Determinants of technology-specific R&D collaboration networks: Evidence from a spatial interaction modelling perspective

Martina Neuländtner[1] and Thomas Scherngell[2]

[1] *martina.neulaendtner@ait.ac.at*
AIT Austrian Institute of Technology GmbH, Giefinggasse 4, 1210 Vienna (Austria)

[2] *thomas.scherngell@ait.ac.at*
AIT Austrian Institute of Technology GmbH, Giefinggasse 4, 1210 Vienna (Austria)

Abstract

It is commonly acknowledged, that the creation of knowledge is the result of interactive, collaborative learning processes among organizations of different types located in different regions. Especially, in a strongly knowledge-based economy built on fast-growing and R&D-intensive technologies such as Key Enabling Technologies (KETs), collaborative knowledge creation is gaining importance to rapidly enable access to external, nation-wide and global new sources of knowledge. With the focus on technology-specific R&D collaboration networks in six KETs, each representing different knowledge bases and modes of (collaborative) knowledge creation, we emphasize the determining role of technology-specific heterogeneities. The objective is to estimate determinants of these technology-specific R&D collaboration networks, focusing on spatial separation and network structural effects. We employ a spatially filtered negative binomial spatial interaction model with a set of 521 regions to identify differences in the determinants of technological knowledge flows, proxied by EU-funded collaborative projects. The results show differences in the relative importance of the determinants. Geographical barriers are significant, and network structural effects are of high importance, but do not remove spatial effects in all KETs. Both spatial and network effects seem to be of higher relevance for more industrial and engineering based than more science based technological fields.

Introduction

Collaborative Research and Development (R&D) activities between firms, universities and research organisations 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 describing such collaborative endeavours (see Barber and Scherngell 2013) and has become a major research domain in manifold aspects. One major research stream is – without doubt – the identification and estimation of determinants affecting structures and dynamics of such networks, often with a geographical focus and accomplished at the regional level of analysis (see Scherngell and Barber 2009; Hoekman et al. 2010; Scherngell and Lata 2013; Lata et al. 2015). However, these works capture R&D, and accordingly the underlying knowledge, in a quite aggregated manner, neglecting technology-specific peculiarities of knowledge creation and interactions, such as technological regimes, as well as different modes of (collaborative) knowledge creation.

This study intends to address this research gap by accounting for technological idiosyncrasies when explaining the constitution and dynamics of R&D collaboration networks. Accordingly, the objective is to estimate determinants of technology-specific R&D collaboration networks, shifting particular attention – as in previous works – to spatial separation effects, such as geographical distance or country borders, but also to network structural effects, such as central positioning, influencing the collaboration probability between two organisations. To address this objective, we employ a spatial interaction modelling approach at the regional level of analysis. The R&D collaboration network under consideration is the project-based 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 of analysis, using a set of 521 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. Semantic technologies are used to assign projects to KETs based on sets of keywords and semantic characteristics.
In what follows, we discuss in some more detail the relevance of technological heterogeneities in the context of R&D collaboration, before we introduce the model, the data, and variables. The study closes with a compact presentation and discussion of the results, and some ideas for a future research agenda.

Technological heterogeneities in R&D collaboration networks

Previous studies that identify determinants of R&D collaboration networks, find quite robust and increasingly stylized results, in particular in terms of their spatial dynamics. Spatial proximity turns out to be an important factor for the constitution of R&D collaboration, also in times of increasing globalisation and new communication technologies (see e.g. Scherngell and Barber 2009). However, while geographical barriers seem still to be significant, they tend to decrease in terms of their relative importance as when compared to other forms of separation, such as technological distance (see e.g. Scherngell and Lata 2013), and/or network structural effects (see Autant-Bernard et al. 2007; Broekel 2012).
Although these insights are interesting and have substantially increased the understanding on structure and dynamics of such networks, the main limitation is the disregard of technological heterogeneities that may influence the relevance and spatial scale of R&D collaboration (see Ponds et al. 2007; Martin and Moodysson 2013). Therefore, this study intends to shift the focus on the debate of the differing role of main determinants for R&D collaboration networks in different technologies. Concerning technological differences, especially novel and fast-growing technologies that spur innovation and technological progress of countries, regions and industries have gained anew interest. 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).
KETs are understood as generic technologies, which are characterised by relatively rapid pervasiveness and growth. They constitute technological inputs for the development of innovation, and by this require high R&D intensity and a high input of skilled labour in their creation. The European Commission defined six KETs: N*anotechnology*, *Micro- and Nanoelectronics*, *Photonics*, *Advanced materials (AM)*, *Advanced manufacturing technology (AMT)* and *Industrial biotechnology* (EC 2009). Due to the specific characteristics of KETs – knowledge intensive, high R&D intensity, rapid innovation cycles, highly skilled employment etc. (EC 2009) – R&D networks are considered of particular importance in order to cope with the high demand for R&D and to gain rapid access to nation-wide and global state of the art knowledge. Specifically, in such globally relevant technologies, 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). Hence, stimulating knowledge creation and interaction in KET fields has become one of the major priorities of the EU industrial policy to accelerate industrial restructuring and change, particularly in structurally weak regions.
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. Collaborative research activities follow different rationales and aims – especially across different technologies, leading to different outcomes of knowledge creation, that have not been accounted for so far in the literature; comprehensive investigations and studies allowing for comparisons on the role of networks in different technological fields are still missing. Hence, we go beyond the beaten track by investigating determinants of technology-specific R&D collaboration networks – proxied by KET fields – where we especially focus on

