: Old Wine in New Bottles? (Volume 26, Number 5, 1992). *Regional Studies* 41(1): 107-121.

Hazır, C. S., J. LeSage and C. Autant-Bernard (2018). The role of R&D collaboration networks on regional knowledge creation: Evidence from information and communication technologies. *Papers in Regional Science* 97(3): 549-567.

Heimeriks, G. and R. Boschma (2014). The path- and place-dependent nature of scientific knowledge production in biotech 1986–2008. *Journal of Economic Geography* 14: 339–364.

Heller-Schuh, B., M. Barber, M. Züger and T. Scherngell (2015). Report on the content and technical structure of the EUPRO infrastructure, AIT Austrian Insitute of Technology, Vienna.

Heppenstall, A. J., A. T. Crooks, L. M. See and M. Batty (2011). *Agent-based models of geographical systems*, Springer Science & Business Media, Berlin Heidelberg New York.

Hoekman, J., K. Frenken and F. Van Oort (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science* 43(3): 721-738.

Hoekman, J., K. Frenken and R. J. Tijssen (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39(5): 662-673.

Howells, J. R. L. (2002). Tacit Knowledge, Innovation and Economic Geography. *Urban Studies* 39: 871–884.

Huggins, R. and P. Thompson (2014). A network-based view of regional growth. *Journal of Economic Geography* 14(3): 511-545.

Jacobs, J. (1969). *The economy of cities*. Random House, New York.

Jaffe, A. B. (1989). Real effects of academic research. *The American Economic Review*: 957-970.

Jaffe, A. B. and M. Trajtenberg (2002). Patents, citations, and innovations: A window on the knowledge economy, MIT press, Boston, MA.

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

Jensen, M. B., B. Johnson, E. Lorenz and B. Å. Lundvall (2007). Forms of knowledge and modes of innovation. *Research Policy* 36: 680–693.

Karaca-Mandic, P., E. C. Norton and B. Dowd (2012). Interaction terms in nonlinear models. *Health Services Research* 47: 255–274.

Kiesling, E., M. Günther, C. Stummer and L. M. Wakolbinger (2012). Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research* 20(2): 183-230.

Kirat, T. and Y. Lung (1999). Innovation and proximity: territories as loci of collective learning processes. *European Urban and Regional Studies* 6(1): 27-38.

Kleinberg, J. M. (1999). Hubs, authorities, and communities. *ACM computing surveys (CSUR)* 31(4es): 5-es.

Lagendijk, A. (2001). Scaling knowledge production: How significant is the region? In Fischer, M. M. and J. Fröhlich (eds) *Knowledge, complexity and innovation systems*. Springer, Berlin, Heidelberg: 79-100.

Lanjouw, J. O. and M. Schankerman (2004). Patent quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal* 114(495): 441-465.

Lata, R., T. Scherngell and T. Brenner (2015). Integration Processes in European Research and Development: A Comparative Spatial Interaction Approach Using Project Based Research and Development Networks, Co-Patent Networks and Co-Publication Networks. *Geographical Analysis* 47(4): 349-375.

LeSage, J. P. and M. M. Fischer (2016). *Spatial regression-based model specifications for exogenous and endogenous spatial interaction*. In: Patuelli, R. and G. Arbia (eds) Spatial Econometric Interaction Modelling Springer International Publishing, Switzerland: 15-36.

LeSage, J., M. M. Fischer and T. Scherngell (2007). Knowledge spillovers across Europe. Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science* 86: 393–421

LeSage, J. P. and R. K. Pace (2009). *Introduction to Spatial Econometrics (Statistics, textbooks and monographs)*. CRC Press, Boca Raton.

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

Leydesdorff, L., F. Alkemade, G. Heimeriks and R. Hoekstra (2015). Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on “photovoltaic cells”. *Scientometrics* 102: 629–651.

Li, Y., W. Vanhaverbeke and W. Schoenmakers (2008). Exploration and exploitation in innovation: Reframing the interpretation. *Creativity and Innovation Management* 17(2): 107-126.

Lipsey, R. G., K. I. Carlaw and C. T. Bekar (2005). Economic transformations: general purpose technologies and long-term economic growth. Oxford University Press, Oxford

Long, S. J. and J. Freese (2006). *Regression models for categorical dependent variables using Stata*. Stata press, Texas.

Lovelace, R. and D. Ballas (2012). *Modelling commuter patterns: a spatial microsimulation approach for combining regional and micro level data.* ERSA conference papers, European Regional Science Association.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics* 22(1): 3-42.

