European Commission

SMART SPECIALISATION: BEYOND PATENTS

Project 2018CE160AT089/090

Final report

December 2019

EUROPEAN COMMISSION

Directorate-General for Regional and Urban Policy

Directorate Smart and Sustainable Growth and Programme Implementation IV

Unit Smart and Sustainable Growth

Contact: Peter Berkowitz

E-mail: REGIO-G1-HEAD-OF-UNIT@ec.europa.eu

European Commission
B-1049 Brussels

Manuscript completed in December 2019

Luxembourg: Publications Office of the European Union, 2019.

EUROPEAN COMMISSION

SMART SPECIALISATION: BEYOND PATENTS

Project 2018CE160AT089/090

Final report

Prepared by

Pierre-Alexandre Balland and Ron Boschma

Directorate-General for Regional and Urban Policy

2019 EN

Acknowledgements

The authors would like to acknowledge the help and support received from Richard Deiss, Jitka Vocaskova, Marek Przeor and Peter Berkowitz in the Regional and Urban Policy Directorate of the European Commission.

The authors take responsibility for the views expressed in the report.

Contents

Smart Specialisation: beyond patents

ABSTRACT

Local capabilities are regarded a key pillar of Smart Specialization policy in the EU. Regions should build on existing capabilities to develop new activities. This makes it crucial to understand how to capture and measure regional capabilities. Studies often rely on one type of capabilities, like technological capabilities based on patent data, and use that information to identify diversification potentials of regions. This implies other types of capabilities in regions are not taken up, resulting in a possible underestimation of the diversification potentials of regions. This report analyzes different capabilities to assess the diversification potentials of 292 European NUTS2 regions by using data on occupations and sectors from the EU Labour Force Survey database. The report finds that regions in Europe differ in terms of capabilities. For instance, a region may score high with respect to diversification potentials in new occupations, but low on new sectors, or new technologies. This report shows how crucial it is to account for different types of diversification opportunities of regions in S3 strategies, instead of relying on one type only. Finally, it shows the diversification potentials of European regions in 7 key technologies, like batteries and hydrogen.

1. INTRODUCTION

From the start, the role of regional capabilities has been regarded as a key pillar of smart specialization (S3) policy (Foray 2015; Balland et al. 2019). Regions should build on existing capabilities to develop new activities, because regional capabilities condition which new activities are feasible to develop. There is strong support in the empirical literature that regions are indeed engaged primarily in related diversification (Boschma 2017).

Scholars have proposed a S3 framework to map diversification opportunities of regions and to formulate an evidence-based S3 strategy (Balland et al. 2019). This framework is based on the concepts of relatedness and complexity. Studies have used patent data to map diversification opportunities of regions, but there is wide consensus that patent data might not fully capture the whole range of knowledge bases and diversification opportunities available to regions. Non-technological knowledge bases (of high and low complexity) also provide diversification opportunities for regions (McCann and Ortega-Argilés, 2015), especially in the more peripheral parts of the EU where one can find less technologically-advanced regions.

Therefore, it is crucial to underline that S3 policy is about diversifying from regional capabilities in general, not from capabilities in high-tech only. A S3 policy framework should therefore assess diversification options for regions that include other forms of capabilities that go beyond patents, such as economic sectors and occupations. In this project, we explore the possibility to add other economic activities into this S3 framework by using data on occupations and sectors in European regions, derived from the EU Labour Force Survey Database. We assess the potentials in each of the 292 European (NUTS 2) regions to develop new occupations and to develop new sectors, following the S3 policy framework of Balland et al. (2019). The rationale behind this policy framework is that a region should aim to target and develop new sectors and new occupations that would involve less risks (because of their higher relatedness with existing sectors and occupations in the region) and generate potentially higher economic benefits (because of their higher complexity).

The report is structured as follows. Section 2 gives a short literature review. Section 3 presents a S3 framework based on the concepts of relatedness and complexity. Section 4 explains how we calculate relatedness between all sectors, and between all occupations, and how we derive a measure of complexity for all occupations and sectors. Section 5 calculates for each region in Europe its diversification potentials with respect to new occupations and new sectors, based on the degree of relatedness between potential new occupations/sectors and existing occupations/sectors in the region, following the S3 policy framework proposed by Balland et al. (2019). Section 6 shows the diversification potentials of all European regions in 7 key technologies, like batteries and hydrogen. Section 7 concludes.

