Spatial and institutional dimensions of research collaboration: a multidimensional profiling of European regions

Tamás Sebestyén
University of Pécs,
Faculty of Business and Economics,
RIERC,
MTA-PTE Innovation and
Economic Growth Research Group,
Hungary
E-mail: sebestyent@ktk.pte.hu

Erik Braun
University of Pécs,
Faculty of Business and Economics,
RIERC,
CERS, Institute for Regional Studies,
Hungary
E-mail: braun.erik@ktk.pte.hu

Zita Iloskics
University of Pécs,
Faculty of Business and Economics,
RIERC,
CERS, Institute for Regional Studies,
Hungary
E-mail: iloskics.zita@ktk.pte.hu

Attila Varga
University of Pécs,
Faculty of Business and Economics,
RIERC,
MTA-PTE Innovation and
Economic Growth Research Group,
Hungary
E-mail: vargaa@ktk.pte.hu

Keywords:
research collaboration networks,
university-industry collaboration,
regional profiling,
cluster analysis

This study extends previous research by providing a large-scale analysis of Framework Program collaboration patterns from a multidimensional perspective. We use detailed data on Framework Program cooperation to map research collaboration patterns across European regions, focusing on the intensity between industry and research institutions. Using these data, we profile European NUTS3 regions along the institutional and spatial dimensions of their collaboration networks. The results show that cooperation intensities correlate among types of collaboration: most of the regions are either weakly or strongly cooperative along most of the cooperation dimensions. However, a significant group of moderately developed regions shows selective collaboration patterns, mostly with an external focus.

Introduction

There has been extensive research on the collaboration networks among European regions from different aspects. This paper contributes to the literature by providing a large-scale analysis of institutional collaboration patterns across European regions based on Framework Program participation. The novelty of our analysis is rooted in the details of the dataset, which allows drawing a research collaboration network of institutions while accounting for the spatial patterns at the NUTS3 level and types of institutions as well. Based on this dataset, we run a clustering exercise to profile European regions according to the characteristics of their collaboration patterns along the spatial and institutional dimensions. The results of this exercise reveal that there are significant differences in patterns across regions, even with similar levels of development. However, collaboration intensity is rarely specific to one or several of these (spatial and institutional) dimensions.

Recognising that innovation is an inherently collaborative process, recent research has highlighted the prominent role of cooperation in innovations (Lundvall 2010). Through innovation, different networks of cooperation can contribute to the development and growth of regions, showing that policies targeting network formation can be effective tools in promoting regional development. In addition to the general understanding that collaborative ties can positively contribute to innovation and growth, results in this field highlight the importance of interregional cooperation (Hoekman et al. 2008, Varga et al. 2014, Sebestyén–Varga 2013). Moreover, these more distant ties of knowledge flows can significantly improve innovation in the lagging regions where the local supply of resources is scarce. The networks provide access to similar resources accumulated elsewhere (Varga–Sebestyén 2017, Páthy 2017, Lengyel–Leydesdorff 2015).

However, the literature on regional innovation systems emphasises that collaborative innovation is leveraged by the cooperation between different actors like firms, universities, research centres, government agencies, financial institutions, and others, supporting innovative process (Jacobs 1969, Henderson 1997, Fritsch–Slavtchev 2010, Becker–Dietz 2004, Csáfordi et al. 2018, D'Ambrosio et al. 2018). Among these, much attention has been given to the relationship between industry actors and universities, showing that these links can be conducive to innovation and development (OECD 2019, Reichert 2019). Entrepreneurial ecosystems also highlight the dense interaction among a diverse set of actors in promoting entrepreneurial activity and innovation (Acs et al. 2017, Alvedalen–Boschma 2017).

Our paper fits into the intersection between these two broad topics: it maps the patterns of research cooperation across Europe at an institutional level, using network analytic techniques. We focus on the patterns of interaction between these institution types, especially universities and industry actors, having the opportunity to distinguish between them. Although several studies have focused on university-industry collaboration, the evidence reported about the patterns and the connections is largely based on regional-level case studies (Cantner–Graf 2006, Guan et al. 2005, Reichert 2019) or restricted to specific technological fields (Guan–Zhao 2013, Owen-Smith et al. 2002, George et al. 2002). Additionally, large-scale studies covering several fields and regions or countries rely on patent data (Balconi et al. 2004, Owen-Smith et al. 2002, OECD 2019). Our paper fills the gap by using data on research collaboration between European institutions and firms, which allows for the analysis of the entire network of collaboration, the embeddedness of actors in this network, and its role in innovation.

While the universities' role in regional development is long studied (Kotosz et al. 2015), several studies highlight the importance of the connections between universities (or higher education institutions, research institutions in general) and industry actors. Using Brazilian data, Rücker Schaeffer et al. (2018) show that innovation activity is higher in areas where universities are present. In a German sample of start-ups, Audretsch and Lehmann (2005) also conclude that higher output of universities (both in terms of students and publications) positively affect the rate of new local start-ups. Similarly, Maietta (2015) shows that the innovative activity of a sample of Italian food and drink companies was positively affected by cooperating with universities. Analysing at a regional level, Ponds et al. (2009) conclude that interregional collaboration networks between universities and companies (measured by co-publications) are important knowledge spillovers in the case of Dutch regions.

Other studies shade the simple positive effect of collaboration. D'Este and Immarino (2010) find that the university's geographical proximity and research quality affect the frequency of joint research collaboration between the institutions. However, Bruneel et al. (2010) emphasise the obstacles in university-industry collaboration, arguing that a wider set of collaborators and trust can overcome these obstacles in previous collaborations. Analysing survey data focusing on RIS3 strategies, Vallance et al. (2017) find that despite the connections between universities and companies, less innovative regions lag in their capability to use these connections as a matching point between scientific research and business activities.

The studies cited previously either use survey data or specific national databases on research collaboration. However, some studies access data on the EU-funded Framework Program collaboration. Caloghirou et al. (2001) analyse university-industry collaboration in Framework Programs and conduct a large-scale analysis between 1983–1996, showing that universities are included in more than 50% of cooperative projects and that their role increases with time. They also highlight that universities of peripheral countries are also important actors in this collaboration network. This result is reinforced by Sousa and Salavisa (2015) in a more recent paper on Portuguese collaborative projects, which shows that universities contribute significantly to these networks and they become increasingly more central.

Although not explicitly addressing university-industry collaboration, other studies provide an overview of Framework Program collaboration. Reillon (2017) gives a comprehensive overview of the different waves of the Framework Programs, while Schluga and Barber (2008) analyse the collaboration network across different FPs, finding that despite the changing programs, the network structure does not change too much. A stable core can be identified, mainly consisting of universities and research institutions. Akcomak et al. (2018) use FP data to draw a collaboration network between regions and show that less developed countries exhibit knowledge convergence.

In this paper, we emphasise that the collaborative ties between project participants of the Framework Programs constitute a network, and analysing the different properties of this network can provide useful insights. However, several studies used this approach to reveal the patterns of innovation. Using patent collaboration data, De Noni et al. (2018) emphasise that intraregional collaboration frequency does not affect patent applications in less innovative regions. However, extra-regional collaboration positively affects, especially in high-performance regions. Additionally, using patent collaboration data, Santoalha (2018) highlights that successful innovation systems are both locally and globally integrated, indicating a balanced set of collaborative ties in both directions. However, with patent data, Dosso and Lebert (2019) show that most central regions are strong innovators. Similarly, based on a survey in a Hungarian region, Juhász (2019) emphasise that spinoff companies are more likely to form local knowledge networks through a dense connection of collaborative ties. Fitjar and Rodriguez-Pose (2019) also show that local and international collaboration positively affect firm-level innovation.

A few studies explicitly consider the structure of the network among participating institutions. Ponds et al. (2009) work with a spatial econometric setup where collaboration networks are considered through the spatial weight matrix. More precisely, the weighted average of R&D of partner regions is considered detrimental to local innovativeness. However, this study relies on publication data as a measure of research collaboration, and the nodes of the network are regions. Although Akcomak et al. (2018) use data on Framework Program cooperation, they also set up the regional level collaboration network. Then they use different centrality measures to analyse convergence patterns of countries. However, Schluga and Barber (2008) analyse the participant-level network of Framework Program collaboration, focusing on the evolution of macroscopic properties (degree distribution, small world properties) of these networks. However, they do not address university-industry collaboration explicitly. Sousa and Salavisa (2015) analyse the network of Portuguese participants in FP projects through their centralities. Although they provide a network-approach to actor-level FP collaboration, their analysis is geographically limited.

Thus, our research extends this study line and provides a large-scale analysis of Framework Program collaboration patterns from a multidimensional perspective. We rely on a participant-level network in our analysis, covering all EU countries through the waves of FP5, FP6, and FP7. Distinguishing between different institutions, we address cooperation between research institutions (higher education institutions and other non-profit research institutions in general) and industry actors. We use these data to profile European NUTS3 regions through a cluster analysis where regions are grouped along with their collaboration patterns. We contribute to the literature by mapping the institutional and spatial dimensions of research collaboration while relying on a large-scale dataset of Framework Program collaborations covering European NUTS3 regions. The institutional dimension is focused on industry actors (companies) and research institutions (including universities): we map all different relations (industry-industry, research-industry, and research-research), with the primary focus on the collaboration patterns between research institutions (including universities) and industry actors.