spatial and network structural effects. We employ a comparative perspective on the effect sizes of the determinants of KET collaboration networks to gain insights in technology-specific heterogeneities.
Satisfying the multidisciplinary character of KETs, cutting across many technological domains on regional, national and subnational levels, makes a spatial network perspective with a spatial interaction modelling approach evident. By this, this research will be the first, bridging networks and KETs from a spatial perspective to examine systematically the (claimed) converging and integrating nature of KETs, explicitly considering the systemic, cross-sectoral and inter-regional character of KETs.

Methodological approach and model

For the estimation of spatial and network structural determinants on technology-specific R&D collaboration networks, we follow earlier research and employ a spatial interaction modelling approach. 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. In general, these types of models comprise three types of factors to explain mean interaction frequencies between spatial locations i and j. The general form of the model can be written as

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

where $\mu_{ij} = E(Y_{ij})$ is the expected mean interaction frequency from i to j, and ε_{ij} is an error about the mean (Fischer and Wang 2011).
In this specific context of application, locations correspond to European metropolitan and remaining non-metropolitan regions, where each location is both origin and destination of interactions. 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 constrains or impedes the interaction, most basically geographical distance (LeSage and Fischer 2016). Hence, the mean interaction frequencies between origin i and destination j are modelled by

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

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 used to specify origin-, destination- and separation functions (see Fischer and Wang 2011), studies investigating R&D networks usually employ univariate (i.e. with only 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

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

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

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

Here, o_i and d_j are measured in terms 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 organisations in a region. 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. Core of the spatial interaction model is the separation function as defined by Equation (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 here takes the specific form of a spatially filtered, negative binomial spatial interaction model (see Scherngell and Lata 2013 in a similar context)[1]. The main 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 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 to the Poisson model that assumes equidispersion (i.e. conditional mean equals the conditional variance), the negative binomial model explicitly corrects for overdispersion[2], by adding a dispersion parameter γ. Hence, the negative binomial spatial interaction model takes the form

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

where $\mu_{ij} = E\left[y_{ij} \middle| O_i, D_j, S_{ij}\right] = \exp\left[O_i(\alpha_1)\, D_j(\alpha_2)\, S_{ij}(\beta)\right]$ and Γ denotes the gamma function with a model parameter γ 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[3]. ESF is based on the mathematical relationship between the Moran's I, as a measure for spatial autocorrelation, and spatial weights 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 weights matrix (see e.g. Fischer and Wang (2011) on construction of spatial weights matrices), and then add these vectors as control variables to the regression model.

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 κ of eigenvectors with MI/MI_{max} larger than 0.25, 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 full empirical model is specified by setting

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

where E_q denotes the selected subset of eigenvectors expanded by means of the Kronecker product associated to the origin variable, and E_r the respective eigenvectors for the destination

[1] 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).

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

[3] 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 neighboring origins or destinations, respectively. Not accounting for spatial autocorrelation leads, similar to overdispersion, to incorrect inferences and hence, wrong significances (Chun 2008).

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.

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 currently 28 EU member states, plus Switzerland and Norway, corresponding to a set of 521 regions. Going beyond previous research, we distinguish metropolitan regions as well as remaining non-metropolitan regions based on the 2013 NUTS version and the 2010 Geostat population grid defined by Eurostat[4].

Dependent variable

As dependent variable EU-funded KET R&D collaboration links are used (see Table 1 for some descriptive statistics). Data is extracted from the EUPRO data base 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 and type or participating organization and their geographical location in the form of organization addresses (see Heller-Schuh et al. 2015 for details). The latter is used to geolocalize participating organizations and to assign them to regions, enabling the observation of region-level R&D collaboration activities[5].