Macal, C. M. and M. J. North (2010). Tutorial on agent-based modelling and simulation. *Journal of simulation* 4(3): 151-162.

Maggioni, M. A., M. Nosvelli and T. E. Uberti (2007). Space versus networks in the geography of innovation: A European analysis. *Papers in Regional Science* 86(3): 471-493.

Maggioni, M. A. and T. E. Uberti (2011). Networks and geography in the economics of knowledge flows. *Quality & Quantity* 45(5): 1031-1051.

Malerba, F. (2002). Sectoral systems of innovation and production. *Research Policy* 31(2): 247-264.

Malerba, F. and L. Orsenigo (1997). Technological regimes and sectoral patterns of innovative activities. *Industrial and Corporate Change* 6(1): 83-118.

Malerba, F. and L. Orsenigo (2000). Knowledge, innovative activities and industrial evolution. *Industrial and Corporate Change* 9(2): 289-314.

Malmberg, A., Ö. Sölvell and I. Zander (1996). Spatial clustering, local accumulation of knowledge and firm competitiveness. *Geografiska Annaler: Series B, Human Geography* 78(2): 85-97.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science* 2: 71–87.

Marek, P., M. Titze, C. Fuhrmeister and U. Blum (2017). R&D collaborations and the role of proximity. *Regional Studies* 51(12): 1761-1773.

Markusen, A. (1996). Sticky places in slippery space: a typology of industrial districts. *Economic Geography* 72(3): 293-313.

Marrocu, E., R. Paci and S. Usai (2013). Productivity growth in the old and new Europe: the role of agglomeration externalities. *Journal of Regional Science* 53(3): 418-442.

Marshall, A. (1890). *Principles of economies*. Macmillan, London.

Marshall, A. (1920). *Industry and trade*. Macmillan, London.

Martin, R. and J. Moodysson (2013). Comparing knowledge bases: on the geography and organization of knowledge sourcing in the regional innovation system of Scania, Sweden. *European Urban and Regional Studies* 20(2): 170-187.

März, S., M. Friedrich-Nishio and H. Grupp (2006). Knowledge transfer in an innovation simulation model. *Technological Forecasting and Social Change* 73(2): 138-152.

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

Mattes, J. (2012). Dimensions of proximity and knowledge bases: Innovation between spatial and non-spatial factors. *Regional Studies* 46(8): 1085-1099.

Maurseth, P. B. and B. Verspagen (2002). Knowledge spillovers in Europe: a patent citations analysis. *Scandinavian Journal of Economics* 104(4): 531-545.

Maynard, D., A. Funk and B. Lepori (2017). Towards an Infrastructure for Understanding and Interlinking Knowledge Co-Creation in European Research. Lecture Notes in Computer Science (Proceedings of ESWC 2017 Workshop on Scientometrics), Springer Verlag.

McCann P, Ortega-Argilés R (2015) Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy. *Regional Studies* 49:1291–1302

McKay, M. D., R. J. Beckman and W. J. Conover (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42(1): 55-61.

McPherson, M., L. Smith-Lovin and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1): 415-444.

Miguélez, E. and R. Moreno (2013). Research networks and inventors' mobility as drivers of innovation: evidence from Europe. *Regional Studies* 47(10): 1668-1685.

Molina-Morales, F. X. and M. T. Martínez-Fernández (2009). Too much love in the neighborhood can hurt: How an excess of intensity and trust in relationships may produce negative effects on firms. *Strategic Management Journal* 30(9): 1013-1023.

Montresor S, Quatraro F (2017). Regional Branching and Key Enabling Technologies: Evidence from European Patent Data. *Economic Geography* 93: 367–396.

Moodysson J (2008). Principles and Practices of Knowledge Creation: On the Organization of “Buzz” and “Pipelines” in Life Science Communities. *Economic Geography* 84: 449–469.

Moreno, R., R. Paci and S. Usai (2005). Geographical and sectoral clusters of innovation in Europe. *The Annals of Regional Science* 39(4): 715-739.

Morescalchi, A., F. Pammolli, O. Penner, A. M. Petersen and M. Riccaboni (2015). The evolution of networks of innovators within and across borders: Evidence from patent data. *Research Policy* 44(3): 651-668.

Morgan, K. (2007). The learning region: institutions, innovation and regional renewal. *Regional Studies* 41(1): 147-S159.

