The Annals of Regional Science (2020) 65:743–773
https://doi.org/10.1007/s00168-020-01002-5

ORIGINAL PAPER

Geographical or relational: What drives technology-specific R&D collaboration networks?

Martina Neuländtner[1] · Thomas Scherngell[1]

Received: 2 August 2019 / Accepted: 18 June 2020 / Published online: 8 July 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, in particular, 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 geographical separation and network structural effects influence technology-specific R&D collaborations between European regions. The results underline both the significance of geographical barriers and network structural effects and confirm that specific network connectivity is able to 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.

JEL Classification O30 · R10 · C10

1 Introduction

Collaborative Research and Development (R&D) activities between firms, universities and research organizations are generally recognized 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 characterizing such collaborative research endeavours and has become a fascinating research domain in manifold aspects (see Scherngell 2013 for an overview). With knowledge creation being 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

✉ Martina Neuländtner
martina.neulaendtner@ait.ac.at

Thomas Scherngell
thomas.scherngell@ait.ac.at

[1] AIT Austrian Institute of Technology GmbH, Giefinggasse 4, 1210 Vienna, Austria

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 the creation of knowledge (Glückler et al. 2017). In this context, not only the general importance of networks is stressed, but also the role of different network structures and topologies. On an organizational level, it is well recognized in the literature that the position of single nodes, e.g. representing firms, and the network structure as a whole have signficant impact on the creation of new knowledge and its diffusion (e.g. Ahuja 2000; Zaheer and Bell 2005; Giuliani 2007), also from a spatial perspective understanding regions as network nodes (e.g. Whittington et al. 2009; Maggioni and Uberti 2011). Studies on the geography of R&D collaboration networks, focusing on the identification and estimation of determinants affecting structures and dynamics, 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 quite 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). While there are a number of studies that separately address geographical or network structural factors when analysing cross-region R&D collaboration networks (see Scherngell 2019 for an overview), there are only very few and 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 differing role of geographical versus relational effects when explaining the constitution and dynamics of R&D collaboration networks. We attempt this in one integrated modelling framework and for a larger geographical area, while at the same time 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 estimate these effects, we employ a negative binomial spatial interaction modelling approach at the regional level, accounting for spatial autocorrelation of the interactions. The R&D collaboration network under consideration is a network of organizations that collaborate in projects funded by the EU framework programme (FP). This network is partitioned into different technological domains and aggregated from the organizational to the regional level, using a set of 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 make use of 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 are not able to capture these technologies. With our focus on networks of KETs, we propose—in contrast to previous research—a finer-grained and policy-relevant perspective when identifying determinants of R&D collaboration networks.

The study departs from related previous research in at least three major 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), but also on the probability for 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 urbanization effects from other effects (e.g. geographical proximity or country borders) in a more robust way.

The remainder of the study is organized as follows. The following section reviews the main elements of the theoretical and empirical debate on determinants of R&D collaboration networks, specifically highlighting the relevance of the focus on geographical and relational characteristics in different technologies. Section 3 shifts specific attention to the role of technological heterogeneities in such networks that have been largely neglected so far in the empirical literature. Section 4 describes the spatial interaction approach used to identify determinants of collaboration, followed by Sect. 5 that sets out the empirical setting. Section 6 discusses the estimation results, before Sect. 7 closes with a summary and some ideas for future research.

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 wide agreement that both 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. 2007; Bathelt and Glückler 2003). Recently, this research interest is mainly motivated by the seminal work by Bathelt and Glückler (2003) suggesting 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 organizational, 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, there exists a large and diverse body of the empirical literature on determinants of R&D collaboration networks in different technological fields and different geographical areas. Despite their differences, spatial proximity turns out to be an important factor for the constitution of R&D collaboration in all these studies, also in times of increasing globalization and new information and 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 the exchange of more tacit knowledge elements. Accordingly, face-to-face interaction in inter-organizational 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 for transmitting uncodified, tacit knowledge in geographical space, complex knowledge is more immobile in geographical space, and accordingly, network effects may become more important 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 important role of geographical space as driver for network dynamics.

However, apart from being geographically close to create and exchange complex and tacit knowledge, being part of a same professional community—such as a research network—may facilitate knowledge creation and transfer; i.e. '*organizational proximity*' (Kirat and Lung 1999; Boschma 2005) or '*organized proximity*' (Torre and Rallet 2005). This type of relational proximity is characterized by common knowledge and knowledge bases (e.g. same scientific community) and by interacting actors that enable interaction and accelerate knowledge creation (Boschma 2005; Torre and Rallet 2005). The EU framework programmes (FP), for instance, feature such kind of 'organized proximity', where firms, universities and research organizations located in various European regions collaborate in all kinds of topics aiming for excellence throughout the European Research Area (ERA) (see Breschi

[1] In the 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).

and Cusmano 2004, followed by many others). Moreover, fostering inter-regional collaboration by means of funding opportunities, such as the EU FPs, have facilitated long-distance collaborations (Scherngell and Lata 2013), highlighting the potential of networks as an organizational arrangement to overcome geographical barriers.

In this vein, a region's position in global R&D networks has been increasingly considered as important in recent years, in particular, 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, in comparison with geographical ones (Glückler et al. 2017).

Inspired from 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 hence, their network embeddedness in terms of the number of collaboration links.[2] However, not only the quantity of a region's collaboration arrangements matters, but also their quality indicating, on the one hand, access to reliable information itself, and, on the other hand, linkages to other partnering organizations 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. in terms of their number and quality of collaboration links. In social network analysis, this is usually referred to as homophily, i.e. social actors are more likely to interlink when they have similar attributes (McPherson et al. 2001). From a regional network perspective, such mechanisms may come into play when considering the amount of critical R&D infrastructure of regions. Global R&D players are usually located in large and advanced regions, such as metropolitan regions, and may be more likely to collaborate with R&D actors of similar size and impact, located themselves in metropolitan regions. Similarly, the opposite may occur with small R&D actors located in less advanced regions.