Our approach is related to a strand of literature focusing on different proximities in shaping collaboration patterns (Broekel–Boschma 2017). Several studies emphasise the role of network distance (Fafchamps et al. 2010), technological proximity (Cunningham–Werker 2012), institutional proximity (Hardeman et al. 2012), academic excellence and informal communication (Jeong et al. 2016) or even similar affiliation background (Rodriguez–Pepe 2008) behind the strength of collaboration intensity. Others highlight the crucial role of geographical proximity in innovation collaboration (Broekel–Boschma 2017, Cunningham–Werker 2012, Hardeman et al. 2012, Autant-Bernard et al. 2007) as the melting pot of several types of proximities. Finally, some authors argue that frequent communication due to advanced info-communication technologies can substitute for missing geographical proximity. Thus, it is easy to establish cooperation between geographically remote actors (Boschma 2005, Hansen 2014). Our study directly refers to geographic distance by setting up the networks on a spatial (regional) basis and explicitly different between intra- and extra-regional collaboration links.

The rest of the paper is structured as follows. First, we discuss the construction of the database, with special emphasis on compiling FP collaboration data, followed by the methods of our clustering exercise. Some descriptive topological indicators are also calculated for the whole collaboration network and the industry and research institution sub-networks separately. Second, we provide the cluster analysis results, discuss the results, elaborate on the changing patterns of regional profiles and, finally, conclude the paper.

Data and method

We briefly discuss the data sources of the analysis, provide a background for the different indicators that we employ in the clustering exercise, and describe the methods used to profile the regions in our sample.

Data sources and network construction

Network data

The starting point of our analysis is the information available on EU-funded Framework Programmes, retrieved from the Cordis database. We use the information on all projects funded in the three waves of the Framework Programmes in this study, FP5, FP6, and FP7, which means that the data cover 1999–2013. The basic unit of these data is a project-participant pair, which means a particular institution (as a participant, e.g., university, company) is involved in a funded project. First, we use the projects' information: the contract numbers of the specific projects are used as unique identifiers. The duration (starting and ending years) of the projects allows us to have a longitudinal approach to the collaboration patterns. Second, information on the participants is used: their location, the NUTS3 level region they belong to, and the institution's type (e.g., higher education institution, industry actor). According to Pálóczi (2016), NUTS3 level are adequate for territorial economy systems than UTS2.

These data passed through two waves of data cleaning. First, we cleaned the regional classification. Although the Cordis dataset provides NUTS3-level categorisation of the participants, this is incomplete and with errors in several cases. We completely re-classification them based on the information on postal codes, addresses, and cities provided in Cordis. Additionally, manual checks were conducted to assign a clean regional code at the NUTS3 level to all institutions. Second, as Cordis's participant identifiers are problematic, we completely re-identified institutions, especially when used across different FP programmes. Using the information on the name, location (region), and address of the participants, we ran a string-matching algorithm to reveal every institution-pair similarity. The same procedure was done manually as well on a subsample of institutions. The latter provided reference-cases where we were sure about similar and different institutions. This reference subsample was then confronted with the algorithmic results to establish an ambiguity range. Institution-pairs with a similarity score below this range were assumed to be different; pairs above this range were assumed identical. Pairs falling into the ambiguity range were manually rechecked to arrive at a clean identification of institutions finally.

In the cleaned dataset, we derived information about every funded project, the project's duration, the participants, their location at the NUTS3 level, and their type being higher education institutions, research institutions, industry actors, or others. This analysis considers only the first three types with merging research and higher education institutions into one category. For simplicity, we will refer to the latter group as research institutions in general.

Before the data manipulation, we note that the data being used reflect an important but specific aspect of innovation (scientific) networks. First, the pooled connections recorded in this dataset reflect research cooperation. While it reflects how and where the generation of new knowledge is attempted, these records do not show whether these attempts are successful or not (in the form of scientific publications or patents). Also, the records are selective. We have information on funded projects, while unsuccessful applications and research collaborations without formal infrastructure are not visible.

Our starting point for data manipulation is the project matrix. The rows of this matrix correspond to institutions, whereas the columns represent projects. A given cell of the matrix would be one if the institution *i* was participating in project *k*. From this project matrix, simple matrix manipulation provides the adjacency matrix A for all years in our sample: $A=PP^T$, where P^T is the transpose of P^T. The resulting A adjacency matrix provides the number of ongoing joint projects between any pair of institutions. This adjacency matrix gives a snapshot of collaboration patterns between institutions with a weighted perspective: we consider the number of joint projects, reflecting the intensity of collaboration, as our starting point for further calculations.

The adjacency matrix A contains information between all pairs of institutions, regardless of their location (region) and type (research institution or industry actor). To explain these features, we use two categorisation vectors. d^T refers to the type of institutions: it has one entry (row) for all institutions. It contains 1 if the given institution is a research institution and 2 if it is a company or an industry actor. Similarly, d^R refers to the location of institutions and one entry (row) contains the index of the region to which the institutions belong.

To ease further exposition, we reshape the adjacency matrix A into an array W which structures connections between institutions and their location and type. Its general element is defined as follows:

$$w_{rfi,qgj} = a_{l_1 l_2} | d^R_{l_1} = r, d^R_{l_2} = q, d^T_{l_1} = f, d^T_{l_2} = g$$

Here, $W_{rfi,qgj}$ describes the number of collaboration projects between institution *i* of type *f* in region *r* and institution *j* of type *g* in region *q*. Here the indices $f,g=1,2$, indicating whether institutions are companies (1) or research institutions (2). Then, $r,q=1,2,\ldots,R$ refer to region indices, while $i,j=1,2,\ldots,I_{f,r}$ reflect the indices of institutions. Note that $I_{f,r}$ is different for all region *r* and institution type *f*, representing the number of institutions of the given type in the given institution.

Analogous to the structure W, representing the weights (intensities) of collaboration between any two institutions, we define the binary version of this structure, B, representing the existence of collaboration between institutions in period *t*, rather than their intensity:

$$b_{rfi,qgj} = \begin{cases} 1, \text{if } w_{rfi,qgj} > 0 \\ 0, \text{otherwise} \end{cases}$$

Here the indices have the same meaning as in the weighted connections.

This notation, using multidimensional structures instead of one large adjacency matrix, allows for a simple exhibition of calculations behind the collaboration properties we analyse. We construct these collaboration indicators along two dimensions. Along the first dimension, we distinguish the relationship of institutions as being local or global. Inter- and intraregional relationships play a different role in knowledge spillovers, and we use this differentiation in our study. Along the second dimension, we differentiate relationships by the type of institutions that participate in the collaboration. As we are focusing on two types of institutions, three categories of relationships appear along this dimension: (i) both institutions are companies, (ii) both institutions are research institutions, (iii) one institution is a company, the other is a research institution. If we consider all possible cases by the two dimensions, we can calculate seven different versions of all collaboration properties for every region.[1] These versions are summarised in Table 1.

The characteristics of a network can be described in many ways by different indicators. In this study, we focus on the collaboration intensity of institutions, then briefly discuss some other structural properties, while other network indicators can be covered in further analyses. We evaluate collaboration intensity with three related measures. First, the simple number of collaborative projects reflects how intense the cooperation is between any two institutions – this is called *strength*. This is the weighted degree centrality of all institutions aggregated at the regional level by different dimensions. We use the weighted connections of institutions in a given region and add them to get an overall measure of connection strength at the regional level. Formally, we have:[2]

$$(1)\ S_r^{in,fg} = \sum_{i,j} w_{rfi,rgj}$$

$$(2)\ S_r^{out,fg} = \sum_{q \neq r,i,j} w_{rfi,qgj}$$

Here, $S_r^{in,fg}$ refers to the strength of connections between institutions of type *f* and type *g*, both belonging to region *r*. Thus, it refers to intraregional connection

[1] We have 7 versions, as within one region the relationship between two institution types are symmetric, while across regions it is not necessarily the case. More precisely, the collaboration intensity between research institutions within a region is the same as between research institutions and companies, however there is a difference between cooperation intensity of local companies with extra-regional research institutions and local research institutions with extra-regional companies.

[2] In equation (1), if *f*=*g* then the right-hand-side must be divided by 2 as in this case we count every link within the same institution type and the same region twice.

strength. The formulae $S_r^{out,fg}$ refer to the strength of connections between institution type *f* in region *r* and institution type *g* outside region *r*.