Moulaert, F. and F. Sekia (2003). Territorial innovation models: a critical survey. *Regional Studies* 37(3): 289-302.

Mowery, D. C., J. E. Oxley and B. S. Silverman (1996). Strategic alliances and interfirm knowledge transfer. *Strategic Management Journal* 17(2): 77-91.

Mowery, D. C. and A. A. Ziedonis (2002). Academic patent quality and quantity before and after the Bayh–Dole act in the United States. *Research Policy* 31(3): 399-418.

Mueller, M., K. Bogner, T. Buchmann and M. Kudic (2017). The effect of structural disparities on knowledge diffusion in networks: an agent-based simulation model. *Journal of Economic Interaction and Coordination* 12(3): 613-634.

Nagaoka, S., K. Motohashi and A. Goto (2010). Patent statistics as an innovation indicator. In Hall, B. H. and N. Rosenberg (eds) *Handbook of the Economics of Innovation*, *Vol 2.* Elsevier, Amsterdam. 1083-1127.

Nelson, R. R. and S. G. Winter (1982). *An evolutionary theory of economic change*. Harvard University Press, Cambridge, MA.

Neuländtner, M. and T. Scherngell (2020). Geographical or relational: What drives technology-specific R&D collaboration networks? *The Annals of Regional Science* 65: 743-773.

Neves, P. C. and T. N. Sequeira (2018). Spillovers in the production of knowledge: A meta-regression analysis. *Research Policy* 47(4): 750-767.

Nikolic, I., K. H. Van Dam and J. Kasmire (2013). Practice. In K. H. Van Dam, I. Nikolic and Z. Lukszo_(eds) *Agent-based modelling of socio-technical systems.* Dordrecht, Springer Netherlands. 9: 73-140.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science* 5(1): 14-37.

Nooteboom, B. (1999). Innovation, learning and industrial organisation. *Cambridge Journal of Economics* 23(2): 127-150.

Nowotny, H., P. Scott and M. Gibbons (2003). Introduction:Mode 2'Revisited: The New Production of Knowledge. *Minerva* 41(3): 179-194.

Ó hUallacháin, B. and T. F. Leslie (2007). Rethinking the regional knowledge production function. *Journal of Economic Geography* 7(6): 737-752.

Paci, R., E. Marrocu and S. Usai (2014). The complementary effects of proximity dimensions on knowledge spillovers. *Spatial Economic Analysis* 9(1): 9-30.

Paci, R. and S. Usai (2000). Technological enclaves and industrial districts: an analysis of the regional distribution of innovative activity in Europe. *Regional Studies* 34(2): 97-114.

Paier, M., M. Dünser, T. Scherngell and S. Martin (2017). Knowledge creation and research policy in science-based industries: an empirical agent-based model. In Vermeulen B. and M. Paier *Innovation Networks for Regional Development*. Springer, Switzerland. 153-183.

Patuelli, R., D. Griffith, M. Tiefelsdorf and P. Nijkamp (2011). Spatial Filtering and Eigenvector Stability: Space-Time Models for German Unemployment Data. *International Regional Science Review* 34(2): 253-280.

Pavitt, K. (1984). Sectoral patterns of technical change: towards a taxonomy and a theory. *Research Policy* 13(6): 343-373.

Pavitt, K. (2005) Innovation processes. In Fagerberg J, Mowery DC, Nelson RR (eds) *The Oxford Handbook of Innovation*. Oxford University Press, Oxford. 86–114.

Phene, A., K. Fladmoe-Lindquist and L. Marsh (2006). Breakthrough innovations in the US biotechnology industry: the effects of technological space and geographic origin. *Strategic Management Journal* 27(4): 369-388.

Ponds, R., F. Van Oort and K. Frenken (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science* 86(3): 423-443.

Ponds, R., F. v. Oort and K. Frenken (2009). Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach. *Journal of Economic Geography* 10(2): 231-255.

Popadiuk, S. and C. W. Choo (2006). Innovation and knowledge creation: How are these concepts related? *International Journal of Information Management* 26(4): 302-312.

Porter, M. E. (1998). *Clusters and the new economics of competition*, Harvard Business Review, Boston.

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

Pyka, A., M. Kudic and M. Müller (2019). Systemic interventions in regional innovation systems: entrepreneurship, knowledge accumulation and regional innovation. *Regional Studies* 53(9): 1321-1332.

Qiu R, Cantwell J (2018). The international geography of general purpose technologies (GPTs) and internationalisation of corporate technological innovation. *Industry and Innovation* 25: 1–24.