The importance of relational or network structural effects on R&D collaborations has been partly addressed in only a few empirical studies up to now, such as for social proximity (Autant-Bernard et al. 2007), 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 limited studies addressing 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

[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.

fostering collaboration across geographical distances by manifesting sustainable network links. Against the background of these theoretical and empirical debates, we pose a 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 hampering 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.

3 Technological heterogeneities in R&D collaboration networks

While technological heterogeneities in terms of differences in knowledge bases and knowledge creation processes have been subjected 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 differing knowledge domains, originally referred to as technological regimes[3]—has been stressed to explain differences across sectors in patterns of innovation and, accordingly, can be considered as highly relevant for R&D collaboration networks as major input for innovation. Malerba (2002) identifies three key dimensions of knowledge related to the notion of technological regimes: degree of *accessibility* (i.e. opportunities of gaining knowledge, e.g. by means of cross-regional network 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 specific properties of the knowledge base, 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

[3] The term '*technological regime*' originates in the work by Nelson and Winter (1982) and characterizes the knowledge environment in which organizations within the same industry are argued to be subject to same technological and knowledge conditions, such as 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 1987).

collaboration networks, as well as varying underlying mechanisms that drive their constitution and dynamics. This motivates our third 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 rather 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), also disregarding 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 give an explanation on 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) emphasize 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) categorizes 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 (1984) derives four types of sectors: *supply-dominated* (e.g. clothing, furniture), *scale-intensive* (e.g. food, cement), *specialized supplier* (e.g. engineering, software and instruments) and *science-based producers* (e.g. chemical industry, biotechnology and electronics). Derived from this discussion, we pose an additional hypothesis:

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 observe 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 as crucial for the development of the EU towards a sustainable, knowledge-based economy (EC 2009, 2012).[4] These are N*anotechnology*, *Microelectronics*, *Photonics*, *Advanced Materials (AM)*, *Advanced Manufacturing Technology (AMT)* and *Industrial Biotechnology* (EC 2009).[5]

Despite the common specificities of KETs (by which they identify as *'key enabling'*), we argue that these distinct technologies differ with respect to their geographical and network impacts on inter-regional R&D collaboration. Note in this context that KETs are empirically found to be strongly spatially concentrated on certain 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).

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 specialization in certain KETs suggest 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 (1984) taxonomy, they can be characterized as either *specialized 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 knowledge 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

[4] 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, 2015).

[5] KETs are understood as generic technologies that are characterized by relatively rapid pervasiveness and growth, high knowledge and R&D intensity, and highly skilled employment (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 nationwide and global state-of-the-art knowledge. Moreover, KETs are claimed to affect the regional capacity of developing new technological specializations (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).

modelling framework (see Sect. 4) for a larger geographical area, namely the whole European territory (see Sect. 5).

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, whereas the interactions are a function of the attributes of the locations of origin, the attributes of the locations destination and the friction (*separation*) between the respective origin and destination. The purpose of such models is to explain the relationships between interaction frequencies of two spatial entities and their (relational) properties (Roy and Thill 2003). In our case, the spatial interactions under consideration are R&D collaboration networks between regions. The general form of the model can be written as

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

where $\mu_{ij} = E\left(Y_{ij}\right)$ 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 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: (1) *origin-specific* factors characterizing the ability of the origins to generate R&D network links, (2) *destination-specific* factors indicating the attractiveness of destinations and (3) *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

$$\mu_{ij} = O_i D_j S_{ij} \quad \text{with} \quad i,j = 1, \ldots, N \tag{2}$$

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 a number of separation variables) exponential functional forms for the separation function. We follow these lines and define

$$O_i = O\left(o_i, \alpha_1\right) = o_i^{\alpha_1} \tag{3}$$

$$D_j = D\left(d_j, \alpha_2\right) = d_j^{\alpha_2} \tag{4}$$

$$S_{ij} = \exp\left[\sum_{k=1}^{K} \beta_k s_{ij}^{(k)}\right] \tag{5}$$

Here, o_i and d_j are measured in terms of variables controlling for the mass in the origin and the destination, respectively. In context of R&D networks, these are often captured by the number of firms or researching organizations in a region. Accordingly, α_1 and α_2 are scalar parameters to be estimated, so that the product of the functions O_iD_j can be simply interpreted as the number of cross-region R&D collaborations which are possible. Core of the spatial interaction model is the separation function as defined by Eq. (5), with K ($k=1, \ldots, K$) 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 in this study takes the specific form of a spatially filtered, negative binomial spatial interaction model (see Scherngell and Lata 2013 in a similar context).[6] The main motivation for this is given by the true integer nature and distributional assumptions on the dependent variable, namely 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 amount 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).

In comparison with the standard Poisson specification that assumes equidispersion (i.e. conditional mean equals the conditional variance), the negative binomial model explicitly corrects for overdispersion,[7] by adding a dispersion parameter θ. Hence, the negative binomial spatial interaction model takes the form (Long and Freese 2006)

$$\Pr\left(Y_{ij} = y_{ij} \middle| \mu_{ij}, \gamma\right) = \frac{\Gamma\left(y_{ij} + \theta\right)}{\Gamma\left(y_{ij} + 1\right)\Gamma(\theta)} \left(\frac{\theta}{\theta + \mu_{ij}}\right)^{\theta} \left(\frac{\mu_{ij}}{\theta + \mu_{ij}}\right)^{y_{ij}} \tag{6}$$

where $\mu_{ij} = E\left[y_{ij} \middle| O_i, D_j, S_{ij}\right] = \exp\left[O_i\left(\alpha_1\right)D_j\left(\alpha_2\right)S_{ij}(\beta)\right]$ 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).

