Regional
Innovation and Entrepeneurship
Research Center

WORKING PAPER SERIES

2021-003

Spatial and Institutional Dimensions of Research collaboration: a Profiling of European regions

Tamás Sebestyén
MTA-PTE Innovation and Economic Growth Research Group
University of Pécs, Faculty of Business and Economic

Erik Braun
MTA-PTE Innovation and Economic Growth Research Group
University of Pécs, Faculty of Business and Economic

Zita Iloskics
MTA-PTE Innovation and Economic Growth Research Group
University of Pécs, Faculty of Business and Economic

Attila Varga
MTA-PTE Innovation and Economic Growth Research Group
University of Pécs, Faculty of Business and Economic

Regional Innovation and Entrepreneurship Research Center
Faculty of Business and Economics
University of Pécs

H-7622, Pécs Rákóczi str. 80.

Phone: +36-72-501-599/23121

www. rierc.ktk.pte.hu

Spatial and Institutional Dimensions of Research collaboration: a Profiling of European regions

Tamás Sebestyén[a,b,c], Erik Braun[a], Zita Iloskics[a], Attila Varga[a,b,c]

Abstract

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

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

1 Introduction

Recognizing that innovation is an inherently collaborative process, recent research has shed light on the prominent role of cooperation in innovative activities (e.g. Lundvall, 2010). Through innovation, different networks of cooperation can thus 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 specifically call attention to the importance of interregional cooperation (e.g. Hoekman et al. 2008, Varga et al. 2014, Sebestyén and Varga 2013). Moreover, these more distant ties of knowledge flows can significantly improve innovation performance in those lagging regions where the local supply of resources used in innovation is scarce, because the networks provide access to similar resources accumulated elsewhere (Varga and Sebestyén, 2017).

On the other hand, the literature on regional innovation systems emphasize that the collaborative nature of innovation is also leveraged by the cooperation between different types of actors like firms, universities, research centers, government agencies, financial institutions and others, providing support to the innovative process (Jacobs 1969, Henderson 1997, Fritsch and Slavtchev 2010, Becker and Dietz 2004, Csáfordi et al. 2018, D'Ambrosio et al. 2018). Among these specific ties, much attention has been given to the relationship between industry actors and universities showing that these links

[a] University of Pécs, Faculty of Business and Economics

[b] Regional Innovation and Entrepreneurship Research Center

[c] MTA-PTE Innovation and Economic Growth Research Group

can be conducive to innovation and development (OECD 2019, Reichert 2019). A somewhat related field, that of entrepreneurial ecosystems also calls attention to the dense interaction among a diverse set of actors in promoting entrepreneurial activity and innovation (Acs et al. 2017, Alvedalen and 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. Having the opportunity to distinguish between different types of institutions, we pay attention to the patterns of interaction between these different institution types, especially universities and industry actors. Although there are several studies focusing on university-industry collaboration, the evidence reported about the patterns and the role of these connections are largely based on regional-level case studies (Cantner and Graf 2006, Guan et al. 2005, Reichert 2019) or restricted to specific technological fields (Guan and Zhao 2013, Owen-Smith et al. 2002, George et al. 2002). Also, large-scale studies covering several fields and regions/countries rely on patent data (Balconi et al. 2004, Owen-Smith et al. 2002, OECD 2019). Our paper fills a 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.

Several studies draw the attention on 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 in those areas where universities are present, innovation activity is higher. In a German sample of startups, 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 startups. Similarily, Maietta (2015) shows that the innovative activity of a sample of Italian food and drink companies was positively affected by cooperating with universities. Putting the analysis on the regional level, Ponds et al. (2009) also conclude that interregional collaboration networks between universities and companies (measured by co-publications) are important channels of knowledge spillovers in the case of Dutch regions.