Rallet, A. and A. Torre (1998). On geography and technology: The case of proximity relations in localized innovation networks. In Steiner, M. *Clusters and regional specialisation: On geography, technology and networks*: Pion London.

Rodríguez-Pose, A. and R. Crescenzi (2008). Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe. *Regional Studies* 42(1): 51-67.

Rodrik, D. (2014). Green industrial policy. *Oxford Review of Economic Policy* 30:469–491.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98(5, Part 2): S71-S102.

Rosenkopf, L. and A. Nerkar (2001). Beyond local search: boundary-spanning, exploration, and impact in the optical disk industry. *Strategic Management Journal* 22(4): 287-306.

Rotolo, D., D. Hicks and B. R. Martin (2015). What is an emerging technology? *Research Policy* 44:1827–1843.

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

Roy, J. R. and J.-C. Thill (2003). Spatial interaction modelling. *Papers in Regional Science* 83(1): 339-361.

Salman, N. and A. L. Saives (2005). Indirect networks: an intangible resource for biotechnology innovation. *R&D Management* 35(2): 203-215.

Savin, I. and A. Egbetokun (2016). Emergence of innovation networks from R&D cooperation with endogenous absorptive capacity. *Journal of Economic Dynamics and Control* 64: 82-103.

Scherngell, T. (2013). The Geography of Networks and R&D Collaborations. Springer, Berlin Heidelberg New York.

Scherngell, T. (2019). The geography of R&D collaboration networks. In Fischer, M. M. and P. Nijkamp *Handbook of Regional Science*. Springer, Berlin.

Scherngell, T. and M. J. Barber (2009). Spatial interaction modelling of cross-region R&D collaborations: Empirical evidence from the 5th EU framework programme. *Papers in Regional Science* 88(3): 531-546.

Scherngell, T. and R. Lata (2013). Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. *Papers in Regional Science* 92(3): 555-577.

Schmoch, U. (2008). Concept of a technology classification for country comparisons. Final report to the world intellectual property organisation (wipo), WIPO.

Schwartz, M., F. Peglow, M. Fritsch and J. Günther (2012). What drives innovation output from subsidized R&D cooperation?—Project-level evidence from Germany. *Technovation* 32(6): 358-369.

Sebestyén, T. and A. Varga (2013). Research productivity and the quality of interregional knowledge networks. *The Annals of Regional Science* 51:155–189.

Sebestyén, T. and A. Varga (2019). Knowledge networks in regional development: an agent-based model and its application. *Regional Studies* 53(9): 1333-1343.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science* 51(5): 756-770.

Singh, J. (2008). Distributed R&D, cross-regional knowledge integration and quality of innovative output. *Research Policy* 37(1): 77-96.

Sorenson, O., J. W. Rivkin and L. Fleming (2006). Complexity, networks and knowledge flow. *Research Policy* 35: 994–1017.

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

Squicciarini, M., H. Dernis and C. Criscuolo (2013). Measuring patent quality. OECD No. 2013/03, Paris.

Tang, L. and P. Shapira (2012). Effects of international collaboration and knowledge moderation on China's nanotechnology research impacts. *Journal of Technology Management in China* 7(1): 94-110.

Tamada, S., Y. Naito, F. Kodama et al (2006). Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics* 68: 289–302.

Ter Wal, A. L. J. (2013). The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. *Journal of Economic Geography* 47(5): 651-668.

Thiele, J. C., W. Kurth and V. Grimm (2014). Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social Simulation* 17(3): 11.

Thiriot, S. and J.-D. Kant (2008). Using Associative Networks to Represent Adopters Beliefs in a Multiagent Model of Innovation Diffusion. *Advances in Complex Systems* 11(02): 261-272.

Tiefelsdorf, M. and B. Boots (1995). The exact distribution of Moran's I. *Environment and Planning A* 27(6): 985-999.

Tiwana, A. (2008). Do bridging ties complement strong ties? An empirical examination of alliance ambidexterity. *Strategic Management Journal* 29(3): 251-272.

Tödtling, F. and M. Trippl (2005). One size fits all?: Towards a differentiated regional innovation policy approach. *Research Policy* 34(8): 1203-1219.

Tödtling, F., P. Lehner and M. Trippl (2006). Innovation in knowledge intensive industries: The nature and geography of knowledge links. *European Planning Studies* 14(8): 1035-1058.