[6] Although the data used have 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).

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

To take the spatial dependence of flows into account, spatial filtering using eigenvectors (ESF) is employed[8] (see 'Appendix 1' 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 of 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 (6), the final empirical model to be estimated is specified by setting

$$\mu_{ij} = \exp(\alpha_0 + \alpha_1 \ln\left(o_i\right) + \alpha_2 \ln\left(d_j\right) + \sum_{k=1}^{K} \beta_k s_{ij}^{(k)} + \sum_{q=1}^{Q} \phi_q E_q + \sum_{r=1}^{R} \varphi_r E_r + \xi_{ij} \quad (7)$$

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. Explanatory variables enter the regression in 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).

5 Data and variables

The main interest of this study is to estimate determinants of technology-specific R&D collaboration networks, with a special focus on spatial separation and network structural effects. The geographical coverage comprises the current 27 EU member states (excluding Malta and Cyprus) plus UK, Switzerland and Norway, corresponding to a set of 505 regions. 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,[9] the remaining non-metropolitan regions are either original NUTS 2 regions, or adapted NUTS 2 regions with

[8] 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).

[9] 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.

respective NUTS 3 regions—belonging to a metropolitan region—removed (see Fig. 1 in 'Appendix 2' for map of metropolitan regions).[10]

5.1 Dependent variable

As dependent variable EU-funded KET R&D collaboration links are used (see Table B1 in 'Appendix 2' for some descriptive statistics). Data are extracted from the EUPRO database[11] comprising systematic information on collaborative research projects of FP1-FP7 as well as Horizon 2020 (until 2016), including information on respective participating organizations, e.g. name, type and their geographical location in the form of organization 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, that is subject to certain governance rules (e.g. each project must have partners from at least two different countries). However, these rules are by far less relevant for the formation of collaboration than their behaviour that is driven by strategic, technological, geographical, cultural and institutional conditions (see Scherngell and Barber 2009).

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 projects.[12] 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 containing the inter- and intra-regional collaboration activities of all region pairs. Figure 2 in 'Appendix 2' illustrates the spatial distribution of the networks revealing the Paris region as dominating hub in all networks, showing the characteristic star-shaped backbone structure. Nevertheless, the networks differ with respect to density, variance in number of collaborations, spatial scales and importance of certain regions (e.g. London in the case of *Nanotechnology* and *Biotechnology*; see Table 3 in 'Appendix 2').

5.2 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 solely specified as the number of organizations participating in

[10] Although the NUTS-2 level perspective is widely used in the 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, for example, the NUTS classification that separates these two geographical spaces in most cases.

[11] 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 geolocalization, standardization and integration with other datasets.

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

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 Fig. 3 in 'Appendix 2' for spatial distribution). For the separation variables, we distinguish between (1) spatial separation variables and (2) network structural separation variables (see 'Appendix 2' for Table 4 with descriptive statistics and Table 5 providing correlation measures between explanatory variables).

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

- As variables accounting for geographical 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 link 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.[13] Whereas the degree centrality simply measures the number of collaboration links of a region, the hub score (Kleinberg's authority centrality[14]) is defined as the principal eigenvector of $A * t(A)$, where A denotes the adjacency matrix of the KET-specific R&D network and 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, as well as difference in the *quality* of these interactions.

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

5.3 Assignment of data items to KETs

The meaningful delimitation of KETs is essential for 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

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

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

a hierarchical system of topical classes for each KET that are each characterized 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 characterized 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.[15]

Second, projects are tagged and then mapped to specific KET subtopics which are aggregated to the six main KETs to extract the KET-specific collaboration networks or 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 subtopics belonging to this specific KET. The similarity score depends basically on the overlap in keywords from the ontology and the text of the project description, whereas the keywords are weighted by their representativeness for a specific topic using pointwise mutual information (PMI) procedures (see Blei 2012). Note that assignment of projects is subjected to a series of robustness and sensitivity analyses (including manual checking of individual cases) to guarantee a sufficiently meaningful and robust result (see Maynard et al. 2017 for details on the assignment procedure).[16] 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).

6 Estimation results

Table 1 displays the estimation results of the spatial interaction models. 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 geographical 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 Sects. 2, 3), since we can observe directly the changes in the spatial effects, when accounting for network structural effects. Each of the two model specifications was executed for

[15] Different natural language processing (NLP) techniques are used to refine sets of keywords and explore interrelations between them (e.g. two generic keywords are marked as stop-words, and combinations of keywords and multi-term keywords are constructed that are specifically relevant for a topic to get a better discrimination (Maynard et al. 2017).