The simple positive effect of collaboration is shaded by other studies. D'Este and Immarino (2010) e.g. find that both geographical proximity and research quality of the university affect the frequency of joint research collaboration between the two types of institutions. On the other hand, Bruneel et al. (2010) draw attention on the obstacles in university-industry collaboration arguing that previous collaboration, a wider set of collaborators and trust can help overcoming these obstacles. Analyzing survey data focusing in RIS3 strategies, Vallance et al. (2017) find that in spite of the connections between universities and companies, less innovative regions lag behind 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 reach out for data on EU-funded Framework Program collaboration. Caloghirou et al. (2001) analyze university-industry collaboration in Framework Programs. They provide a large-scale analysis between 1983 and 1996, showing that universities are included in more than 50% of cooperative projects and their role increases with time. They also point out that universities of peripheral countries are also important actors of this collaboration network. This result is reinforced by Sousa and Salavisa (2015) in a more recent paper on Portugese collaborative projects showing that universities play a key role in these networks and they become more and more central.

Although not explicitly addressing university-industry collaboration, other studies provide an overview on Framework Program collaboration. Reillon (2017) gives a comprehensive overview of the different waves of the Framework Programs, while Schluga and Barber (2008) analyze the properties of the collaboration network across different FPs, finding that in spite of the changing programs, the structure of the network does not change too much and a stable core can be identified which mainly consist 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 emphasize that the collaborative ties between project participants of the Framework Programs constitute a network and analyzing the different properties of this network can provide useful insights. However, there are several studies which used this approach to reveal the patterns of innovation. Using patent collaboration data, De Noni et al. (2018) emphasize that the frequency of intraregional collaboration does not affect patent applications in less innovative regions. However, extraregional collaboration has a positive effect especially with high-performance regions. Also using patent collaboration data, Santoalha (2018) highlights that successful innovation systems are both locally and globally integrated which means a balanced set of collaborative ties in both directions. Still with patent data Dosso and Lebert (2019) show that most central regions are the strong innovators. On a similar line, but based on a survey done in a Hungarian region Juhász (2019) emphasize 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 take into account the structure of the network among participating institutions. Ponds et al. (2009) work with a spatial econometric setup where collaboration networks are taken into account through the spatial weight matrix. More precisely, the weighted average of R&D of partner regions are taken into account as a determinant of local innovativeness. However, this study relies on publication data as a measure of research collaboration and the nodes of the network are regions. Although Akcomnak et al. (2018) use data on Framework Program cooperation, they also set up the collaboration network at the regional level. Then they use different centrality measures to analyze convergence patterns of countries. On the other hand, Schluga and Barber (2008) analyze the participant-level network of Framework Program collaboration, focusing on the evolution of some macroscopic properties (degree distribution, small world properties) of these networks. However, they do not address university-industry collaboration explicitly. Sousa and Salavisa (2015) analyze 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.

Our study thus extends this line of study by providing a large scale analysis of Framework Program collaboration patterns from a network perspective. We can rely on a participant-level network in our analysis which covers all EU countries through the waves of FP5, FP6 and F7. By differentiating between different types of institutions we are also able to address cooperation between universities (higher education institutions and research institutions in general) and industry actors. We use this data to provide a profiling of European NUTS3 regions through a clustering analysis where regions are grouped along their collaboration patterns. We contribute to the literature by mapping the institutional and spatial dimensions of research collaboration networks, while relaying 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 a primary focus on the collaboration patterns between research institutions (including universities) and industry actors.

The paper is structured as follows. In Section 2 we discuss the construction of the database, with special emphasis on the background of compiling FP collaboration data, then the methods behind our clustering exercise is given. Section 3 provides the results of the clustering analysis, discusses the results and elaborates on the changing patterns of regional profiles. Finally, Section 4 concludes the paper.

2 Data and method

In this section we briefly discuss the data sources of the analysis, provide background for the different indicators that we employ in the clustering exercise and also describe the methods used to do the profiling of the regions in our sample.

2.1 Data sources and network construction

2.1.1 Network data

The starting point of our analysis is the information available on EU-funded Framework Programmes. This information can be retrieved from the Cordis database. For the analysis in this paper, we use information on all projects funded in the three waves of the Framework Programmes: FP5, FP6 and FP7. The basic unit of this data is a project-participant pair. This means a particular institution (as participant, e.g. university, company) being involved in a funded project. First, we use information on the projects: the contract numbers of the specific projects are used as unique identifiers, and the duration (starting and ending years) of the projects allow us to have a longitudinal approach on the collaboration patterns. Second, information on the participants is used: their location, as the NUTS3 level region they belong to and the type of the institution (e.g. higher education institution, industry actor).