Second, we also calculate the *density* of cooperation: the number of connections observed in the network compared to the maximum possible number of connections. The density indicators are calculated using the binary cooperation patterns recorded in the structure B. Density is then calculated in an analogous way to strength:[3]

$$(3)\ D_r^{in,fg} = \frac{\sum_{i,j} b_{rfi,rgj}}{I_{f,r} I_{g,r}}$$

$$(4)\ D_r^{out,fg} = \frac{\sum_{q\neq r,i,j} b_{rfi,qgj}}{I_{1,r} \sum_{q\neq r} I_{1,q}}$$

Although strength and density are straightforward ways to measure cooperation intensities, we have to consider that the strength will be very low in small regions with few actors. In contrast, density can be very high due to the nature of these indicators. In large regions, we face the opposite problem. Therefore, we include a third measure, the *average strength* of collaboration, the average number of connections. Formally, we get average strengths by

$$(5)\ \hat{S}_r^{in,fg} = \frac{\sum_{i,j} w_{rfi,rgj}}{I_{f,r}}$$

$$(6)\ \hat{S}_r^{out,fg} = \frac{\sum_{q\neq r,i,j} w_{rfi,qgj}}{I_{f,r}}$$

Here the notation resembles that used for density and strength.

Table 1

Summary of collaboration indicators by type and location of participants

	Local company	Local research institution	Global company	Global research institution
Local company	$D_r^{in,11}, S_r^{in,11}, \hat{S}_r^{in,11}$	$D_{r,t}^{in,12}, S_r^{in,12}, \hat{S}_r^{in,12}$	$D_r^{out,11}, S_r^{out,11}, \hat{S}_r^{out,11}$	$D_r^{out,12}, S_r^{out,12}, \hat{S}_r^{out,12}$
Local research institution	Same as Local company vs. Local research institution	$D_r^{in,22}, S_r^{in,22}, \hat{S}_r^{in,22}$	$D_r^{out,21}, S_r^{out,21}, \hat{S}_r^{out,21}$	$D_r^{out,22}, S_r^{out,22}, \hat{S}_r^{out,22}$

As summarised in Table 1, we have three indicators for all types of relationships describing three different aspects of collaboration strength. Although these three indicators focus on different aspects, higher values always imply more intensive relationships. Therefore, we use the three indices (strength, density, average strength) together and control for the number of institutions in different regions.

[3] In equation (3), if $f=g$ then the denominator becomes $I_{f,r}(I_{f,r}-1)$ as we do not count self loops (an institution's connection to itself) as a possible link.

To compress information, we present the results of a composite indicator of the three different values: density, strength, and average strength. This indicator is the simple average of standardised (mean 0 and standard deviation 1) values. However, the clustering analyses are run on the three indicators separately.

In addition to the indicators derived (and to be used in the clustering exercise later), we exploit the information on the whole network of institutions and describe this network through some indicators about this network's topology. We calculate the following metrics:

– Size: which is the number of nodes in the network (*N*).
– Density: the number of edges relative to the maximum possible number of edges ($\sum_{rfi,rgj} b_{rfi,rgj} /(N^2 - N)$).
– Average (weighted) degree: the number of collaborative projects an average institution has ($\sum_{rfi,rgj} w_{rfi,rgj} /N$).
– Transitivity: the tendency of the network to form closed triangles (the number of closed triangles relative to the maximum possible number of such triangles (see Barrat et al. 2004).
– Modules: the number of separate components of the network (see Blondel et al. 2008).
– Size of the giant component: the share of nodes in the largest component (module) relative to the total number of nodes.
– Degree distribution: the estimated exponent of the degree distribution (see Newman 2007).

Data on the level of development

Previously, we discussed those measures used to capture the collaboration patterns among institutions. These indicators are augmented by information on the development level and the regions' innovation capacity to provide a more comprehensive view of collaboration patterns. The development level is measured by per capita GDP, while the innovation activity is captured by patent per head. These two indicators are available at the NUTS3 level for a large set of European countries. Thus, we can add it to the collaboration measures, which are also calculated at this regional disaggregation.

The GDP per capita data are retrieved from the OECD database for 2000–2013. This database contains a moderate amount of missing data that were replaced using the Eurostat database. For two regions (LT025 and CY000), we used the GDP per capita data from the Eurostat database at current prices and applied the 2015 PPP and the GDP deflator (with the base year 2015) to make it consistent. In the case of Cyprus, we converted the base year of the deflator from 2010 to 2015 prices. For some Polish regions, we used NUTS 2 data. In Switzerland, data are only available at the national level until 2008, so we used the average national data for the regions between 2000 and 2007.

In the case of innovation activity, data are retrieved from the Eurostat database in patents per million inhabitants. These data are available for the majority of the European NUTS3 regions. In cases where data are missing, but the patent count is available, we calculated the patent per million inhabitants using the earliest available population data. When the patent counts indicated low patenting activity, per capita patents were replaced by zeros in cases of missing data. Finally, in the rest of the cases with missing data, we used information on the NUTS2 level. This estimation was employed when regions were either not included in the Eurostat database or had no available population data while the patent count was significantly different from zero.

The final dataset

After the data preparation process, we obtained an extensive database of collaboration links between different institutions. The main characteristics of these data are summarised in Table 2. Although the CORDIS data classify institutions into industry, higher education, research institutions, government institutions, and others, in this analysis, we use only industry and higher education plus research institutions together and omit government institutions and those participants classified as other. The variables used in our analysis covers most of the data.

Table 2

Summary statistics of collaboration data

	Total in database	Used in the analysis
Number of participants	56,597	56,473
Number of projects	51,187	
Number of industry actors	27,509	27,474
Number of higher education and research institutions	10,561	10,527
Number of regions	1,419	1,378

Based on this collaboration data, we constructed the collaboration indicators summarised in Table 1. We augmented them with GPD and patent data for a more comprehensive dataset at the regional level. The descriptive statistics of these indicators can be found in Table A1 in the Appendix.

Table 3 shows the topological indicators of the cooperation network, as discussed in this paper. The table shows average values over the 1999–2013 period, the framework programs FP5, FP6, and FP7. These indicators reflect the macroscopic characteristics of the whole network of institutions. We employ detailed information about the type of institutions and present these indicators for three different cases: (i) the network between industry actors (companies), (ii) the network between higher education institutions or research institutions, and (iii) the network between both of these institutions. In terms of the previously introduced notation, in case (iii), we use the whole W matrix; in case (i), those rows and columns of W are extracted for which *r,q*=1; and in case (ii), those row and columns are extracted for which *r,q*=2.

Table 3

Global topological indicators of the cooperation network (average values over the period 1999–2013)

	Research institution network	Industry network	Whole network
Size (nodes)	10 561	27 509	38 070
Density, %	0.2105	0.0106	0.0332
Average (weighted) degree	32.0052	3.1866	16.1470
Transitivity	0.2516	0.4234	0.1892
Number of modules	6 366	20974	26630
Share of nodes in the largest component, %	39.68	20.83	29.94
Power law exponent	1.8783	2.8913	2.0339

The results show that these networks are quite sparse: density is lower than 1% in all cases. However, the density of the network among research institutions tends to be significantly greater than that of companies. This reveals that universities and other research institutions provide the connecting core of this cooperation network. Concurrently, industry actors are typically less connected with fewer joint projects and cooperate with only a few partners. This is also reflected by the average degree (average number of projects): research institutions are involved in more projects than industry actors. Transitivity measures the extent to which nodes tend to cluster in smaller, tightly connected groups. While the industry actor network's density is much lower than that of research institutions, transitivity moves in the other direction. As density is a natural reference point for interpreting transitivity (it is more likely to find closed triangles in a denser network), we can conclude that industry actors tend to cluster more intensively than research institutions. The fact that transitivity is significantly lower in the whole network than in the separated networks shows that closed triangles (clustering) tends to occur within the same type of actors, i.e., it is more likely to find similar institution types (either companies or research institutions) on the three nodes of these triangles. Although the networks' density is significantly above the threshold where a giant component emergence is expected, the largest connected component is moderate in all three cases. Consistent with the higher average degree, roughly 40% of the nodes belong to the largest component in the research institution network. However, we would expect nearly all nodes to be part of the giant component in a random network. The average degree of the industry network is much closer to the threshold value of one, and the size of the largest component is smaller but still significantly lower than expected in a random network. Finally, we fitted a power law on the degree distributions of these networks. We found exponents for these power laws close to 2, indicating that these networks show scale-free properties. The exponent in the industry network is higher, revealing that this network is more polarised between small, weakly connected actors and hubs.

Profiling through clustering

This study aims to provide a mapping of European regions according to their collaboration patterns and development level. We accomplish this goal by running a standard cluster analysis based on the indicators introduced so far. The result provides a grouping of regions into relatively homogeneous clusters where collaboration patterns and the development levels are relatively similar.