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

Table 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	−0.245***	−0.144***	−0.250***	−0.145***	−0.148***	−0.189***	−0.213***	−0.097***	−0.222***	−0.123***	−0.083***	−0.157***
[β_1]	(0.008)	(0.012)	(0.008)	(0.014)	(0.010)	(0.007)	(0.008)	(0.012)	(0.008)	(0.015)	(0.010)	(0.007)
Country border effects	−0.153***	−0.250***	−0.162***	−0.199***	−0.212***	−0.213***	−0.185***	−0.298***	−0.192***	−0.233***	−0.281***	−0.235***
[β_2]	(0.020)	(0.033)	(0.021)	(0.039)	(0.026)	(0.018)	(0.020)	(0.033)	(0.021)	(0.039)	(0.026)	(0.018)
Metropolitan region	0.210***	0.071***	0.144***	−0.021	0.110***	0.153***	0.186***	0.034*	0.135***	−0.033	0.120***	0.131***
[β_3]	(0.010)	(0.017)	(0.011)	(0.020)	(0.013)	(0.009)	(0.010)	(0.017)	(0.011)	(0.020)	(0.013)	(0.009)
Gap in degree centralities	–	–	–	–	–	–	−0.137***	−0.178***	−0.117***	−0.062***	−0.272***	−0.148***
[β_4]							(0.005)	(0.009)	(0.006)	(0.012)	(0.007)	(0.005)
Gap in hub score	–	–	–	–	–	–	−0.170**	−1.291***	0.238***	−0.962***	−0.156*	−0.280***
[β_5]							(0.054)	(0.084)	(0.059)	(0.095)	(0.066)	(0.048)
Constant [α_0]	−5.906***	−5.999***	−5.638***	−6.473***	−6.055***	−6.154***	−5.701***	−5.997***	−5.421***	−6.625***	−5.663***	−5.883***
	(0.056)	(0.090)	(0.061)	(0.106)	(0.075)	(0.052)	(0.058)	(0.089)	(0.062)	(0.107)	(0.074)	(0.054)
Dispersion [θ]	1.117***	0.733***	0.921***	0.747***	0.806***	1.238***	1.144***	0.760***	0.933***	0.755***	0.852***	1.293***
	(0.015)	(0.015)	(0.012)	(0.020)	(0.013)	(0.015)	(0.015)	(0.015)	(0.012)	(0.021)	(0.014)	(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***

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, and *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 all six KETs to allow the comparison between 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 time period (i.e. summing up FP7 and H2020) due to the extremely high number of zeros challenging a reasonable estimation.

In our discussion, we shift explicit 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 1), i.e. the number of organizations 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 effect 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,[17] 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 organizations located in these regions.

This is a somewhat sobering outcome in a European integration and policy context. While country border effects seem to diminish in networks of the FP as a whole (Scherngell and Lata 2013), in KETs—that are considered as 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 organizations by +0.7% in the field of *Microelectronics* that exhibits the smallest effect and +23% in *Nanotechnology* with the largest 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: (1) *Nanotechnology* and *Photonics* and (2) *Microelectronics*, *Advanced*

[17] 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).

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 and 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 effect and the weakest effect of geographical distance. Hence, R&D collaborations in *Nanotechnology* and *Photonics* are much more localized 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 are 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 larger 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 the 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 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* exhibit the smallest effect (change of −0.6%).

The effects of the gap in hub score point in 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 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.

Reviewing 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. This is what Torre and Rallet (2005) refer to as the '*logic* *of similarity'* of organized proximity. Although this conception originally refers to the organizational level, it may also be applied to the regional level. In context of our results, this could be interpreted insofar as regions that are similar in terms of research infrastructure, types of researching organizations, technological profiles, etc. share same frameworks and systems of representation, which facilitate the ability for organizations located in these regions to interact. This holds true for research-intensive regions with large numbers of organizations, 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 searching 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 amount of small- and medium-sized organizations with a mediocre amount of network links, in contrary to large technology hubs and industry clusters in need for equivalent partners to engage in cross-regional R&D activities.

Strikingly, considering 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 fairly strong proxy effects between geographical and relational effects. Both geographical distance and the country border effect change in opposite directions, increasing the impact of 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 for long-distance collaborations increases as well.

Resuming these results in 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 the 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. Specifically, R&D collaborations have more of a localized character in *Nanotechnology* and *Photonics* and are relatively global in *Microelectronics* and *AMT*. In terms of relational effects, especially *Microelectronics* stands out with a distinguished authority- and hub-structured network, whereas on the contrary, findings for *Photonics* indicate a 'hub and spoke' structured network. With respect to hypothesis 2b, we cannot—at least with our focus on six KETs in this study—confirm our assumption that geographical effects have a stronger negative impact in engineering-oriented fields, whereas network structural effects are more important for science-oriented fields. In fact, *Advanced Materials* and *AMT*—both being characterized 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.

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 collaboration 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 such networks in space and time, covering increasingly large geographical areas and time periods.[18]

Empirical studies investigating determinants of R&D collaboration networks—mostly done at the regional level of analysis—have so far brought the interesting results (see Scherngell 2019 for an overview), pointing to the still important 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

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

technological differences that may be prevalent across these results. Such technological heterogeneities are assumed to play a major role, given the different knowledge bases and knowledge creation regimes underlying different technological fields, and accordingly, different collaboration behaviours.

This study has addressed this research gap, aiming to identify spatial, as well as 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, specifically 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 context of the 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 sobering 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: (1) *Nanotechnology* and *Photonics*, and (2) *Microelectronics, Advanced Materials* 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 characterized 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, the results underline that network effects are able to compensate for geographical barriers—throughout 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 that are 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 indeed can observe significant 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, whereas science-oriented fields are more driven by network structural effects, is not supported by the findings.