Tödtling, F., M. Trippl and V. Desch (2021). New directions for RIS studies and policies in the face of grand societal challenges. *GEIST – Geography of Innovation and Sustainability Transitions, 2021*(01), GEIST Working Paper series.

Torre, A. and A. Rallet (2005). Proximity and localization. *Regional Studies* 39(1): 47-59.

Trippl, M., F. Tödtling and L. Lengauer (2009). Knowledge sourcing beyond buzz and pipelines: evidence from the Vienna software sector. *Economic Geography* 85(4): 443-462.

Tsai, T. and J. Gill (2013). Interactions in generalized linear models: theoretical issues and an application to personal vote-earning attributes. *Social Sciences* 2: 91–113.

Uzzi, B. and R. Lancaster (2003). Relational embeddedness and learning: The case of bank loan managers and their clients. *Management Science* 49(4): 383-399.

Valente, T. W., K. Coronges, C. Lakon and E. Costenbader (2008). How correlated are network centrality measures? *Connections* 28: 16.

Van der Panne, G. (2004). Agglomeration externalities: Marshall versus Jacobs. *Journal of Evolutionary Economics* 14(5): 593-604.

van Raan, A. F. (2017). Patent citations analysis and its value in research evaluation: A review and a new approach to map technology-relevant research. *Journal of Data and Information Science* 2(1): 13-50.

Varga, A., D. Pontikakis and G. Chorafakis (2014). Metropolitan Edison and cosmopolitan Pasteur? Agglomeration and interregional research network effects on European R&D productivity. *Journal of Economic Geography* 14: 229–263.

Varga, A. and T. Sebestyén (2017). Does EU Framework Program participation affect regional innovation? The differentiating role of economic development. *International Regional Science Review* 40(4): 405-439.

Vermeulen, B. and A. Pyka (2018). The role of network topology and the spatial distribution and structure of knowledge in regional innovation policy: a calibrated agent-based model study. *Computational Economics* 52(3): 773-808.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics* 10(2): 365-391.

Wang, Z., Z. Yao, G. Gu, F. Hu and X. Dai (2014). Multi-agent-based simulation on technology innovation-diffusion in China. *Papers in Regional Science* 93(2): 385-408.

Wanzenböck, I., T. Scherngell and T. Brenner (2014). Embeddedness of regions in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications. *The Annals of Regional Science* 53(2): 337-368.

Wanzenböck, I., T. Scherngell T and R. Lata (2015) Embeddedness of European Regions in European Union-Funded Research and Development (R&D) Networks: A Spatial Econometric Perspective. Regional Studies 49: 1685–1705.

Wanzenböck, I. and P. Piribauer (2018). R&D networks and regional knowledge production in Europe: Evidence from a space-time model. *Papers in Regional Science 97: 1-24*.

Wanzenböck, I., M. Neuländtner and T. Scherngell (2020). Impacts of EU funded R&D networks on the generation of key enabling technologies: Empirical evidence from a regional perspective. *Papers in Regional Science* 99(1): 3-24.

Wasserman, S. and K. Faust (1994). *Social network analysis: Methods and applications*. Cambridge University Press, Cambridge.

Whittington, K. B., J. Owen-Smith and W. W. Powell (2009). Networks, propinquity, and innovation in knowledge-intensive industries. *Administrative Science Quarterly* 54(1): 90-122.

Wilden, R., J. Hohberger, T. M. Devinney and D. Lavie (2018). Revisiting James March (1991): whither exploration and exploitation? *Strategic Organization* 16(3): 352-369.

Winter, S. G. (1987). Knowledge and competence as strategic assets. In David J, and J. Teece (eds) *The Competitive Challenge: Strategies for Industrial Renewal.* Ballinger Publishing, Cambridge, MA.

Winter, S. G. (1997). Knowledge and competence as strategic assets. In Klein, A. (ed) *The strategic management of intellectual capital.* Elsevier, Amsterdam: 165-187.

Wolfe, D. A. and M. S. Gertler (2004). Clusters from the inside and out: local dynamics and global linkages. *Urban Studies* 41(5-6): 1071-1093.

Yli-Renko, H., E. Autio and H. J. Sapienza (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal* 22(6-7): 587-613.

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

Zitt, M., S. Ramanana-Rahary, E. Bassecoulard and F. Laville (2003). Potential science-technology spillovers in regions: An insight on geographic co-location of knowledge activities in the EU. *Scientometrics* 57(2): 295-320.