This data had to go through two waves of data cleaning. First, we had to clean regional classification. Although the Cordis dataset provides NUTS3-level categorization of the participants, this is not complete and come with errors in several cases. We did a complete re-classification in this respect based on the information on postal codes, addresses and cities provided in Cordis. If this information was not enough, manual checks were done to assign a clean regional code at the NUTS3 level to all institutions. Second, as the participant identifiers provided by Cordis are similarly problematic, especially to be used across different FP programmes, we did a complete re-identification of institutions. Using information on the name, location (region) and address of the participants we run a string-matching algorithm to reveal the similarity of every institution-pair. The same procedure was done manually as well on a subsample of institutions. The latter provided reference-cases where we were sure about which institutions are the same and which are different. This reference subsample was then confronted with the algorithmic results, in order 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 to be identical. Pairs falling into the ambiguity range were manually checked again to finally arrive at a clean identification of institutions.

In the cleaned dataset we have information about every funded project, the duration of the project, the participants of the project, their location at the NUTS3 level and their type being higher education institution, research institution, industry actor or other. In this analysis we consider only the first three types with merging research institutions and higher education institutions into one category. For simplicity, we will refer to the latter group as research institutions in general.

Our starting point for data manipulation is the project matrix $\mathbf{P}_t$ the rows of this matrix correspond to institutions whereas the columns represent projects. A given cell of the matrix is one if institution i was participating in project k in year t. From this project matrix, simple matrix manipulation provides the adjacency matrices $\mathbf{A}_t$ for all years in our sample: $\mathbf{A}_t = \mathbf{P}_t\mathbf{P}_t^{\mathrm{T}}$, where $\mathbf{P}_t^{\mathrm{T}}$ is the transpose of $\mathbf{P}_t$. The resulting $\mathbf{A}_t$ adjacency matrix provides the number of ongoing joint projects between any pair of institutions in year t. Being our starting point for further calculations, these adjacency matrices give a snapshot of collaboration patterns between institutions with a weighted perspective: we account for the number of joint projects in all years, reflecting the intensity of collaboration.

The above mentioned adjacency matrix $\mathbf{A}_t$ contains information between all pairs of institutions, regardless of their location (region) and type (research institution or industry actor). In order to account for these features, we use two categorization vectors. $\mathbf{d}^T$ refers to the type of institutions: it has one entry (row) for all institutions and contains 1 of the given institution is a research institution and 2 if it is a company/industry actor. Similarly, $\mathbf{d}^R$ refers to the location of institutions and one entry (row) contains the index of the region the institution belongs to.

In order to ease further exposition, we reshape the adjacency matrix $\mathbf{A}_t$ into an array $\mathbf{W}_t$ which structures connections between institutions along their location and type as well. The general element of it is defined as follows:

$$w_{rfi,qgj,t} = a_{l_1 l_2} | d_{l_1}^R = r, d_{l_2}^R = q, d_{l_1}^T = f, d_{l_2}^T = g$$

In other words, $w_{rfi,qgj,t}$ describes the number of joint collaboration projects in year t 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, $r, q = 1,2, \dots, R$ refer to region indices, while $i, j = 1,2, \dots, I_{f,r}$ 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 $\mathbf{W}_t$, representing the weights (intensities) of collaboration between any two institutions, we define the binary version of this structure, $\mathbf{B}_t$, representing the existence of collaboration between institutions in period t, rather than their intensity:

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

where the indices have the same meaning as in the case of the weighted connections.

The notation above, using multidimensional structures instead of one large adjacency matrix, allows for a simple exhibition of calculations behind the network properties we analyze. We construct these network property indicators along two dimensions. Along the first dimension, we distinguish the relationship of institutions as being local or global. As we mentioned in the introduction, 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 a companies, (ii) both institutions are a research institutions, (iii) one institution is a company, the other is a research institution. If we take into account all possible cases by the two dimensions, we can calculate seven different versions of all network property for every region.[1] These versions are summarized in Table 1 below.