From a policy perspective, the findings are of high interest with respect to the interplay between geographical and relational effects and their relative importance for the different KETs, which requires tailored policy measures; specifically, the potential of networks to reduce geographical barriers is of great interest, encouraging further policies, in particular, for lagging regions, supporting the participation in networks. However, in light of the differing configuration of the effects across KETs, some differing policy conclusions could be considered across them. On the one hand, we identify KETs with relatively high geographical barriers (Nanotechnology and Photonics) hindering R&D collaborations, pointing towards the existence of regional technological clusters that require cluster-oriented policy measures to strengthen regional research infrastructure and accelerate regional knowledge creation. However, with relational effects being of general importance for R&D collaborations—as suggested by the findings—policymakers may aim at providing incentives for organizations within such clusters to establish new national and supra-national R&D collaboration links. This enables knowledge exchange and diffusion among the clusters, enhancing the regional knowledge base. On the other hand, for KETs that exhibit relatively lower geographical barriers (*Microelectronics, Advanced Materials* and *AMT*) for R&D collaborations, policymakers may rather focus on establishing strong and sustainable inter-regional R&D collaboration networks, rather than creating new network links. This entails providing incentives for organizations to collaborate with partners from geographically peripheral and less embedded regions, since the findings of this study suggest that large differences in number and quality of network links are considerable barriers for R&D collaborations between regions, which needs to be actively addressed by European policymakers.

Finally, 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 specific 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 utilizing the effect estimates from this study in a simulation approach—may provide better understanding of the results presented here, in particular as what concerns the differing determinants and their magnitude in different technological fields.

Acknowledgements We are grateful to the anonymous reviewers for their valuable comments on this manuscript. This work was supported by the research projects RISIS2 (Research Infrastructure for and Innovation Policy Studies 2, H2020, Grant Agreement No. 824091), KNOWMAK (H2020, Grant Agreement No. 726992) and MULTIREG (FWF Austrian Science Fund, Grant Agreement No. P28936-G27).

Appendix 1: 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 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

$$W' = \left(I - \iota\iota' \frac{1}{N}\right) W \left(I - \iota\iota' \frac{1}{N}\right) \tag{8}$$

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 $E_n = \left(E_1, E_2, \ldots, E_N\right)$ and their associated N eigenvalues $\lambda = \left(\lambda_1, \lambda_2, \ldots, \lambda_N\right)$. 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.

Appendix 2: Descriptive statistics

See Tables 2, 3, 4 and 5; Figs. 1, 2, 3.

Table 2 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	2774	1820	2464	323	1076	3534
# Inter-regional collaborations	77,245	23,940	64,506	10,364	38,678	109,329
# Organizations	5189	1820	4559	1298	2363	5912

# Denotes 'number', *Nano* Nanotechnology, *Micro* Microelectronics, *AM* Advanced Materials, *AMT* Advanced Manufacturing Technologies, *Ind. Biotech.* Industrial Biotechnology

Table 3 R&D collaboration network characteristics of six KETs

	Nano	Micro	Photonics	AM	AMT	Biotech
Number of edges	19,510	8278	17,754	5709	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 centralization	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 centralization	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

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

Table 4 Descriptive statistics of regression variables

	Nano	Micro	Photonics	AM	AMT	Biotech
Number of R&D collaborations (dependent variable)						
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						
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						
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						
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						
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

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

Table 5 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	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	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	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

Table 5 (continued)

	Origin/destination	Geogr. distance	Country border	Metro region	Gap degree centralities	Gap in hub score
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	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	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