The most common clustering techniques are the k-means and the hierarchical clustering methods (MacQueen 1967, Hartigan–Wong 1979, Kodinariya–Makwana 2013). The k-means clustering is suitable in cases where outliers are a problem; however, it provides different results for every calculation because of this method's random initial state. Hierarchical clustering gives consistent results; however, it is sensitive to outliers. Considering the indicators' descriptive statistics and the presence of outliers in our sample, we apply the k-means clustering technique. This algorithm classifies regions so that the Euclidean distance between normalised indicator values of a given region is the closest to the group-centre among all groups (Hartigan–Wong 1979). After the necessary standardisation (zero mean and unit standard deviation), the algorithm has the following steps (MacQueen 1967): (i) we determine the number of clusters k to use; (ii) the algorithm randomly creates k clusters and determines the centres; (iii) it adds the observations to the group the centre of which is the closest to the observation; (iv) it recalculates the centres of groups; (v) it repeats points (iii) and (iv) until the classification does not change. Due to the random initial conditions, we repeat the process 100 000 times. We use a fitness measurement to select the best grouping by minimising the total within-cluster sum of squared distances (*TSS*) between observations and cluster centres (Hartigan–Wong 1979). The final point is to provide an accurate number of clusters for the algorithm; however, there is no unambiguously optimal method to determine this value. We use Elbow-method, the simplest and most practical solution (Kodinariya–Makwana 2013). This method also uses the *TSS* and determines the optimal number of clusters where adding one more cluster does not decrease the *TSS* significantly. The results of this method can be found in Figure A1 in the Appendix.

Results and discussion

We present and discuss the clustering exercise done with the method and indicators described earlier. In the clustering exercise, we used all collaboration and development indicators separately. However, for the sake of conciseness, the results are presented with the collaboration indicators grouped into the seven composite indicators along the different connection types shown in Table 1.

Table 3 shows the clustering exercise results where all collaboration indicators are included together with GDP and patent per capita. The table entries reflect the extent to which the given indicator (column) in the given cluster (row) is above or below the average. These values represent a standard normal distribution where the zero means reflect the overall average and the standard deviation is one. Values lower than zero thus reflect below-average cluster mean in the given indicator, while values higher than zero reflect above-average cluster mean. The first three indicators refer to intraregional collaboration intensities regarding the two institution types, and the next four show extra-regional connection intensities. In the latter case, industry-research institution relationships can differ depending on whether local industry actors cooperate with research institutions outside the region or vice versa. The three columns at the right-hand side refer to the GDP per capita, patent per capita, and the number of regions belonging to the given cluster. Negative values show below average. Positive values show above-average performance. Shading shows the extent to which values are below average (blue) or above average (red).

Table 4

Characteristics of clusters

Cluster	Intra-regional connections			Inter-regional connections				Development level		Number of regions
	I – I	*U – U*	*I – U*	*I – I*	*U – U*	*I (loc) – U*	*U (loc) – I*	*GDP*	*Patent*	
A	–0,3308	–0,3439	–0,3677	–1,1047	–0,6620	–1,1357	–0,6578	–0,8956	–0,9131	312
B	–0,2159	–0,3467	–0,3903	–0,0819	–0,6720	–0,1556	–0,6534	0,1227	0,3999	375
C	7,9405	–0,3511	–0,4111	–0,2204	–0,4827	–0,5102	–0,7223	–0,7813	–0,4637	2
D	–0,0576	–0,3458	–0,3508	1,3245	–0,5362	1,2901	–0,5356	0,1675	0,2244	113
E	–0,0799	–0,0861	–0,0228	0,1487	0,5152	0,1738	0,4417	0,0499	0,0058	313
F	1,3116	–0,3511	4,8581	–0,7359	0,4299	–0,4376	0,4343	0,1065	–0,3394	6
G	–0,1954	3,1659	0,0860	–0,3839	0,6536	–0,3938	0,3463	–0,2036	–0,0410	13
H	–0,1701	0,6345	0,5568	0,1032	2,1289	0,1672	2,2774	0,2905	0,2291	53
I	0,6757	0,6896	0,9242	0,7771	1,1696	0,9267	1,1860	0,8916	0,5703	131
J	1,9163	2,3501	2,1307	1,3628	1,3619	1,5171	1,5132	1,1708	0,4066	60

The algorithm extracted 10 clusters in this case. Figure A1 in the Appendix shows how the *TSS* value changes with the number of clusters, and it shows a fracture at 10 clusters, which clearly shows that adding more clusters is not meaningful. The clusters are roughly ordered from A to J, so that cluster A is very sparsely connected while cluster J is very densely connected in all respects. We briefly characterise all clusters in what follows. We can form three main groups from the ten clusters that give the following discussion's organising structure.

Group I – Non-cooperative regions. This group contains clusters A, B, and C, 689 regions altogether. All of them are weakly connected; however, there are some differences.

- **Cluster A – Non-cooperative, less developed**. In this cluster, regions do not show intensive cooperation either within or outside the region. This is a relatively large group with 312 regions, and as marked by the below-average values of GDP and patenting, these are also the less developed regions.
- **Cluster B – Non-cooperative, developed**. This cluster also shows typically low interaction strength in all types of collaboration. However, compared to cluster A, entries are less negative; thus, institutions have a denser interaction in this group than those in cluster A. But the most important differentiating factor in cluster B is that both GDP per capita and patent per capita are above-average in this group. So this is a group of relatively developed regions, while their embeddedness in collaboration networks is sparse. This is also a large group with 375 regions.
- **Cluster C – Non-cooperative, less developed with strong local industry**. This group is similar to cluster A, with all but one indicator showing significantly below-average values. This group also contains less developed regions with low cooperation levels, except intraregional collaboration within industry actors. The latter type of collaboration is extremely high in this group: the local industry-industry cooperation is the highest in the whole sample. However, only two regions belong to this cluster, which means that while the extreme local industrial cooperation is a natural reason for this group to be separated, it is more realistic to treat these two regions as a special subgroup of cluster A.

Group II – Externally focused, moderately developed regions. This group contains clusters D and E with 426 regions.

- **Cluster D – Externally focused, industry-based.** We find regions in this cluster where companies are the predominant networks; however, cooperation is typically extra-regional. While intra-regional cooperation is below-average in all types, local companies have strong cooperation with both companies and research institutions outside the region. This is a relatively large cluster with 113 regions, relatively developed, but not among the most developed ones. However, research institutions have sparse connections in this type of region, both locally and externally.
- **Cluster E – Externally focused, research-based.** This cluster is similar to cluster D because local cooperation is weak in all respects, but external

cooperation is above average. However, the basis of external cooperation is shifted: while in cluster D, local industry actors show strong collaboration, in cluster E, local research institutions provide the basis for external links. However, local companies still show above-average external cooperation. Still, this cooperation's strength is not that high as in cluster D. This is one of the most sizeable groups with 313 regions. While overall, these regions show a moderate development level, compared to cluster D, both GDP per capita and innovation are lower.

Group III – Cooperative regions. This group contains clusters F to J, with 263 regions.

– **Cluster F – Locally industrial, externally research-based.** This cluster shows intensive local cooperation; it is built around local companies. While industry-industry and industry-research institution cooperation is high, we find weak connections between local research institutions in this cluster. However, local research institutions show above-average connections for external links, while local companies are weakly connected. Although not sizeable, this group shows slightly above-average GDP per capita levels but low patenting activity.
– **Cluster G – Research-based, less developed.** Research institutions dominate this cluster. While we see strong local connections between research institutions, other local connections are weak. Research institutions also dominate external connections. This cluster consists of less developed regions (13 regions).
– **Cluster H – Research-based, developed.** This cluster is similar to cluster G; however, it consists of more developed regions: GDP and patent per capita are significantly above average. This is the first cluster where almost all collaboration indicators are above average, except local intra-industry cooperation. However, the local research institutions dominate the cooperation networks, both internally and externally: the external cooperation intensity of research institutions are among the highest in the whole sample.
– **Cluster I – Cooperative, developed.** In this quite a large cluster (131 regions), we find developed regions with strong cooperative patterns along all types of collaborations. Although research institutions (primarily in external links) still show some dominance, cooperation patterns are quite homogenous in this cluster.
– **Cluster J – Super cooperative, developed**. This group is similar to cluster I with even stronger networks. Local collaboration is strong, while these regions also show a higher GDP per capita; however, patent per capita is slightly smaller than in cluster I. The difference between cluster J and cluster I is more quantitative than qualitative.

The general picture of this clustering exercise is that the majority of the regions show homogenous networking patterns. 50% show below-average, while 18% show above-average collaboration intensity in all connection types with a few exceptions. While those regions which show homogenously above-average collaboration intensities are almost exclusively in developed regions (above average GDP and patent per capita). A significant group (27%) of relatively developed regions with below-average collaboration intensities in all types exists.

Along with these homogenous regions, a significant 32% of regions show selective collaboration patterns. These are typically moderately developed regions and are frequently characterised by a university dominance, especially in the regions which belong to the less developed segment. The majority of these moderately developed regions belong to cluster D (8%) and cluster E (23%), which show external focus with weak intra-regional collaboration. However, the relatively more developed (and especially innovative) cluster D relies on local companies that keep extra-regional connections. However, the relatively less developed cluster E relies mainly on local universities and research institutions, embedded in extra-regional cooperation.

The collaboration between industry actors and research institutions is of particular interest in this study. This clustering shows that local interaction between research institutions and companies are typically concurrent with extra-regional cooperation. Those regions with dense internal interactions between the two types of actors also show strong cooperation between local research institutions and outside companies. Cooperation between local companies and outside actors is less frequent: apart from the two strongly connected clusters, there is only one cluster (cluster D), which shows intensive cooperation between local companies and external research institutions. Companies rarely reach out to external partners. There is only one cluster in which this is observed apart from the two strongly cooperative clusters, where, naturally, all connection types are strong. However, external links are mainly dominated by research institutions.