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

	Nano	Micro	Photonics	AM	AMT	Biotech
# All links	271441	271441	271441	271441	271441	271441
# Positive links	18434	6923	19229	2641	9169	32867
% Zero links	93.21	97.45	92.916	99.03	96.62	87.89
# Intra-regional collaborations	909	358	1147	44	266	2251
# Inter-regional collaborations	48852	15500	54052	3634	17246	129634
# Organizations	2221	850	2391	305	1035	3911

Notes: Nano = Nanotechnology, Micro = Micro- and Nanoelectronics, AM = Advanced materials, AMT = Advanced manufacturing technologies, Ind. Biotech. = Industrial biotechnology

To construct the dependent variable, we consider the 7th FP and H2020, i.e. 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[6]. 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.

[4] 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.

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

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

Independent variables

As described in the previous section, the independent variables comprise three types of variables: 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 joint EU-funded FP projects in region i and j in a distinct KET field. For the separation variables, we distinguish between (i) spatial separation variables, (ii) network structural separation variables, as well as (iii) a measure for technological separation. 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.

The separation variables constitute the core of the spatial interaction model since they can be interpreted as determinants of the technology-specific R&D collaboration networks. The focus of this study lies on the spatial as well as network structural determinants:

- As variables accounting for spatial separation effects, *first*, the geographical distance, measured as the great circle distance, indicating the shortest distance between two regions i and j, *second*, 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*, 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 and *second*, the gap in the hub score between the two regions i and j, are included. Whereas, the degree centrality simply measures the number of collaboration links of a region, the hub score (Kleinberg's authority centrality[7]) is defined as the principal eigenvector of $A\ t(A)$, where A is the adjacency matrix of the KET-specific R&D network and hence, indicates whether a region holds reliable information on the topic of interest and at the same time is linked to other regions, themselves with reliable information. Together, the two variables account for differences in the quantity of collaboration links, as well as difference in the quality of these interactions.
- As a last separation variable, the technological distance is included. It is constructed by using patent data drawn from the IFRIS-PATSTAT, which provides KET-specific structured information on patent applications including details on the patent itself, e.g. date of application and technology classes, as well as information on applicants and inventors, such as their names and location[8]. Technological distance between two regions i and j is defined as the correlation between the vectors of patent applications in 353 KET-subtopics (see '*Assignment of data items to KETs'* for details); the technological distances between regions is KET-specific.

Assignment of data items to KETs

The meaningful delimitation of KETs in EUPRO and IFRIS-PATSTAT 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 a both datasets under consideration. Thus, we employ the classification approach developed in the EU-funded project KNOWMAK that provides an ontology for KETs, comprising a hierarchical system of topical classes for each KET that are characterised by a set of weighted keywords. The data items are assigned to these topical classes, based on a scoring system that evaluates the similarity of a text (in our case an abstract of a FP project or patent) to a keyword set of a specific KET-related subtopic. By this, projects and patents are tagged to specific KET subtopics which are aggregated to the six main

---

[7] Equals the authority score for undirected graphs.

[8] IFRIS-PATSTAT is based on the PATSTAT database (developed by the European patent office) and accessible via RISIS (risis.eu). It shifts attention to organisation names cleaning of PATSTAT, geolocalisation, and most importantly in our context, assignment of patents to KETs.

KETs to extract the six KET-specific collaboration networks and patenting activities for the analysis at hand. Note that assignment of projects is subject to a series of robustness and sensitivity analysis (including manual checking of individual cases) to guarantee a sufficiently meaningful and robust result (see also Maynard et al. 2017)[9].

Estimation results

In Table 2, the estimation results of the spatial interaction models are displayed. 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 geographical distance and the country border effect as separation measures, the second column comprises the results for the full model (model 2) including an additional set of spatial and network structural separation measures. Each of the two model specifications was executed for 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 γ-parameter suggests the preference of a negative binomial model over the Poisson specification. Moreover, a likelihood ratio test was conducted testing the spatially filtered negative binomial model against the non-filtered version, clearly pointing towards the filtered adaptation for all models. 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 focus on the separation variables. As can be seen from Table 2, the origin- and destination variables that just control for the mass in the origin and the destination region are significant and higher than one, i.e. the number of organisations active in a KET in a region obviously 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 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 Industrial biotechnology (for a coefficient of -0.277 this equals to a change of -0.24% given by its exponential[10], followed by Nanotechnology (with a factor change of 0.78; i.e. a change of - 0.22%), the effect for Advanced materials (AM) is the smallest with a factor change of 0.90 (i.e. a change of -0.10%).