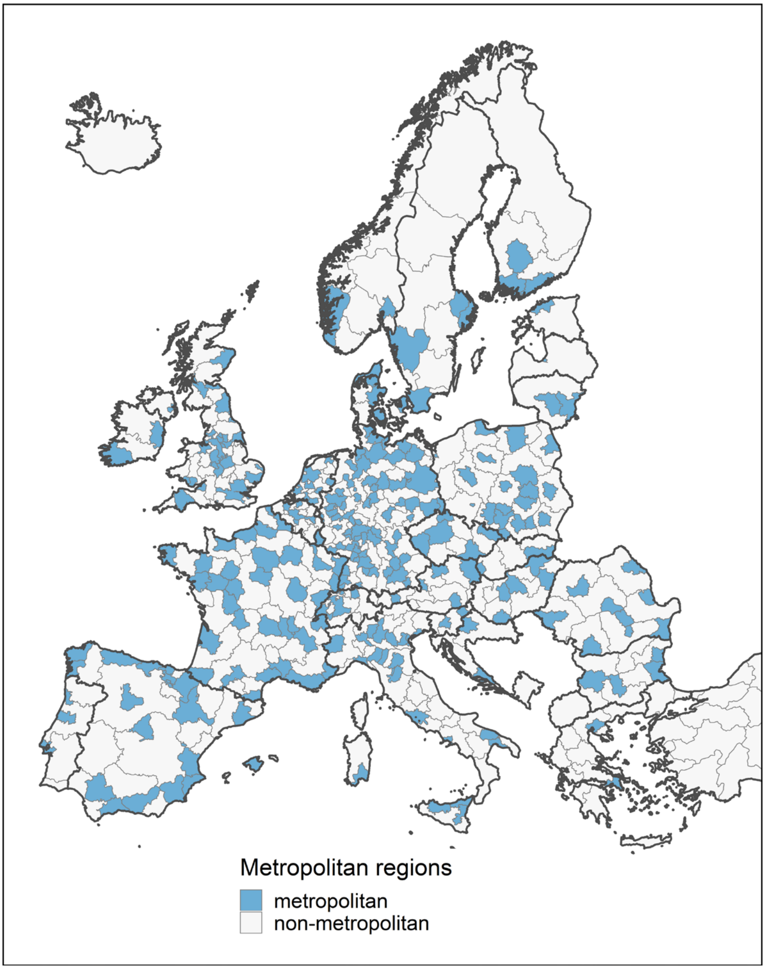
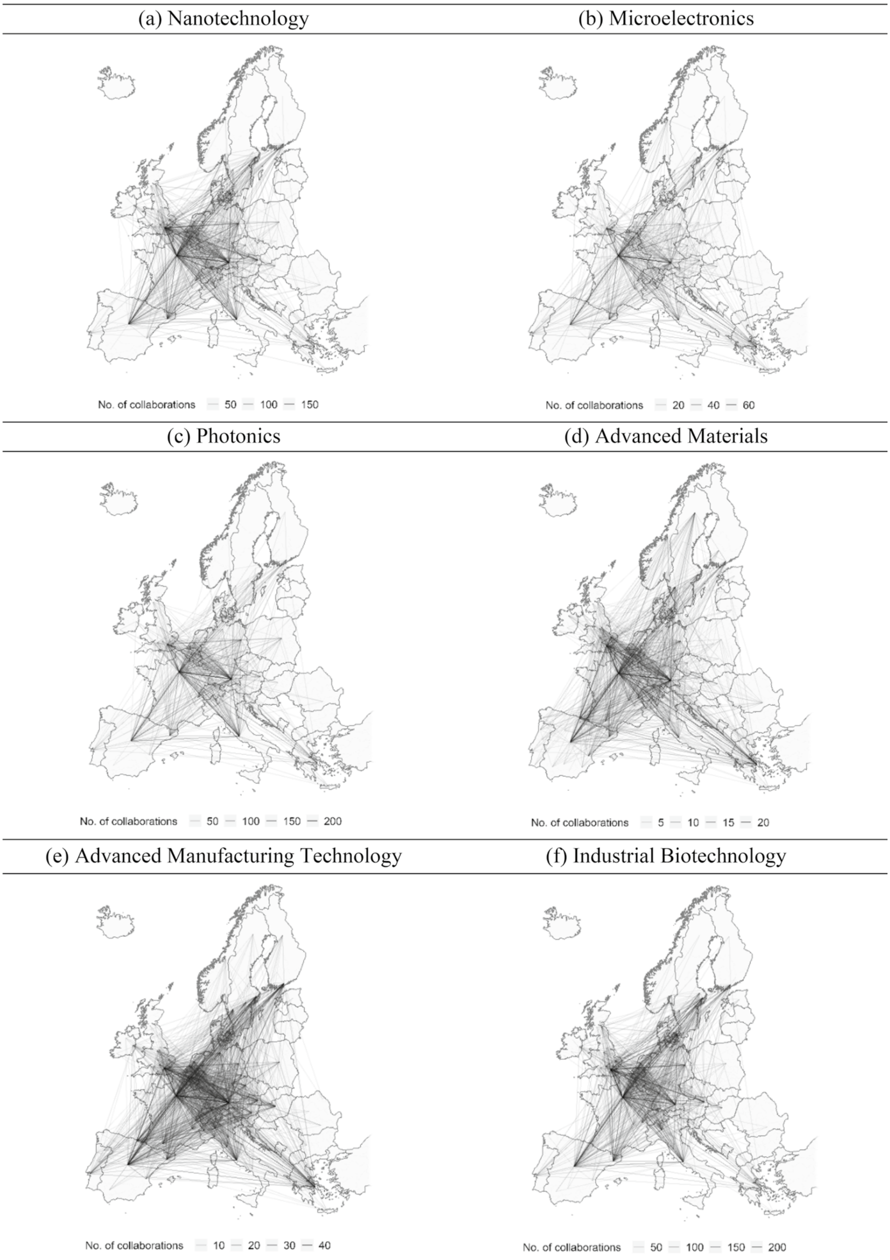


Fig. 1 Metropolitan and non-metropolitan regions



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

Fig. 2 Spatial R&D networks of Key Enabling Technologies (2007–2016)