[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 regions is the same as between research institutions and companies, but there is a difference between cooperation intensity of local companies with extra-regional research institutions and local research institutions with extra-regional companies.

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

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

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

In the formulae above, $S_{r,t}^{in,fg}$ refer to the strength of connections between institutions of type f and type g, both belonging to region r. Thus, it refers to intraregional connection strength. The formulae $S_{r,t}^{out,fg}$ 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: how many connections do we observe 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 $\mathbf{B}_t$. Density is then calculated in an analogous way to strength:[3]

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

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

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

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

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

where the notation resembles that used for density and strength.

---

[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.

[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.

Table 1. Summary of network indicators by type and location of participants

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

As summarized in Table 1, we have three indicators for all types of relationships describing three different aspects of the strength of network connections. Although these three indicators focus on different aspects, higher values always point to more intensive relationships. As a result, we use the three indices (strength, density, average strength) together, thus we can control for the number of institutions in different regions. In order to compress information, we will present the results of a composite indicator of the three different values: density, strength and average strength. This indicator is the simple average of standardized (mean 0 and standard deviation 1) values. However, the clustering analyses are run on the three indicators separately.

2.1.2 Data on the level of development

In the previous section, we discussed those measures which we use to capture the collaboration patterns among institutions. In order to have a more comprehensive view on collaboration patterns, these network indicators are augmented by information on the development level and the innovation capacity of regions. 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 network measure which are also calculated for this regional disaggregation.

Data on GDP per capita is retrieved from the OECD database between 2000 and 2013. This database contains a moderate number of missing data which were replaced from 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 base year 2015) to make it consistent with the rest of the data. In the case of Cyprus, we needed to convert the base year of the deflator from 2010 to 2015 prices. For some Polish regions we used NUTS 2 data. In Switzerland, data is only available at the national level until 2008, so we used the average national data for the regions as well between 2000 and 2007.

In the case of innovation activity, data is retrieved from the Eurostat database in the form of patents per million inhabitants – this data is available for the majority of the European NUTS3 regions. In those cases where this data is missing, but patent count is available, we calculated the patent per million inhabitants using the earliest population data which is available. In other cases of missing data, when the patent counts indicated low patenting activity, per capita patents were replaced by zeros. 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 there was no available population data while the patent count was significantly different from zero.

2.1.3 The final dataset

After the data preparation process discussed above, we end up with an extensive database of collaboration links between different institutions. The main characteristics of this data are summarized in Table 2. Although the CORDIS data classifies institutions into industry, higher education, research institutions, government institutions and other, in this analysis we use only industry and higher education plus research institutions together and omit government institutions and those participants which are dominantly classified as other. It is clear that 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	56597	56473
Number of projects	51187	
Number of industry actors	27509	27474
Number of higher education and research inst.	10561	10527
Number of regions	1419	1378

On the basis of this collaboration data we constructed the network indicators summarized in Table 1, and augmented them with GPD and patent data to have a more comprehensive dataset at the regional level. The descriptive statistics of these indicators can be found in Table A1 in the Appendix.

2.2 Profiling through clustering

The goal of this study is to provide a mapping of European regions according to their collaboration patterns and development level. We accomplish this goal by running a standard clustering analysis on the basis of the indicators introduced so far. The result of this exercise will provide 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 and Makwana 2013). The k-means clustering is suitable in those cases, where outliers are a problem, however, because of the random initial state of this method, it provides different results for every calculation. Hierarchical clustering gives consistent results, but it is very sensitive to outliers. Considering the descriptive statistics of the indicators and the presence of outliers in our sample, we apply the k-means clustering technique. This algorithm classifies regions into groups so that the Euclidean distance between normalized indicator values of a given region is the closest to the group-center among all groups (Hartigan-Wong, 1979). 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 centers; (iii) it adds the observations to the group the center of which is the closest to the observation; (iv) it recalculates the centers 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 and use a fitness measurement to select the best grouping, by minimizing the total within-cluster sum of squared distances (TSS) between observations and cluster centers (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 the Elbow-method, which is the simplest and most practical solution (Kodinariya and 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.