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 is a rather negative outcome in a European integration and policy context. While country border effects seem to diminish in networks of the FP as a whole as evidenced by 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. Interestingly, here the negative effects are the lowest for Nanotechnology and Industrial biotechnology, i.e. the more science-based fields, while AM shows by far the highest negative effect. For region pairs located in different countries the expected number of collaborations is hypothetically decreased by 49% in the case of AM.

Model (2) adds the technological distance and network structural separation variables, as well as the metropolitan region dummy as additional spatial separation. Interestingly, the interpretation of the coefficient for the variables already included in model (1) stays the same,

---

9 Details on the semantic approach and also the technical tools are given at knowmak.eu

10 A change of one kilometre in geographical distance results in an expected count decrease by a factor of $\exp(-0.277) = 0.758$ which implies a change of -0.24% see Long and Freese (2006).

Table 2. 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 [$\alpha_1 = \alpha_2$]	1.411*** (0.007)	1.756*** (0.014)	1.427*** (0.008)	2.631*** (0.035)	1.775*** (0.014)	1.389*** (0.005)	1.458*** (0.009)	1.963*** (0.017)	1.482*** (0.009)	2.940*** (0.041)	1.910*** (0.016)	1.441*** (0.006)
Geographical distance [β_1]	-0.243*** (0.011)	-0.235*** (0.017)	-0.208*** (0.011)	-0.107*** (0.028)	-0.210*** (0.016)	-0.277*** (0.008)	-0.196*** (0.011)	-0.162*** (0.018)	-0.150*** (0.011)	-0.018 (0.029)	-0.173*** (0.016)	-0.233*** (0.008)
Country border effects [β_2]	-0.089*** (0.030)	-0.126** (0.050)	-0.131*** (0.029)	-0.398*** (0.082)	-0.189*** (0.044)	-0.094*** (0.022)	-0.118*** (0.030)	-0.235*** (0.052)	-0.186*** (0.029)	-0.503*** (0.086)	-0.225*** (0.044)	-0.124*** (0.021)
Technological distance [β_3]	-	-	-	-	-	-	-0.352*** (0.069)	-0.065 (0.120)	-0.358*** (0.070)	0.008 (0.198)	-0.184 (0.105)	-0.295*** (0.050)
Metropolitan region [β_4]	-	-	-	-	-	-	0.229*** (0.016)	0.238*** (0.027)	0.117*** (0.016)	-0.064 (0.045)	0.000 (0.024)	0.182*** (0.011)
Gap in degree centralities [β_5]	-	-	-	-	-	-	-0.128*** (0.009)	-0.277*** (0.014)	-0.166*** (0.008)	-0.440*** (0.027)	-0.128*** (0.013)	-0.148*** (0.006)
Gap in hub score [β_6]	-	-	-	-	-	-	-0.885*** (0.077)	-1.574*** (0.120)	-0.562*** (0.077)	-0.079 (0.194)	-1.518*** (0.116)	-0.963*** (0.057)
Constant [α_0]	-5.467*** (0.078)	-5.962*** (0.127)	-5.829*** (0.080)	-7.848*** (0.209)	-6.298*** (0.115)	-5.492*** (0.059)	-5.407*** (0.080)	-6.090*** (0.127)	-5.723*** (0.081)	-7.835*** (0.214)	-6.328*** (0.116)	-5.341*** (0.060)
Dispersion [γ]	1.052 (0.020)	1.712 (0.019)	1.303 (0.015)	2.070 (0.027)	1.770 (0.016)	0.922 (0.016)	1.038 (0.020)	1.739 (0.017)	1.291 (0.014)	2.503 (0.020)	1.750 (0.015)	0.860 (0.017)
Likelihood ratio test	924.4***	438.5***	601.7***	356.4***	564.1***	1628.5***	832.6***	370.3***	517.1***	265.9***	497.1***	1479.9***

Notes: The dependent variable is the number of EU-funded R&D collaborations between two regions; each ten origin and destination spatial filters as specified in the text are included as explanatory variables; the number of observations is 271441; 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 = Micro- and Nanoelectronics, AM = Advanced materials, AMT = Advanced manufacturing technologies, Ind. Biotech. = Industrial biotechnology in terms of significance and direction; however, geographical distance decreases in magnitude when adding the additional variables, i.e. geographical distance may partly be a proxy for the other effects reflected in the additional variables, in particular the metropolitan dummy. The latter indicates links between two metropolitan regions. The estimate is positive and significant for the KETs Nanotechnology, Microelectronics, Photonics and Industrial Biotechnology. This implies that two metropolitan regions 'increase' the expected number of collaborations of their organizations by +12% in the field of Photonics, that exhibits the smallest effect, and +27% in Microelectronics with the largest effect (compared to links between non-metropolitan regions and links between metropolitan and non-metropolitan regions).