Regarding local industry-research collaboration, we see four clusters where this collaboration type is significantly above average: clusters I and J, strongly cooperative regions with all types being intensive, and clusters F and H. In cluster H, the local network is dominated by research institutions, with local intra-industry cooperation being sparse. Cluster F seems to be the opposite, where local cooperation between research institutions is sparse; however, industry-industry and industry-research links are strong. While externally, both clusters are dominated by connections of research institutions, the locally research-based cluster H belongs to the more developed regions. However, the local industry-based cluster F contains slightly above-average GDP per capita and below-average patenting activity. However, the full picture must contain cluster H containing 5% of the regions; there are only six regions (0.6%) in cluster F.

As the k-means clustering method is dependent on the pre-determined number of clusters, we also run a robustness check and calculate for nine and eight clusters. The results of these calculations can be found in Tables A2 and A3 in the Appendix, similar to Table 3 for ten clusters. The clusters seem to be robust—in particular, the large clusters remain the same, and some smaller clusters are merged when the number of predefined clusters decreases. This shows that the main picture from the clustering presented in Table 3 is robust for the main categories and the large groups. In contrast, it can still highlight some specific groups with characteristics different from the large ones.

Figure 1

Map of European NUTS3 regions according to their FP cooperation patterns, group level

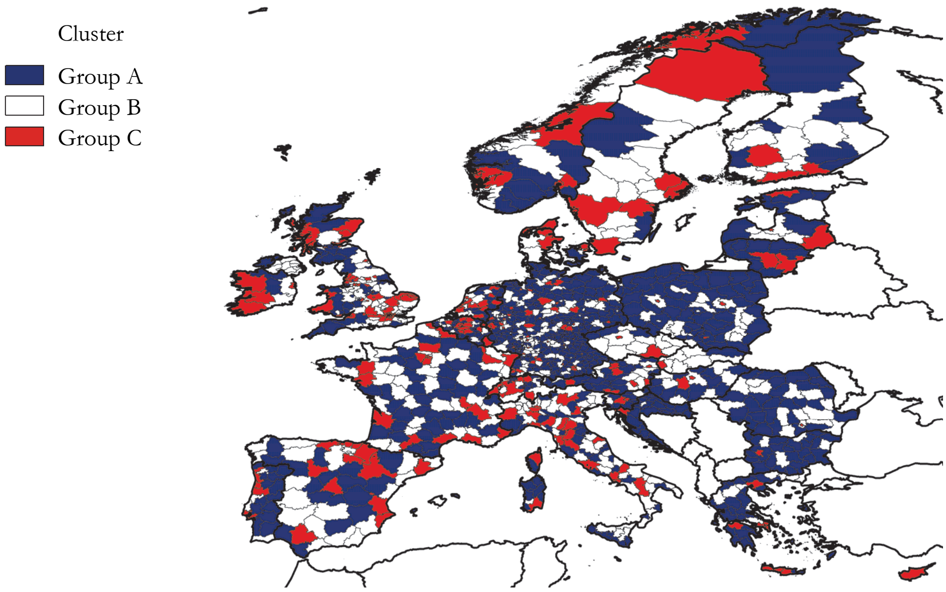


Figure 1 augments the previous analysis by showing the map of European NUTS3 regions with the aggregated results of the cluster analysis: regions are coloured according to the three large groups they belong to, consistent with the colouring in Table 4; the blue shades refer to the less cooperative regions, while red shades reflect the more cooperative ones.

Figure 1 shows that while blue regions are dominant in Eastern Europe, there are also many of these regions in Western Europe. This reinforces the findings in Table 3: cluster A contains non-cooperative and less developed regions. These mainly correspond to the Eastern regions. Cluster B, however, consists of regions that show the above-average level of development while they are still non-cooperative. Figure 1 shows that the latter regions are also scattered around Western

Europe, although these are not the most central or developed parts. However, it is rare to find red, i.e., above-average cooperative regions in the Eastern parts. These are mainly capital regions as Warsaw, Bratislava, Sofia, and Ljubljana. The areas around some relatively industrialised cities such as Krakow in Poland, Brno in the Czech Republic, and Vilnius and Kaunas in Lithuania. However, several Eastern Europe regions belong to Group II, with moderate development level and primarily externally oriented collaboration patterns.

A more detailed picture can be found in Figures A2 to A5 in the Appendix, which shows the ten different clusters (A2 show all clusters while A3, A4, and A5 show clusters of the three main groups). These pictures reinforce that cluster A contains regions from Eastern Europe, while cluster B contains Western Europe regions. Within the moderately developed regions in Group II, there are two sizable clusters D and E, with an external collaboration focus: local cooperation is weak in these regions; however, external cooperation is strong. However, while in cluster D (which is more developed on average), local companies dominate these networks, local universities and research institutions dominate cluster E. The observations show that the university-based model characterises the more developed part of CEE countries (especially the Czech Republic, Slovakia, and Hungary), a significant part of Scandinavian countries, and less central regions in Western Europe. Cluster D, with an industrial focus on external cooperation, can be found all across Europe. Still, these regions are found more frequently in Germany, Poland, Romania, and Italy.

Conclusions

This paper uses a unique dataset to map research collaboration patterns across European regions. This dataset, building on information in collaborations in Framework Program projects, allows us to draw the network of collaboration along an institutional and spatial dimension. While regarding the former, we focused on industry actors and research institutions (universities) as two main types of institutions and the collaboration among them. In the latter, we could reach a relatively detailed, NUTS3 regional level. This institutional detail provides an opportunity to focus on the collaboration patterns between industry actors (companies) and research-focused actors (universities, research institutions), the subject of several prior studies.

Using this dataset, we calculated different collaboration intensity indicators at the regional level, and then we employed cluster analysis to provide a map of collaboration patterns across Europe. In this cluster analysis, we integrated indicators of the development level and innovative activity of regions to gain a detailed picture.

The main finding is that cooperation intensities typically correlate to types of collaborations (institutional and spatial dimensions): most regions are either weakly or strongly cooperative along most of the cooperation dimensions. However, there is some selectiveness in this respect.

First, it became clear that while development moves together with cooperation intensity, there is a large group of developed (typically Western European) regions that are weakly cooperative. Second, there is a heterogeneous group of regions between cooperative and non-cooperative ones in the middle of the development scale. Their cooperative patterns are selective, either institutionally or spatially. In the latter group, we found that most of the regions are externally focused, with strong external collaboration intensities and weak local ones. The research institutions dominate external focus in most cases. Still, several regions base their external collaboration on local industry actors.

Our results are threefold regarding the specific collaboration pattern between industry and research institutions (universities in particular). First, consistent with the correlation mentioned, these specific collaboration links across different types of actors seem to systematically appear together with other types of cooperation: those regions show strong research links across the two types of actors, which are also strongly cooperative in other dimensions. Within-region collaboration between industry actors and research institutions is found to be rare outside strongly cooperative and developed regions. Extra-regional collaboration between the two different institution types is frequent; however, local universities cooperate with companies across the borders. Still, some regions show an industry-based cooperation network.

There are three lines along which this research can be extended. First, a longitudinal analysis of these cooperation patterns is viable, showing different regions between different groups or clusters of collaboration patterns. Second, using econometric techniques, these data can infer the role of the institutional and spatial dimensions of collaboration patterns in shaping regional innovativeness. While our results highlight the differences in network embeddedness of regions with similar development levels and the strong correlation in the strength of different dimensions of collaboration, they have limited capabilities to drive policy decisions. However, further research along the previously mentioned line should help decision-makers design policies tailored to the specific needs and circumstances of regions to boost innovation through cooperation. Third, pointing again to the limitations of the framework program data used in this study, the analysis can be extended to other forms of collaboration like publication or patenting networks.

Acknowledgements

This publication/research has been supported by the European Union and Hungary and co-financed by the European Social Fund through the project EFOP-3.6.2-16-2017-00017, titled "Sustainable, intelligent and inclusive regional and city models".