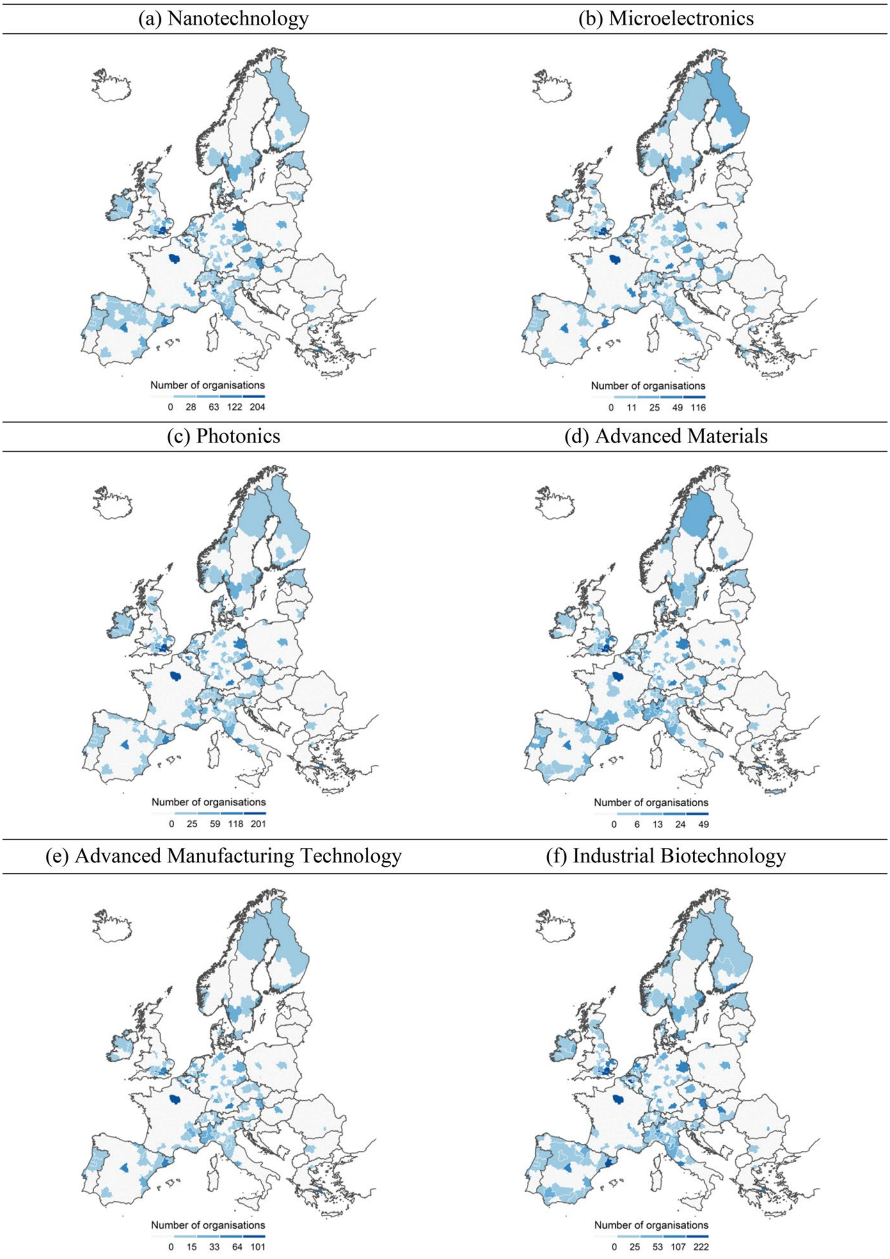


Fig. 3 Spatial distribution of organizations in Key Enabling Technologies (2007–2016)

References

Ahuja G (2000) Collaboration networks, structural holes, and innovation: A longitudinal study. Adm Sci Q 45(3):425–455
Asheim BT, Coenen L (2005) Knowledge bases and regional innovation systems: comparing Nordic clusters. Res Policy 34(8):1173–1190
Autant-Bernard C, Mairesse J, Massard N (2007) Spatial knowledge diffusion through collaborative networks. Pap Reg Sci 86(3):341–350
Balland P-A, Rigby D (2017) The geography of complex knowledge. Econ Geogr 93(1):1–23
Barthélemy M (2011) Spatial networks. Phys Rep 499(1–3):1–101
Bathelt H, Glückler J (2003) Toward a relational economic geography. J Econ Geogr 3(2):117–144
Bathelt H, Malmberg A, Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Prog Hum Geogr 28(1):31–56
Bell GG, Zaheer A (2007) Geography, networks, and knowledge flow. Organ Sci 18(6):955–972
Bergé LR (2017) Network proximity in the geography of research collaboration. Pap Reg Sci 96(4):785–815
Blei DM (2012) Probabilistic topic models. Commun ACM 55(4):77–84
Boschma R (2005) Proximity and innovation: a critical assessment. Reg Stud 39(1):61–74
Boschma RA, Ter Wal AL (2007) Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy. Ind Innov 14(2):177–199
Breschi S, Cusmano L (2004) Unveiling the texture of a European research area: emergence of oligarchic networks under EU Framework Programmes. Int J Technol Manage 27(8):747–772
Breschi S, Malerba F, Orsenigo L (2000) Technological regimes and Schumpeterian patterns of innovation. Econ J 110(463):388–410
Broekel T, Boschma R (2012) Knowledge networks in the Dutch aviation industry: the proximity paradox. J Econ Geogr 12(2):409–433
Broekel T, Graf H (2012) Public research intensity and the structure of German R&D networks: a comparison of 10 technologies. Econ Innov New Technol 21(4):345–372
Broekel T, Hartog M (2013) Explaining the structure of inter-organizational networks using exponential random graph models. Industry and Innovation 20(3):277–295
Cameron C, Trivedi P (1998) Models for count data. Cambridge University Press, Cambridge
Chun Y (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. J Geogr Syst 10(4):317–344
Dijkstra L (2009) Metropolitan regions in the EU. Regional Focus. Brussels, European Union-Regional Policy. 01/2009
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. European Commission, Brussels
EC (2012) A European strategy for key enabling technologies: a bridge to growth and jobs. European Commission, Brussels
EC (2015) KETs: Time to act. Final report of the high-level expert group on key enabling technologies. European Commission, Brussels
Evangelista R, Meliciani V, Vezzani A (2018) Specialisation in key enabling technologies and regional growth in Europe. Econ Innov New Technol 27:273–289
Fischer MM, Wang J (2011) Spatial data analysis: models, methods and techniques. Springer, Berlin
Fleming L, Sorenson O (2001) Technology as a complex adaptive system: evidence from patent data. Res Policy 30(7):1019–1039
Freeman C (1982) The economics of industrial innovation. Frances Pinter, London, UK
Gilsing V, Nooteboom B, Vanhaverbeke W, Duysters G, van den Oord A (2008) Network embeddedness and the exploration of novel technologies: technological distance, betweenness centrality and density. Res Policy 37(10):1717–1731
Giuliani E (2007) The selective nature of knowledge networks in clusters: evidence from the wine industry. J Econ Geogr 7(2):139–168
Glückler J (2006) A relational assessment of international market entry in management consulting. J Econ Geogr 6(3):369–393