2. REGIONAL CAPABILITIES AND DIVERSIFICATION: BEYOND PATENTS

There is a large body of literature that refers to local capabilities to explain why regions specialize. Recently, there has been a shift of focus from capabilities sustaining existing specializations in regions, to capabilities that lay the foundation of new specializations in regions. Local capabilities can give birth to new activities by providing a pool of local resources, such as similar knowledge, skills and institutions. However, at the same time, they also set limits to what can be achieved in this diversification process. If a region does not possess the capabilities required for a new activity, it will be much harder and more risky to develop it. Therefore, one expects regions to diversify into new activities related to existing local activities, to build on their local capabilities. By contrast, unrelated diversification requires a complete transformation of local capabilities, which is accompanied by high transition costs and high risks of failure, and thus less likely to happen.

Scholars have incorporated the relatedness concept in S3 policy (Boschma 2014; Foray 2015; McCann and Ortega-Argilés 2015). This is because studies show that regions tend to develop new activities related to their existing activities (Hidalgo et al. 2007; Neffke et al. 2011; Boschma 2017; Hidalgo et al. 2018). These studies cover a range of new activities, such as new products or industries (Neffke et al. 2011; Boschma et al. 2013; Essletzbichler 2015; Boschma and Capone 2016; Guo and He 2017; Montresor and Quataro 2017; Zhu et al. 2017; Coniglio et al. 2018; He et al. 2018; He and Zhu 2019), new technologies (Kogler et al. 2013; Colombelli et al. 2014; Rigby 2015; Boschma et al. 2015; Tanner 2016; Balland et al. 2019) and new professions (Muneepeerakul et al. 2013; Shutter et al. 2016).

Balland et al. (2019) made the point that S3 policy should also look at the type of new activities that are promoted, as some activities might provide more growth potential for regions than other activities. Growth potential of new activities may be measured in various ways. The economic complexity literature (Hausmann and Hidalgo 2011) argues it is better for a region to develop new activities that are non-ubiquitous, meaning few other countries have the same specializations, which may indicate that they are highly complex and difficult to copy or imitate. More complex activities also tend to generate higher economic benefits for regions, as recent studies show (Davies and Mare 2019; Mewes and Broekel 2019).

However, despite the strong economic incentive to develop more complex technologies, it appears to be very difficult for regions in general to develop more complex technologies (Balland et al. 2019). Balland and Rigby (2017) investigated the geography of complex knowledge in the US, and observed that more complex technologies concentrate in the most urbanized regions. Fusillo et al. (2019) found that technologies that recombine diverse sources of knowledge are more concentrated in cities than would be expected from city size alone. So, it seems that the most advanced urban regions are better capable of developing the most complex technologies. More in general, Balland et al. (2019) found that regions are more likely to diversify into more complex technologies when related to existing technologies in a region. This suggests relatedness is needed to increase the complexity of a regional economy.

When applying this relatedness concept to S3 policy, many studies tend to employ patent data to identify diversification potentials for regions. What these studies take up is technological capabilities, and especially high-tech capabilities. For instance, patent data can be very useful to identify diversification potentials of regions in specific key technologies, like Artificial Intelligence. However, to get a more comprehensive picture of diversification potentials of regions in the EU, there is a need to broaden the capability measure, and to go beyond patents. This is also needed to capture better capabilities in peripheral regions that patent only to a limited extent, and which are focused more on low/medium-tech and low complex activities. This would respond to some degree to critique that S3 is expected to favor advanced regions rather than peripheral regions (McCann and Ortega-Argilés, 2015; Morgan 2015).

In this project, we explore the possibility to add other activities into this S3 framework by using data on occupation and sectors, each of which might take up different types of capabilities. We compare EU regions and investigate whether they are specialized in particular capabilities. Finally, we will map for each EU region whether they score high or low regarding their diversification potentials in new occupations and new sectors.