Appendix

Table A1

Descriptive statistics of network and development indicators (full sample, 1999–2013)

	Number of regions	Mean	Standard deviation	Maximum	Minimum
$D_{r,t}^{in,11}$	1,378	0.0024	0.0145	0.3333	0.0000
$D_{r,t}^{out,11}$	1,378	0.0001	0.0001	0.0007	0.0000
$D_{r,t}^{in,22}$	1,378	0.0051	0.0196	0.4000	0.0000
$D_{r,t}^{out,22}$	1,378	0.0011	0.0018	0.0174	0.0000
$D_{r,t}^{in,12}$	1,378	0.0028	0.0109	0.2000	0.0000
$D_{r,t}^{out,12}$	1,378	0.0002	0.0002	0.0012	0.0000
$D_{r,t}^{out,21}$	1,378	0.0001	0.0002	0.0021	0.0000
$S_{r,t}^{in,11}$	1,378	0.9686	4.9116	81.5333	0.0000
$S_{r,t}^{out,11}$	1,378	61.6364	179.2731	2,702.7330	0.0000
$S_{r,t}^{in,22}$	1,378	1.9103	15.9753	521.4000	0.0000
$S_{r,t}^{out,22}$	1,378	241.1269	818.6080	17,363.7300	0.0000
$S_{r,t}^{in,12}$	1,378	1.9904	11.1553	276.2667	0.0000
$S_{r,t}^{out,12}$	1,378	66.5479	200.7637	3,586.4000	0.0000
$S_{r,t}^{out,21}$	1,378	66.5363	226.3616	4,143.6670	0.0000
$\hat{S}_{r,t}^{in,11}$	1,378	0.0303	0.0664	0.6216	0.0000
$\hat{S}_{r,t}^{out,11}$	1,378	2.2002	2.0402	22.4952	0.0000
$\hat{S}_{r,t}^{in,22}$	1,378	0.1067	0.3484	4.5737	0.0000
$\hat{S}_{r,t}^{out,22}$	1,378	15.1038	27.3401	283.9667	0.0000
$\hat{S}_{r,t}^{in,12}$	1,378	0.0271	0.0577	0.7120	0.0000
$\hat{S}_{r,t}^{out,11}$	1,378	2.1769	2.0567	23.7429	0.0000
$\hat{S}_{r,t}^{out,21}$	1,378	4.1844	7.4991	65.8667	0.0000
GDP per capita	1,378	34,770.8800	21,207.9400	468,013.5000	7,315.3570
No. of patent	1,378	126.2648	174.5742	1,964.7420	0.0000

Table A2

Characteristics of clusters, with 9 clusters pre-set

Clus-ter	Intra–regional connections			Inter–regional connections				Development level		Number of regions
	I – I	*U – U*	*I – U*	*I – I*	*U – U*	*I (loc) – U*	*U (loc) – I*	*GDP*	*Patent*	
A	–0,3279	–0,3420	–0,3373	–1,1024	–0,6529	–1,1302	–0,6535	–0,8975	–0,9174	316
B	–0,2159	–0,3466	–0,3903	–0,0801	–0,6723	–0,1575	–0,6536	0,1155	0,3920	375
C,F	4,9729	–0,3511	3,4765	–0,3246	0,3615	–0,3782	0,3234	0,0145	–0,2575	5
D	–0,0576	–0,3458	–0,3508	1,3245	–0,5362	1,2901	–0,5356	0,1675	0,2244	113
E	–0,0811	–0,0871	–0,0097	0,1503	0,5185	0,1821	0,4486	0,0811	0,0344	314
G	–0,1954	3,1659	0,0860	–0,3839	0,6536	–0,3938	0,3463	–0,2036	–0,0410	13
H	–0,1701	0,6345	0,5567	0,1032	2,1289	0,1672	2,2774	0,2905	0,2291	53
I	0,6844	0,7013	0,9287	0,7825	1,1734	0,9329	1,1911	0,8747	0,5584	129
J	1,9163	2,3501	2,1307	1,3628	1,3618	1,5171	1,5132	1,1708	0,4066	60

Table A3

Characteristics of clusters, with 8 clusters pre-set

Clus-ter	Intra–regional connections			Inter–regional connections				Development level		Number of regions
	I – I	*U – U*	*I – U*	*I – I*	*U – U*	*I (loc) – U*	*U (loc) – I*	*GDP*	*Patent*	
A	–0,3281	–0,3359	–0,3376	–1,1021	–0,6524	–1,1306	–0,6536	–0,8932	–0,9137	317
B	–0,2167	–0,3398	–0,3904	–0,0814	–0,6709	–0,1567	–0,6520	0,1185	0,3956	377
C,F	4,9729	–0,3511	3,4765	–0,3246	0,3615	–0,3782	0,3234	0,0145	–0,2575	5
D	–0,0576	–0,3458	–0,3508	1,3245	–0,5362	1,2901	–0,5356	0,1675	0,2244	113
E	–0,0872	–0,0718	–0,0115	0,1473	0,5427	0,1730	0,4669	0,0755	0,0292	321
H	–0,1480	1,2019	0,5726	0,0327	1,9669	0,1210	2,0564	0,2631	0,2102	58
I	0,7004	0,6907	0,9379	0,7956	1,1696	0,9449	1,1967	0,8640	0,5549	127
J	1,9163	2,3501	2,1307	1,3628	1,3618	1,5171	1,5132	1,1708	0,4066	60

Figure A1

The optimal number of clusters with the elbow method

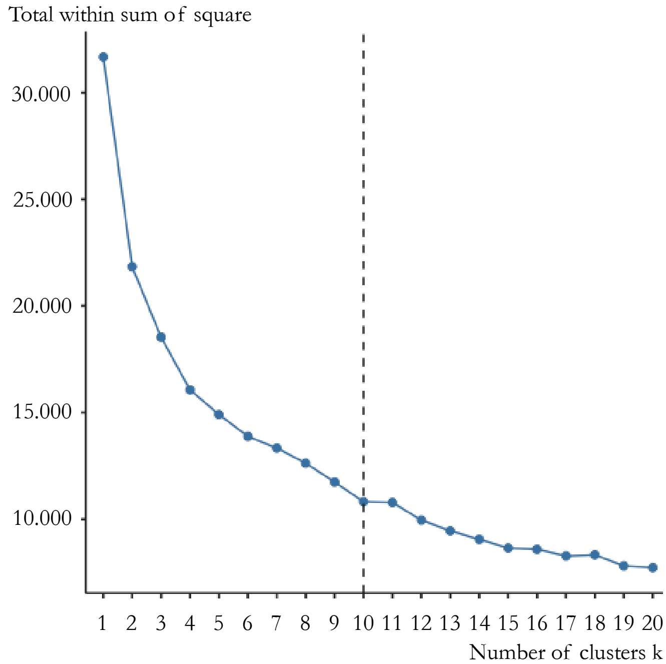


Figure A2

Map of European NUTS3 regions according to their FP cooperation patterns, all clusters (A to J) included

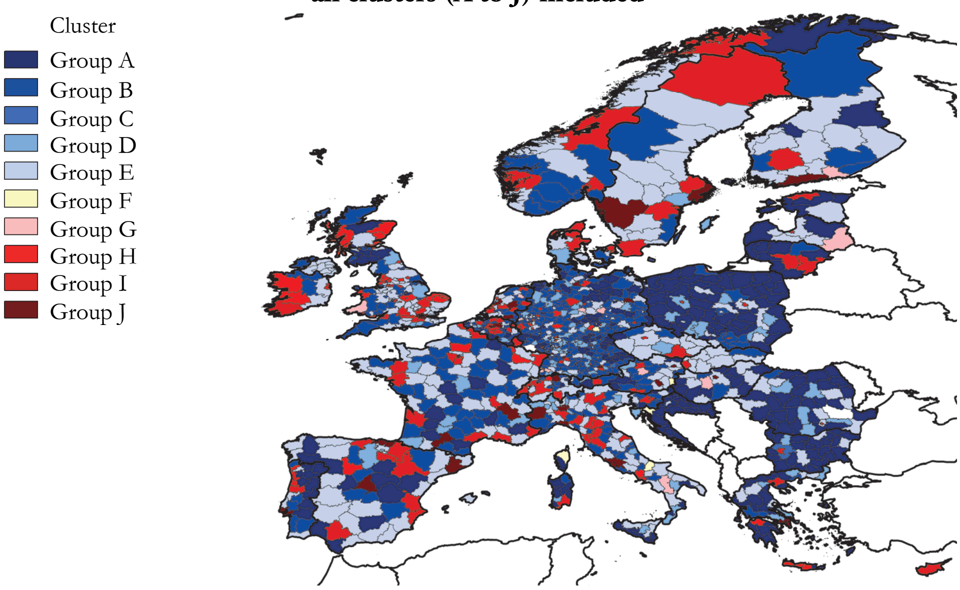


Figure A3

Map of European NUTS3 regions according to their FP cooperation patterns, Group I clusters (A to C)

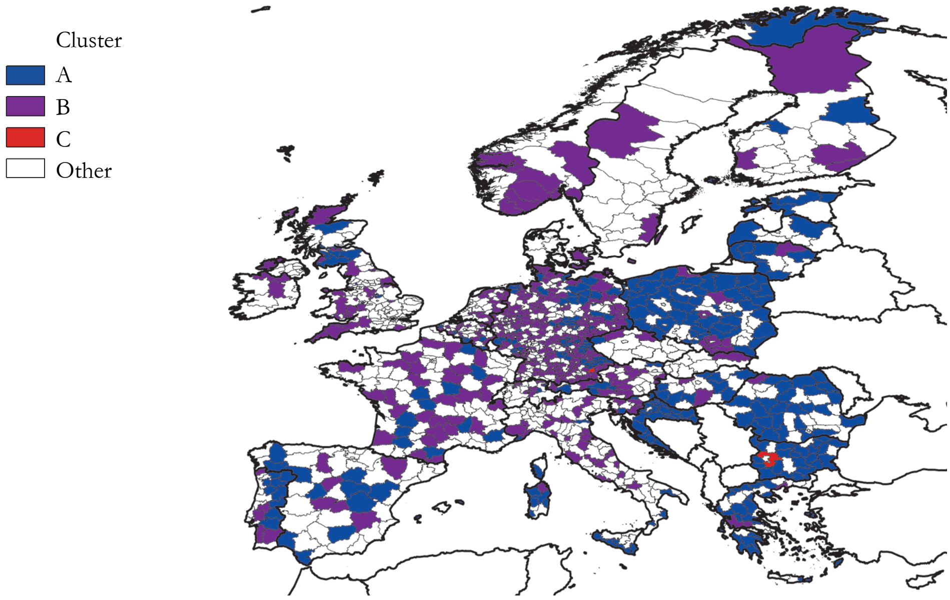


Figure A4

Map of European NUTS3 regions according to their FP cooperation patterns, Group II clusters (D and E)