Glückler J, Lazega E, Hammer I (2017) Exploring the interaction of space and networks in the creation of knowledge: an introduction. Knowledge and networks. Springer, Cham, pp 1–21
Griffith D, Chun Y (2014) Spatial autocorrelation and spatial filtering. In: Fischer MM, Nijkamp P (eds) Handbook of regional science. Springer, Berlin, pp 1477–1507
Hansen MT, Løvås B (2004) How do multinational companies leverage technological competencies? Moving from single to interdependent explanations. Strateg Manag J 25(8–9):801–822
Heller-Schuh B, Barber M, Züger M, Scherngell T (2015) Report on the content and technical structure of the EUPRO infrastructure, AIT Austrian Insitute of Technology
Hoekman J, Frenken K, Tijssen RJ (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. Res Policy 39(5):662–673
Huggins R, Thompson P (2014) A network-based view of regional growth. J Econ Geogr 14(3):511–545
Kirat T, Lung Y (1999) Innovation and proximity: territories as loci of collective learning processes. Eur Urban Reg Stud 6(1):27–38
Kleinberg JM (1999) Hubs, authorities, and communities. ACM Comput Surv (CSUR) 31(4):5
Lata R, Scherngell T, Brenner T (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. Geogr Anal 47(4):349–375
LeSage JP, Fischer MM (2016) Spatial regression-based model specifications for exogenous and endogenous spatial interaction. In: Patuelli R, Arbia G (eds) Spatial econometric interaction modelling. Springer, Berlin, pp 15–36
Long SJ, Freese J (2006) Regression models for categorical dependent variables using Stata. Stata Press, New York
Maggioni MA, Uberti TE (2011) Networks and geography in the economics of knowledge flows. Qual Quant 45(5):1031–1051
Maggioni MA, Nosvelli M, Uberti TE (2007) Space versus networks in the geography of innovation: a European analysis. Pap Reg Sci 86(3):471–493
Malerba F (2002) Sectoral systems of innovation and production. Res Policy 31(2):247–264
Malerba F, Orsenigo L (2000) Knowledge, innovative activities and industrial evolution. Ind Corp Change 9(2):289–314
Marek P, Titze M, Fuhrmeister C, Blum U (2017) R&D collaborations and the role of proximity. Reg Stud 51(12):1761–1773
Martin R, Moodysson J (2013) Comparing knowledge bases: on the geography and organization of knowledge sourcing in the regional innovation system of Scania, Sweden. Eur Urban Reg Stud 20(2):170–187
Mattes J (2012) Dimensions of proximity and knowledge bases: innovation between spatial and non-spatial factors. Reg Stud 46(8):1085–1099
Maynard D, Funk A, Lepori B (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
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Ann Rev Soc 27(1):415–444
Montresor S, Quatraro F (2017) Regional branching and key enabling technologies: evidence from European patent data. Econ Geogr 93:367–396
Morescalchi A, Pammolli F, Penner O, Petersen AM, Riccaboni M (2015) The evolution of networks of innovators within and across borders: evidence from patent data. Res Policy 44(3):651–668
Nelson RR, Winter SG (1982) An evolutionary theory of economic change. Harvard University Press, Cambridge
Pavitt K (1984) Sectoral patterns of technical change: towards a taxonomy and a theory. Res Policy 13(6):343–373
Ponds R, Van Oort F, Frenken K (2007) The geographical and institutional proximity of research collaboration. Pap Reg Sci 86(3):423–443
Popadiuk S, Choo CW (2006) Innovation and knowledge creation: How are these concepts related? Int J Inf Manage 26(4):302–312
Rallet A, Torre A (1998) On geography and technology: the case of proximity relations in localized innovation networks. Clusters and regional specialisation: on geography, technology and networks. Pion, London
Roy JR, Thill J-C (2003) Spatial interaction modelling. Papers in Regional Science 83(1):339–361

Salman N, Saives AL (2005) Indirect networks: an intangible resource for biotechnology innovation. R&D Management 35(2):203–215
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. Handbook of Regional Science. M. M. Fischer and P. Nijkamp. Springer, Berlin
Scherngell T, Barber MJ (2009) Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. Pap Reg Sci 88(3):531–546
Scherngell T, Lata R (2013) Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. Pap Reg Sci 92(3):555–577
Storper M, Venables AJ (2004) Buzz: face-to-face contact and the urban economy. J Econ Geogr 4(4):351–370
Tiefelsdorf M, Boots B (1995) The exact distribution of Moran's I. Environ Plan A 27(6):985–999
Tödtling F, Lehner P, Trippl M (2006) Innovation in knowledge intensive industries: the nature and geography of knowledge links. Eur Plan Stud 14(8):1035–1058
Torre A, Rallet A (2005) Proximity and localization. Reg Stud 39(1):47–59
Trippl M, Tödtling F, Lengauer L (2009) Knowledge sourcing beyond buzz and pipelines: evidence from the Vienna software sector. Econ Geogr 85(4):443–462
Uzzi B, Lancaster R (2003) Relational embeddedness and learning: the case of bank loan managers and their clients. Manage Sci 49(4):383–399
Wanzenböck I, Piribauer P (2018) R&D networks and regional knowledge production in Europe: evidence from a space-time model. Pap Reg Sci 97:1
Wanzenböck I, Scherngell T, Brenner T (2014) Embeddedness of regions in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications. Ann Reg Sci 53(2):337–368
Wanzenböck I, Neuländtner M, Scherngell T (2020) Impacts of EU funded R&D networks on the generation of key enabling technologies: empirical evidence from a regional perspective. Pap Reg Sci 99(1):3–24
Whittington KB, Owen-Smith J, Powell WW (2009) Networks, propinquity, and innovation in knowledge-intensive industries. Adm Sci Q 54(1):90–122
Winter SG (1987) Knowledge and competence as strategic assets. In: Teece DJ (ed) The competitive challenge: strategies for industrial renewal. Ballinger Publishing, Cambridge, MA
Zaheer A, Bell GG (2005) Benefiting from network position: firm capabilities, structural holes, and performance. Strateg Manag J 26(9):809–825