3. S3 FRAMEWORK BASED ON RELATEDNESS AND COMPLEXITY

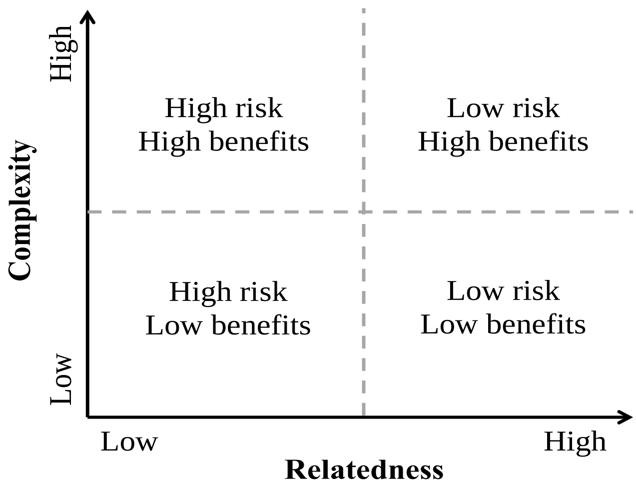
We use a new S3 framework proposed by Balland et al. (2019) that can map diversification opportunities of regions based on the local presence of relevant capabilities. This framework is based on two concepts: relatedness and complexity.

Relatedness provides an indicator of the cost of diversifying from existing activities to a new activity in a region. What does relatedness mean? Economic activities are considered related when they share similar capabilities. Cars and motor cycles are closely related because they rely on similar knowledge and skills in the field of engineering. In contrast, activities like pig production and nuclear energy are unrelated, as they have nothing in common in terms of capabilities: the knowledge and skills required for pig production are completely different from nuclear energy production. So, the more related a potential new activity is to existing activities in a region, the lower the costs (and thus less risky) to develop this new activity.

But potential new economic activities are not only targeted by regions in S3 because they have relevant capabilities to develop those. S3 also aims to develop new economic activities that bring more complexity to the economy of regions. What does complexity mean? Economic activities are complex when they rely on a wide range of sophisticated capabilities that are hard to combine. As a result, complex activities are almost impossible to imitate and therefore can provide a source of high economic rents. Activities that are simple to copy tend to be of little economic value (Hidalgo and Hausmann, 2009). Complexity provides a way of assessing the potential economic benefits of diversifying into a new activity: the higher the economic complexity of this activity, the higher the potential economic benefits.

Balland et al. (2019) developed a S3 policy framework that assesses diversification opportunities of regions in terms of their scores on relatedness (costs) and complexity (benefits). This is illustrated in Figure 3. An attractive S3 policy supports potential new activities that occupy the north-east quadrant, for these activities promise above average returns and can be developed at relatively low risk. This stands in contrast to a S3 focus on potential new activities situated in the south-west quadrant which represent low complex activities (low benefits) that are hard to develop because the region lacks the relevant capabilities (high risk). The north-west quadrant reflects a potentially high-benefit S3 strategy, but it is not rooted in regional capabilities: it aims at developing new activities from scratch and therefore such S3 policy is more likely to fail. The fourth policy option focuses on activities in the south-east quadrant, which reflects a relatively low risk strategy because it builds on related capabilities but these potential activities offer little economic benefits.

Figure 3. S3 policy framework



Source: Balland et al. 2019, p. 1259

4. MEASURES OF RELATEDNESS AND COMPLEXITY

In this S3 framework, one can assess diversification opportunities of regions in new technologies, new occupations and new sectors. This requires measuring the degree of relatedness between technologies, between occupations, and between sectors. Moreover, it requires measuring the degree of complexity of technologies, occupations and sectors. As the main focus of this report (beyond patents) is on occupations and sectors, we explain how we measured relatedness and complexity for occupations and sectors. For this purpose, we made a special extraction of occupation and sector data from the EU Labour Force Survey Database[1]. We use data on 40 occupations (2-digit ISCO) and 88 sectors (2-digit NACE).

Measuring relatedness between sectors/occupations

As explained before, activities are considered related when they share similar capabilities. We derive relatedness between sectors/occupations from normalized employment co-location patterns. Two sectors/occupations are considered related if they are simultaneously over-represented in the same regions, as in Hidalgo et al. (2007). See Appendix 1 for details on the relatedness measures.

Figures 1 and 2 show the degree of relatedness between all 2-digit sectors and all 2-digit occupations respectively for the EU as a whole. Colours indicate groups of highly related activities[2]. Some sectors/occupations are positioned more central in the networks, meaning they are share similar capabilities with many other sectors/occupations, while other sectors/occupations are related with only a few sectors/occupations, and thus positioned more in the periphery of these networks. For individual regions, we also measure the degree of relatedness between an occupation/sector and all other occupations/sectors in a region (Hidalgo et al., 2018; Balland et al., 2019). The higher this measure of 'relatedness density', the more related an occupation/sector is to other occupations/sectors in a region (Appendix 1).