3 Results and discussion

This section presents and discusses the clustering exercise done with the method described in 2.2 and on the indicators presented in 2.1. In the clustering exercise we used all network and development indicators separately, but for the sake of conciseness the results are presented with the network indicators grouped into the 7 composite indicators along the different connection types shown in Table 1.

Table 3 shows the results of the clustering exercise where all network indicators are included together with GDP and patent per capita. The entries in the table reflect the extent to which the given indicator (column) in the given cluster (row) is above/below the average. These values represent a standard normal distribution where the zero mean reflects 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 with respect to 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 vales show above average performance. Shading shows the extent to which values are below average (red) or above average (blue).

Table 3. Characteristics of clusters

	Intra-regional connections			Inter-regional connections				Development level		Number of regions
	I - I	*R - R*	*I - R*	*I - I*	*R - R*	*I (loc) - R*	*R (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. 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 characterize all clusters in what follows. From the 10 clusters we can form 3 main groups which give the organizing structure of the following discussion.

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

- **Cluster A – Non-cooperative, less developed**. In this cluster regions do not show intensive cooperation neither within nor outside the region. This is a relatively numerous 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, so in this group institutions have a bit more dense interaction 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 regions which are relatively developed while their embeddedness in collaboration networks is sparse. This is also a numerous group with 375 regions.
- **Cluster C – Non-cooperative, less developed with strong local industry**. This group is very similar to cluster A with all but one indicators showing significantly below-average values. So this group also contain less developed regions with low levels of cooperation, 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 here within the whole sample. However, there are only two regions belonging 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 actors of networks, but cooperation is typically extra-regional. While intra-regional cooperation is below-average in all types, local companies have very strong cooperation with both companies and research institutions outside the region. This is a relatively numerous cluster with 113 regions, being relatively developed, but not among the most developed ones. On the other hand, research institutions have very sparse connections in this type of regions, both locally and externally.
- **Cluster E – Externally focused, research-based.** This cluster is similar to cluster D in the sense that 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 this cluster E local research institutions provide the basis for external links. However, local companies still show above average external cooperation, but the strength of this cooperation 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 level of development, compared to cluster D both GDP per capita and innovation is 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, but this 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, when it comes to external links, local research institutions show above average connections while local companies are weakly connected. This group, although not sizeable, show slightly above average GDP per capita levels but very low patenting activity.
- **Cluster G – Research-based, less developed.** This cluster is dominated by research institutions. While we see strong local connections between research institutions, other local connections

are weak. External connections are also dominated by research institutions. While not that numerous (13 regions), this cluster consists of less developed regions.

- **Cluster H – Research-based, developed.** This cluster is similar to cluster G, but it consists of more developed regions: both GDP and patent per capita are significantly above average. This is the first cluster where almost all network indicators are above average, except local intra-industry cooperation. However, cooperation networks are still dominated by local research institutions 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 numerous 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 very similar to cluster I with even stronger networks. Especially local collaboration is very strong, while these regions also show a somewhat higher GDP per capita, but 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 quite 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 developed regions (above average GDP and patent per capita), there is a significant group (27%) of relatively developed regions with below average collaboration intensities in all types.

Along with these homogenous regions, there is a significant, 32% of regions which show selective collaboration patterns. These are typically moderately developed regions and they are frequently characterized by a university dominance, especially in the case of those 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 which keep extra-regional connections, while the relatively less developed cluster E relies mainly on local universities and research institutions which are embedded in extra-regional cooperation.

As mentioned earlier, one particular interest in this study is the collaboration between industry acotrs and research institutions. It seems from the clustering above that local interaction between research institutions and companies typically goes hand in hand with extra-regional cooperation: those regions which have dense internal interactions between the two types of actors, also show strong cooperation between local research institutions and outside companies. Cooperation between local companies ad 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. It is generally true that 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. On the other hand, external links are mainly dominated by research institutions.