Turning to technological distance that accounts for technological effects that may determine the collaborative activities of organisations located in the regions of interest, only for the KETs Nanotechnology, Photonics and Industrial Biotechnology the coefficient is of significance in explaining inter-regional R&D collaborations. Especially, these KETs are characterised in the literature as cross-sectoral technologies that combine approaches from physics, chemistry and biology, materials science and electrical engineering (Aschhoff et al. 2010). Comprising many scientific fields, and hence being quite heterogeneous in their technological focus suggests possibly large technological distances that have hampering effects on the collaborative activities between regions – as evidenced by the significantly negative estimates for these KETs.

The addition of network structural effects is another element where this study goes beyond previous research. The results are highly interesting, *first*, the spatial effects remain largely unchanged, showing that geographical variables are not simply proxies of underlying network structural effects, and *second*, that such network structural effects are important. 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. The effects of the gap in hub score point towards the same direction, being negative and significant for all KETs except AM. However, the gap in hub score can be interpreted as the gap in *quality* of the region in providing knowledge and enabling knowledge access. In other words, collaboration probability between two regions decreases when their difference in terms of quantity and quality of links increases, i.e. hubs are more likely to connect with other hubs than to connect with peripheral regions, which is described as preferential attachment mechanisms from a network perspective.

KET-specific differences seem to be of minor relevance in terms of the gap in degree centralities, i.e. quantity of the links. However, for the gap in the hub score, i.e. the quality of links, we find some notable differences between the KETs. In Micro- and Nanoelectronics and Advanced manufacturing technology (AMT) hub score effects are by far highest, suggesting a distinguished authorities- and hub-structured network for these KETs.

Concluding remarks

The investigation of structures and dynamics of R&D collaboration networks has become one of the most important research domains in Science, Technology and Innovation (STI) studies, 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 time periods (see, e.g. risis.eu).

Empirical studies investigating determinants of R&D collaboration networks – mostly done at the regional level of analysis – have so far brought highly interesting results (see Scherngell 2019 for an overview), pointing to the still important role of geographical barriers (geographical distance and/or country borders) and technological determinants, such as technological distance. However, the studies so far did not dig yet into 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 directly addressed this research gap, aiming to identify 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 521 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 previous research and from a European policy perspective. *First*, 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, and also not tremendously high, the significant country border effects are somewhat negative 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. *Second*, network structural effects turned out to be indeed an important additional determinant that has been neglected in previous works, in particular, pointing to the existence of preferential attachment mechanisms, i.e. regions of similar network embeddedness are more likely to collaborate than regions with a high gap in network embeddedness. Accordingly, lagging regions in terms of network centrality face statistically significant barriers to attach to more prominent regions in the network. *Third*, 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. The more science-based KETs (Nanotechnology and Biotechnology) seem to be less affected by geographical barriers than the more engineering and industrial driven fields (Advanced materials and Advanced manufacturing technology). For the latter, network structural effects seem to be of relatively higher importance, i.e. science-based fields may be more open to non-conventional network partners than industry driven fields.

Some ideas for a future research agenda come to mind. *First*, the semantic approach to assign R&D projects to technologies is based on a first version as described in Maynard et al. (2017). Updated versions of the ontology may alter the results. This needs to be checked in terms of robustness of the results. *Second*, 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. *Third*, 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.

Acknowledgements

This research greatly benefited from the data base infrastructure built up in the two EU-funded projects RISIS (risis2.eu) and KNOWMAK (knowmak.eu).

References

Aschhoff, B., D. Crass, K. Cremers, C. Grimpe, C. Rammer, F. Brandes, F. Diaz-Lopez, R. K. Woolthuis, M. Mayer and C. Montalvo (2010). European competitiveness in key enabling technologies. *Background Report to the EU Competitiveness Report*.

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

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.

Broekel, T. (2012). Collaboration intensity and regional innovation efficiency in Germany—a conditional efficiency approach. *Industry and Innovation* 19(2): 155-179.

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

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

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

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

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.

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.

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.

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

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.

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

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.

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.

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.

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

Scherngell, T. (2019). The Geography of Networks and R&D Collaboration. A review. *AIT working paper* 01/2019.

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.