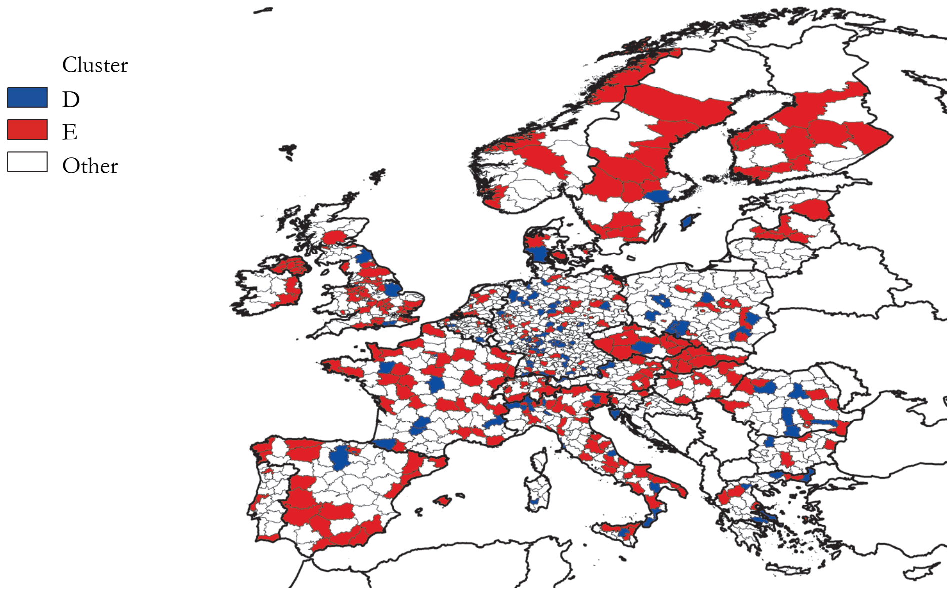
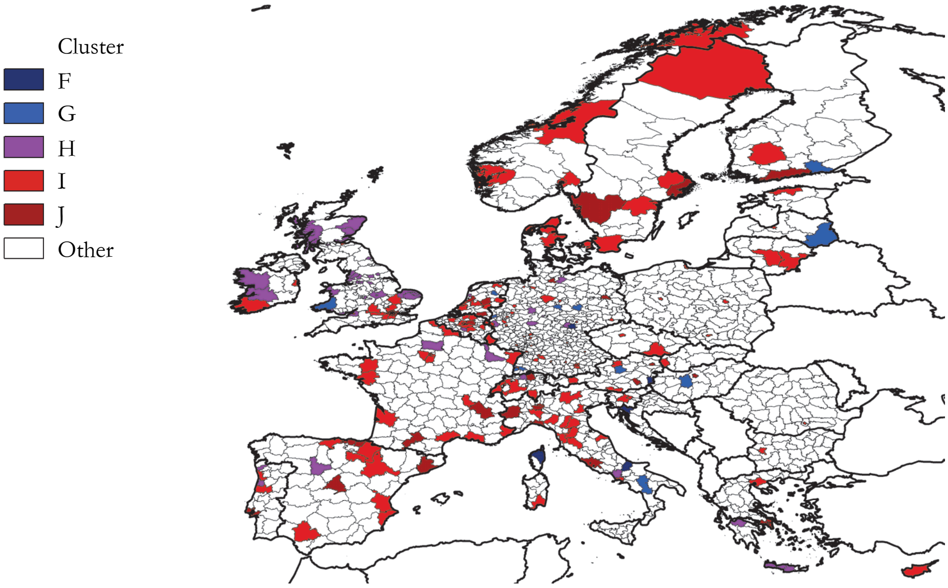


Figure A5

Map of European NUTS3 regions according to their FP cooperation patterns, Group III clusters (F to J)



REFERENCES

ACS, J. Z.–STAM, E.–AUDRETSCH, D. B.–O'CONNOR, A. (2017): The lineages of the entreprenerial ecosystem approach *Small Business Economics* 49 (1): 1–10. https://doi.org/10.1007/s11187-017-9864-8

AKCOMAK, S.–ERDIL, E.–CETINKAYA, U. Y. (2018): *Knowledge convergence in European regions: Towards cohesion?* MERIT Working Papers 027, United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology, Maastricht, The Netherlands.

ALVEDALEN, J.–BOSCHMA, R. (2017): A critical review of entrepreneurial ecosystems research: towards a future research agenda *European Planning Studies* 25 (6): 887–903. https://doi.org/10.1080/09654313.2017.1299694

AUDRETSCH, D. B.–LEHMANN, E. E. (2005): Does the Knowledge Spillover Theory of Entrepreneurship hold for regions? *Research Policy* 34 (8): 1191–1202. http://dx.doi.org/10.1016/j.respol.2005.03.012.

AUTANT-BERNARD, C.–BILLAND, P.–FRACHISSE, D.–MASSARD, N. (2007): Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies *Papers in Regional Science* 86 (3): 495–519. https://doi.org/10.1111/j.1435-5957.2007.00132.x

BALCONI, M.–BRESCHI, S.–LISSONI, F. (2004): Networks of inventors and the role of academia: an exploration of Italian patent data *Research Policy* 33 (1): 127–145. https://doi.org/10.1016/S0048-7333(03)00108-2

BARRAT, A.–BARTHELEMY, M.–PASTOR-SATORRAS, R.–VESPIGNANI A. (2004): The architecture of complex weighted networks *Proc. Natl. Acad. Sci. USA* 101 (11): 3747–3752. https://doi.org/10.1073/pnas.0400087101

BECKER, W.–DIETZ, J. (2004): R&D cooperation and innovation activities of firms – evidence for the German manufacturing industry *Research Policy* 33 (2): 209–223. https://doi.org/10.1016/j.respol.2003.07.003

BLONDEL, V. D.–GUILLAUME, J-L.–LAMBIOTTE, R.–LEFEBVRE, E. (2008): Fast unfolding of communities in large networks *Journal of Statistical Mechanics: Theory and Experiment* P10008 https://doi.org/10.1088/1742-5468/2008/10/P10008

BOSCHMA, R. (2005): Proximity and innovation: a critical assessment *Regional Studies* 39 (1): 61–74. https://doi.org/10.1080/0034340052000320887

BROEKEL, T.–BOSCHMA, R. (2017): The cognitive and geographical structure of knowledge links and how they influence firms' innovation performance *Regional Statistics* 6 (2): 3–26. https://doi.org/10.15196/RS06201

BRUNEEL, J.–D'ESTE, P.–SALTER, A. (2010): Investigating the factors that diminish the barriers to university–industry collaboration *Research Policy* 39 (7): 858–868. http://dx.doi.org/10.1016/j.respol.2010.03.006.

CALOGHIROU, Y.–TSAKANIKAS, A.–VONORTAS, N. S. (2001): University-industry cooperation in the context of the European Framework Programmes *The Journal of Technology Transfer* 26: 153–161. https://doi.org/10.1023/A:1013025615518

CANTNER, U.–GRAF, H. (2006): The network of innovators in Jena: An application of social network analysis *Research Policy* 35 (4): 463–480. https://doi.org/10.1016/j.respol.2006.01.002

CUNNINGHAM, S. W.–WERKER, C. (2012): Proximity and collaboration in European nanotechnology *Papers in Regional Science* 91 (4): 723–742. https://doi.org/10.1111/j.1435-5957.2012.00416.x

CSÁFORDI, Z.–LŐRINCZ, L.–LENGYEL, B.–KISS, K. M. (2018): Productivity spillovers through labor flows: productivity gap, multinational experience and industry relatedness *The Journal of Technology Transfer* 45 (1) 86–121. http://dx.doi.org/10.1007/s10961-018-9670-8.

D'AMBROSIO, A.–MONTRESOR, S.–PARRILLI, M. D.–QUATRARO, F. (2018): Migration, communities on the move and international innovation networks: an empirical analysis of Spanish regions *Regional Studies* 53 (1): 6–16. http://dx.doi.org/10.1080/00343404.2018.1426850.