With respect to local industry-research collaboration, we see four clusters where this collaboration type is significantly above average: clusters I and J which are 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 spare. Cluster F seem to be the opposite, where local cooperation between research institutions is sparse but 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 while the locally industry-based cluster F contain regions with slightly above-average GDP per capita and below average patenting activity. However, the full picture must contain that cluster H contains 5% of the regions, there is only 6 regions (0.6%) in cluster F.

Figure 1. Map of European NUTS2 regions according to their FP cooperation patterns, group level

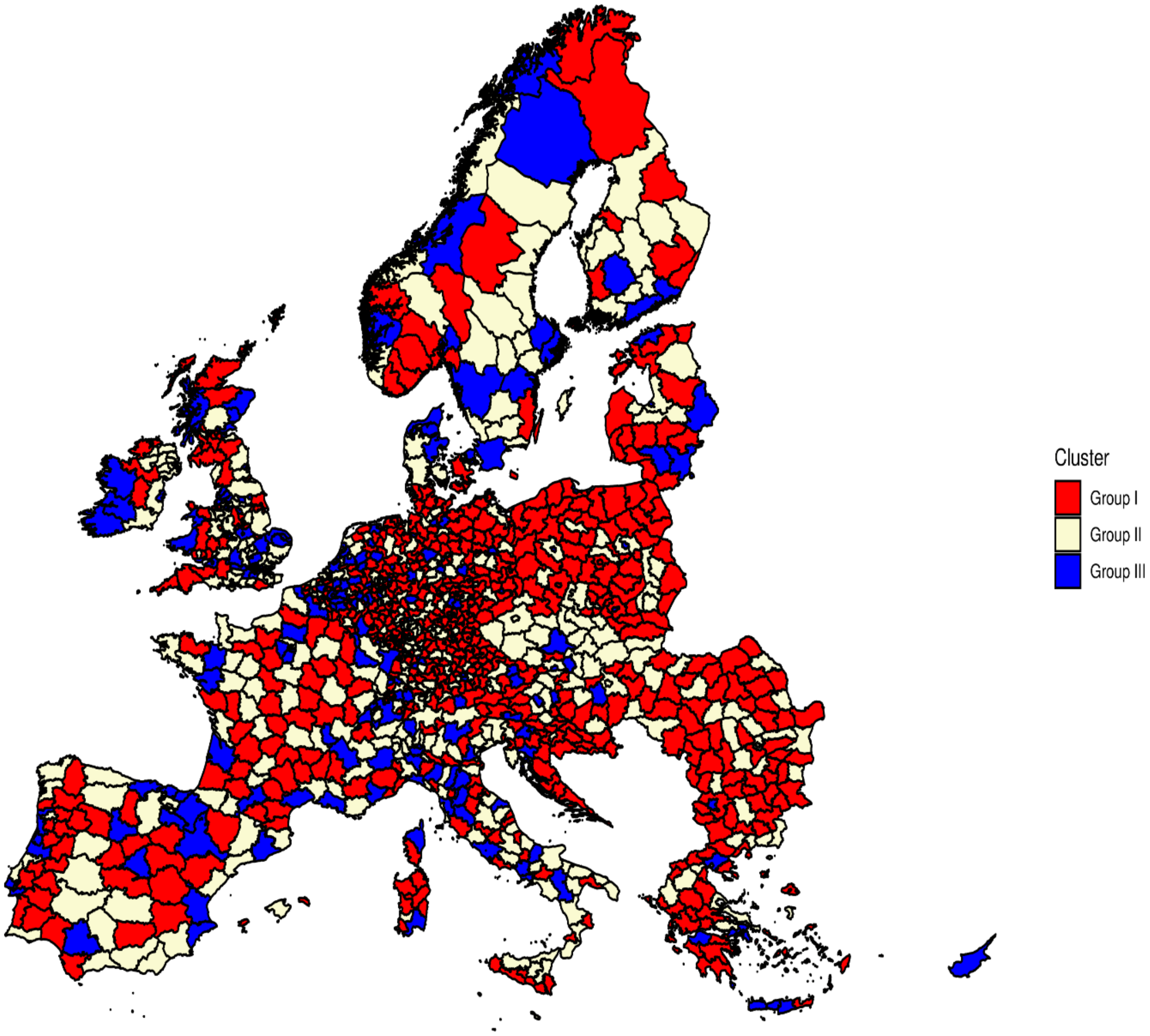

Figure 1 augments the previous analysis by showing the map of European NUTS3 regions with the aggregated results of the clustering analysis: regions are colored according to the three large groups they belong to. Somewhat in line with the coloring in Table 4, the red shades refer to the less cooperative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

The picture shows that while red regions are dominant in Eastern Europe, there are also a considerable amount of this type of regions in the Western part of the continent. 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 which show above-average level of development, while they are still non-cooperative. It is visible in the figure that the latter regions are also scattered around Western Europe, although these are certainly not the most central/developed parts. On the other hand, it is very rare to find blue, i.e. above-average cooperative regions in the Eastern parts. These are mainly capital regions as Warsaw, Bratislava, Sofia and Ljubljana, and the areas around some relatively industrialized cities such as Krakow in Poland, Brno in the Czech Republic and Vilnius and Kaunas in Lithuania. However, several regions in Eastern Europe belong to Group II, with moderate development level and primarily externally oriented collaboration patterns.

A more detailed picture can be found in Figure A1 in the Appendix, which shows the 10 different clusters on the same map. This picture reinforces that cluster A contains regions from Eastern Europe while cluster B contains regions from Western Europe. Within the moderately developed regions in Group II, there are two sizable clusters D and E, with an external collaboration focus: local cooperation is week in these regions, but external cooperation is strong. However, while in cluster D (which are a bit more developed on average) local companies dominate these networks, local universities and research institutions are dominant in cluster E. The observations show that the university-based model seem to characterize the more developed part of CEE countries (especially the Czech Republic, Slovakia and Hungary), a significant part of Scandinavian countries and also less central regions in Western Europe. Cluster D with a more industrial-focus in external cooperation can be found also all around Europe, but these regions are showing up more frequently in Germany, Poland, Romania and Italy.