D'ESTE, P.–IAMMARINO, S. (2010): The spatial profile of university-business research partnerships *Papers in Regional Science* 89 (2): 335–350. http://dx.doi.org/10.1111/j.1435-5957.2010.00292.x.

DE NONI, I.–ORSI, L.–BELUSSI, F. (2018): The role of collaborative networks in supporting the innovation performances of lagging-behind European regions *Research Policy* 47 (1): 1–13. http://dx.doi.org/10.1016/j.respol.2017.09.006.

DOSSO, M.–LEBERT, D. (2019): The centrality of regions in corporate knowledge flows and the implications for Smart Specialisation Strategies *Regional Studies* 54 (10): 1366–1376. http://dx.doi.org/10.1080/00343404.2019.1619923.

FAFCHAMPS, M.–VAN DER LEIJ, M. J.–GOYAL, S. (2010): Matching and network effects *Journal of the European Economic Association* 8 (1): 203–231. https://doi.org/10.1111/j.1542-4774.2010.tb00500.x

FITJAR, R. D.–RODRÍGUEZ-POSE, A. (2019): Where cities fail to triumph: The impact of urban location and local collaboration on innovation in Norway *Journal of Regional Science* 60 (1): 5–32. http://dx.doi.org/10.1111/jors.12461.

FRITSCH, M.–SLAVTCHEV, V. (2010): How does industry specialization affect the efficiency of regional innovation systems? *The Annals of Regional Science* 45 (1): 87–108. https://doi.org/10.1007/s00168-009-0292-9

GEORGE, G.–ZAHRA, S. A.–WOOD, D. R. (2002): The effects of business–university alliances on innovative output and financial performance: a study of publicly traded biotechnology companies *Journal of Business Venturing* 17 (6): 577–609. https://doi.org/10.1016/S0883-9026(01)00069-6

GUAN, J. C.–YAM, R. C. M.–MOK, C. K. (2005): Collaboration between industry and research institutes/universities on industrial innovation in Beijing, China *Technology Analysis & Strategic Management* 17 (3): 339–353. https://doi.org/10.1080/09537320500211466

GUAN, J.–ZHAO, Q. (2013): The impact of university–industry collaboration networks on innovation in nanobiopharmaceuticals *Technological Forecasting and Social Change* 80 (7): 1271–1286. https://doi.org/10.1016/j.techfore.2012.11.013

HANSEN, T. (2014): Juggling with Proximity and Distance: Collaborative Innovation Projects in the D anish Cleantech Industry *Economic Geography* 90 (4): 375–402. https://doi.org/10.1111/ecge.12057

HARDEMAN, S.–FRENKEN, K.–NOMALER, Ö.–TER WAL, A. (2012): A proximity approach to territorial science systems. In: *EUROLIO Conference on 'Geography of Innovation'* pp. 24–26., Saint-Etienne, France.

HARTIGAN, J. A.–WONG, M. A. (1979): Algorithm AS 136: A k-means clustering algorithm *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1): 100–108. https://doi.org/10.2307/2346830

HENDERSON, V. (1997): Medium size cities *Regional Science and Urban Economics* 27 (6): 583–612. https://doi.org/10.1016/S0166-0462(96)02169-2

HOEKMAN, J.–FRENKEN, K.–VAN OORT, F. (2008): *Collaboration networks as carriers of knowledge spillovers: Evidence from EU27 regions* DIME Working Paper in the series on 'Dynamics of Knowledge Accumulation, Competitiveness, Regional Cohesion and Economic Policies, FP7 Project.'

JACOBS, J. (1969): *The economy of cities* Vintage, New York.

JEONG, D.–CHO, K.–PARK, S.–HONG, S. (2016): Effects of knowledge diffusion on international joint research and science convergence: Multiple case studies in the fields of lithium-ion battery, fuel cell and wind power *Technological Forecasting and Social Change* 108: 15–27. https://doi.org/10.1016/j.techfore.2016.03.017

JUHÁSZ, S. (2019): Spinoffs and tie formation in cluster knowledge networks *Small Business Economics* http://dx.doi.org/10.1007/s11187-019-00235-9.

KODINARIYA, T. M.–MAKWANA, P. R. (2013): Review on determining number of cluster in K-means clustering *International Journal* 1 (6): 90–95.

KOTOSZ, B. G.–LUKOVICS, M.–MOLNÁR, G.–ZUTI, B. (2015): How to measure the local economic impact of universities? Methodological overview *Regional Statistics* 5 (2): 3–19. http://dx.doi.org/10.15196/RS05201

LENGYEL, B.–LEYDESDORFF, L. (2015): The effects of FDI on innovation systems in Hungarian regions: Where is the synergy generated? *Regional Statistics* 5 (1): 3–24. http://dx.doi.org/10.15196/RS05101

LUNDVALL, B. Å. (2010) (ed.): *National Systems of Innovation: Toward a Theory of Innovation and Interactive Learning* Anthem Press, London, UK.

MACQUEEN, J. (1967): Some methods for classification and analysis of multivariate observations *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1 (14): 281–297.

MAIETTA, O.W. (2015): Determinants of university–firm R&D collaboration and its impact on innovation: A perspective from a low-tech industry *Research Policy* 44 (7): 1341–1359. http://dx.doi.org/10.1016/j.respol.2015.03.006.

NEWMAN, M. E. J. (2005): Power laws, Pareto distributions and Zipf's law *Contemporary Physics* 46 (5): 323–351. http://dx.doi.org/10.1080/00107510500052444

OECD (2019): *University-industry collaboration: New Evidence and Policy Options* OECD Publishing, Paris.

OWEN-SMITH, J.–RICCABONI, M.–PAMMOLLI, F.–POWELL, W. W. (2002): A comparison of U.S. and European university-industry relations in the life sciences *Management Science* 48 (1): 24–43. https://doi.org/10.1287/mnsc.48.1.24.14275

PÁLÓCZI, G. (2016): Researching commuting to work using the methods of complex network analysis *Regional Statistics* 6 (1): 3–22. https://doi.org/10.15196/RS06101

PÁTHY, Á. (2017): Types of development paths and the hierarchy of the regional centres of Central and Eastern Europe *Regional Statistics* 7 (2): 124–147. https://doi.org/10.15196/RS070202

PONDS, R.–OORT, F. V.–FRENKEN, K. (2009): Innovation, spillovers and university-industry collaboration: an extended knowledge production function approach *Journal of Economic Geography* 10 (2): 231–255. http://dx.doi.org/10.1093/jeg/lbp036.

REICHERT, S. (2019): *The role of universities in regional innovation ecosystems* European University Association, Brussels.

REILLON, V. (2017): *EU framework programmes for research and innovation: Evolution and key data from FP1 to Horizon 2020 in view of FP9* European Parliamentary Research Service. http://dx.doi.org/10.2861/60724

RODRIGUEZ, M. A.–PEPE, A. (2008): On the relationship between the structural and socioacademic communities of a coauthorship network *Journal of Informetrics* 2 (3): 195–201. https://doi.org/10.1016/j.joi.2008.04.002

SANTOALHA, A. (2018): Technological diversification and smart specialisation: the role of cooperation *Regional Studies* 53 (9): 1269–1283. http://dx.doi.org/10.1080/00343404.2018.1530753.

SCHAEFFER, P. R.–FISCHER, B.–QUEIROZ, S. (2018): Beyond education: The role of research universities in innovation ecosystems *Foresight and STI Governance* 12 (2): 50–61. http://dx.doi.org/10.17323/2500-2597.2018.2.50.61.

SCHLUGA, T. R.–BARBER, M. J. (2008): R&D collaboration networks in the European Framework Programmes: data processing, network construction and selected results *International Journal of Foresight and Innovation Policy* 4 (3/4): 321. http://dx.doi.org/10.1504/ijfip.2008.017583.

SEBESTYÉN, T.–VARGA, A. (2013): Research productivity and the quality of interregional knowledge networks *Annals of Regional Science* 51: 155–189. https://doi.org/10.1007/s00168-012-0545-x

SOUSA, C.–SALAVISA, I. (2015): *International R&D networks in renewable technologies – Evidence from the Portuguese participation in European Framework Programmes* Centro de Estudos sobre a Mudança Socioeconómica e o Território, Lisboa. http://dx.doi.org/10.15847/dinamiacet-iul.wp.2015.11.

VALLANCE, P.–BLAŽEK, J.–EDWARDS, J.–KVĚTOŇ, V. (2017): Smart specialisation in regions with less-developed research and innovation systems: A changing role for universities? *Environment and Planning C: Politics and Space* 36 (2): 219–238. http://dx.doi.org/10.1177/2399654417705137.

VARGA, A.–PONTIKAKIS, D.–CHORAFAKIS, G. (2014): Metropolitan Edison and cosmopolitan Pasteur? Agglomeration and interregional research network effects on European R&D productivity *Journal of Economic Geography* 14: 229–263. http://dx.doi.org/10.1093/jeg/lbs041

VARGA, A.–SEBESTYÉN, T. (2017): Does EU Framework Program participation affect regional innovation? The Differentiating Role of Economic Development *International Regional Science Review* 40 (4): 405–439. https://doi.org/10.1177%2F0160017616642821