4 Conclusion

In this paper we used 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 in the former respect we focused on industry actors sand research institutions (universities) as two main types of institutions and the collaboration among them, in the latter respect we were able to go down to a relatively detailed, NUTS3 regional level. This institutional detail provides an opportunity also to focus on the collaboration patterns between industry actors (companies) and research-focused actors (universities, research institutions) which has been the subject of several studies before.

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

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

First, it became clear that while the level of development roughly moves together with cooperation intensity, there is a numerous group of relatively developed (typically Western European) regions which are weakly cooperative. Second, there is a quite heterogeneous group of regions between cooperative and non-cooperative ones which are typically in the middle of the development scale and their cooperative patterns are selective either institutionally or spatially. In the latter group, we found that that most of the regions are externally focused, with strong external collaboration intensities and weak local ones. This external focus is dominated by research institutions in the majority of the cases, but there is a visible amount of regions which base their external collaboration on local industry actors.

With respect to the specific collaboration pattern between industry and research institutions (universities in particular) our results are threefold. First, in line with the correlation mentioned above, these specific collaboration links across different types of actors seem to systematically show up 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. Local, within-region collaboration between industry actors and research institutions are found to be very rare outside strongly cooperative and developed regions. Extra-regional collaboration between the two different institution types is more frequent, but in these cases typically local universities cooperate with companies across the borders. Still, some regions show an industry-based cooperation network.

There are two lines along which this research can be readily extended. First, a longitudinal analysis of these cooperation patterns is viable showing the route of different regions between different groups/clusters of collaboration patterns. Second, using econometric techniques this data can be used to infer on the role of the institutional and spatial dimensions of collaboration patterns in shaping regional innovativeness.

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.

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.

Alvedalen, J., Boschma, R. (2017): A critical review of entrepreneurial ecosystems research: towards a future research agenda. European Planning Studies, 25(6): 887–903.

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

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, 2004, 127-145.

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.

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. Available at: 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), pp. 463-480.

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. Available at: 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. Available at: 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. Available at: 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, 1–11. Available at: http://dx.doi.org/10.1080/00343404.2019.1619923.

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. Ann Reg Sci (2010) 45: 87.

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.

Guan, J., Zhao, Q. (2013): The impact of university–industry collaboration networks on innovation in nanobiopharmaceuticals. Technological Forecasting and Social Change, 80(7)

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.

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.

Henderson, V. (1997): Medium size cities. Reg Sci Urban Econ 27:583–612.

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.

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.

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

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).

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. Available at: http://dx.doi.org/10.1016/j.respol.2015.03.006.

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), pp. 24-43.

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. Available at: http://dx.doi.org/10.1093/jeg/lbp036.

Reichert, S. (2019): The Role of Universities in Regional Innovation Ecosystems. European University Association.

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. September. https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_IDA(2017)608697

Rücker Schaeffer P., Fischer B., Queiroz S. (2018): Beyond Education: The Role of Research Universities in Innovation Ecosystems. Foresight and STI Governance. 12. évf. 2. sz. 50–61. o. http://dx.doi.org/10.17323/2500-2597.2018.2.50.61.

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.

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.

Sousa, C., Salavisa, I. (2015): International R&D networks in renewable technologies – Evidence from the Portuguese participation in European Framework Programmes. 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.

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) pp. 1-35.

Appendix

Table A1. Descriptive statistics of network and development indicators (full sample, 1999-2013)

	Number of regions	Mean	Standard deviation	Max	Min
$D_{r,t}^{in,11}$	1378	0.0024	0.0145	0.3333	0.0000
$D_{r,t}^{out,11}$	1378	0.0001	0.0001	0.0007	0.0000
$D_{r,t}^{in,22}$	1378	0.0051	0.0196	0.4000	0.0000
$D_{r,t}^{out,22}$	1378	0.0011	0.0018	0.0174	0.0000
$D_{r,t}^{in,12}$	1378	0.0028	0.0109	0.2000	0.0000
$D_{r,t}^{out,12}$	1378	0.0002	0.0002	0.0012	0.0000
$D_{r,t}^{out,21}$	1378	0.0001	0.0002	0.0021	0.0000
$S_{r,t}^{in,11}$	1378	0.9686	4.9116	81.5333	0.0000
$S_{r,t}^{out,11}$	1378	61.6364	179.2731	2702.7330	0.0000
$S_{r,t}^{in,22}$	1378	1.9103	15.9753	521.4000	0.0000
$S_{r,t}^{out,22}$	1378	241.1269	818.6080	17363.7300	0.0000
$S_{r,t}^{in,12}$	1378	1.9904	11.1553	276.2667	0.0000
$S_{r,t}^{out,12}$	1378	66.5479	200.7637	3586.4000	0.0000
$S_{r,t}^{out,21}$	1378	66.5363	226.3616	4143.6670	0.0000
$\widehat{S}_{r,t}^{in,11}$	1378	0.0303	0.0664	0.6216	0.0000
$\widehat{S}_{r,t}^{out,11}$	1378	2.2002	2.0402	22.4952	0.0000
$\widehat{S}_{r,t}^{in,22}$	1378	0.1067	0.3484	4.5737	0.0000
$\widehat{S}_{r,t}^{out,22}$	1378	15.1038	27.3401	283.9667	0.0000
$\widehat{S}_{r,t}^{in,12}$	1378	0.0271	0.0577	0.7120	0.0000
$\widehat{S}_{r,t}^{out,11}$	1378	2.1769	2.0567	23.7429	0.0000
$\widehat{S}_{r,t}^{out,21}$	1378	4.1844	7.4991	65.8667	0.0000
GDP per capita	1378	34770.8800	21207.9400	468013.5000	7315.3570
No. of patent	1378	126.2648	174.5742	1964.7420	0.0000

Figure A1. Map of European NUTS3 regions according to their FP cooperation patterns, cluster level