[1] This extraction is based on the best sample available for a given year, and the variables corresponds either to one single quarter (generally quarter 2), or to a sub-sample distributed along the year (sample size equivalent to one reference quarter).

[2] The algorithm is computed using the igraph R package and implements the multi-level modularity optimization algorithm for finding community structure See VD Blondel, J-L Guillaume, R Lambiotte and E Lefebvre: (http://arxiv.org/abs/arXiv:0803.0476 for the details.

Figure 1. Industry Space (economic sectors)

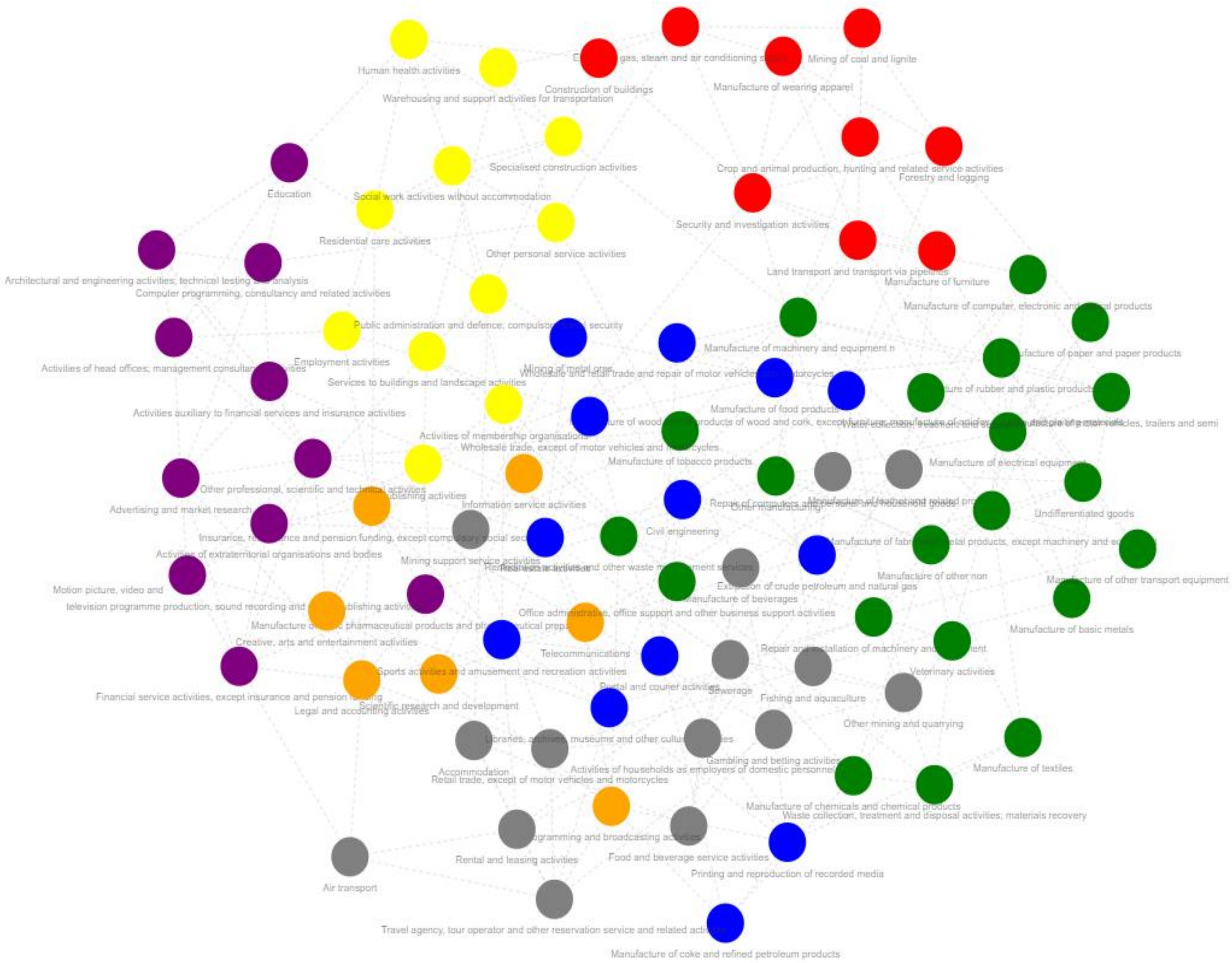
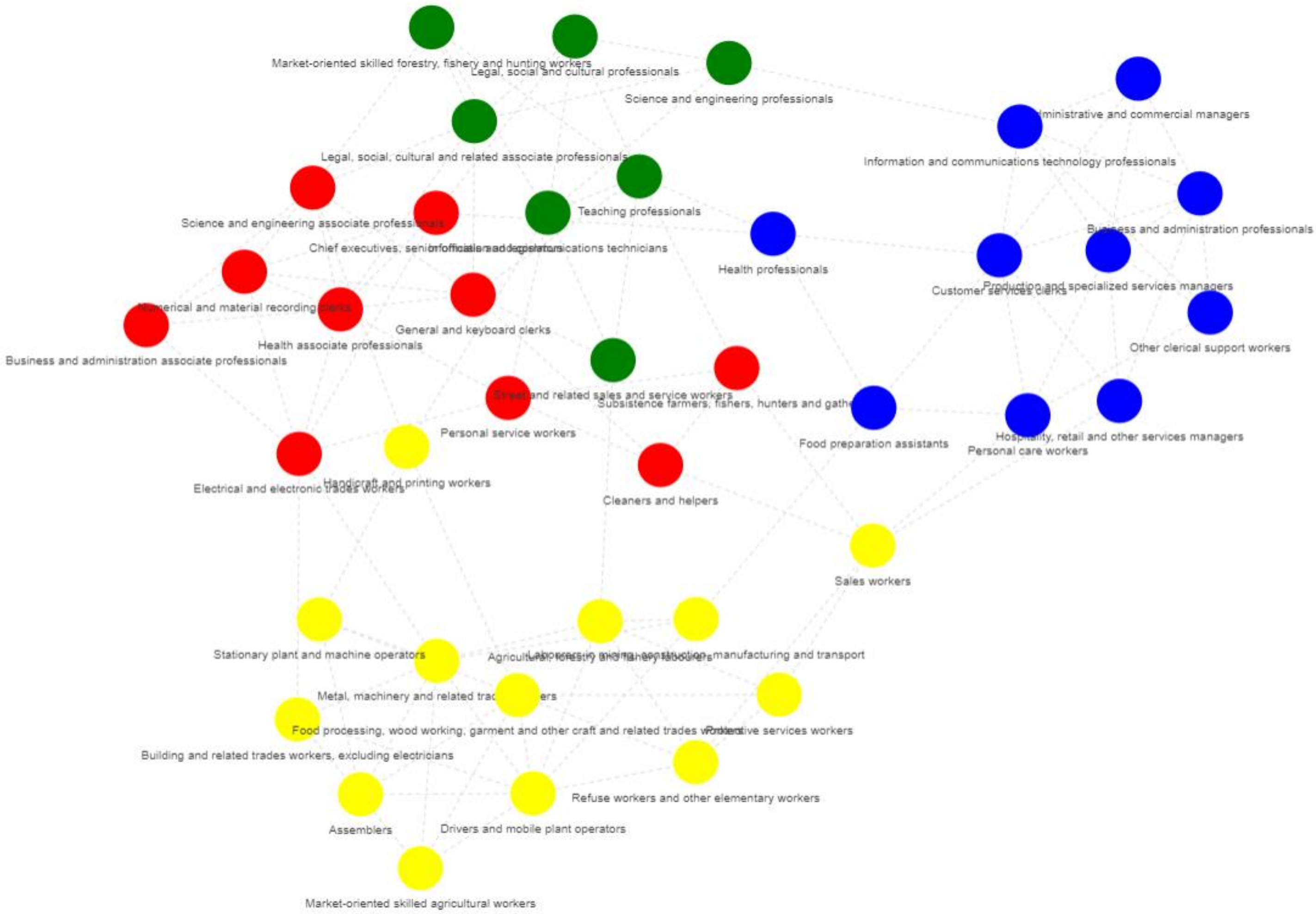


Figure 2. Occupation Space



Measuring the complexity of sectors/occupations

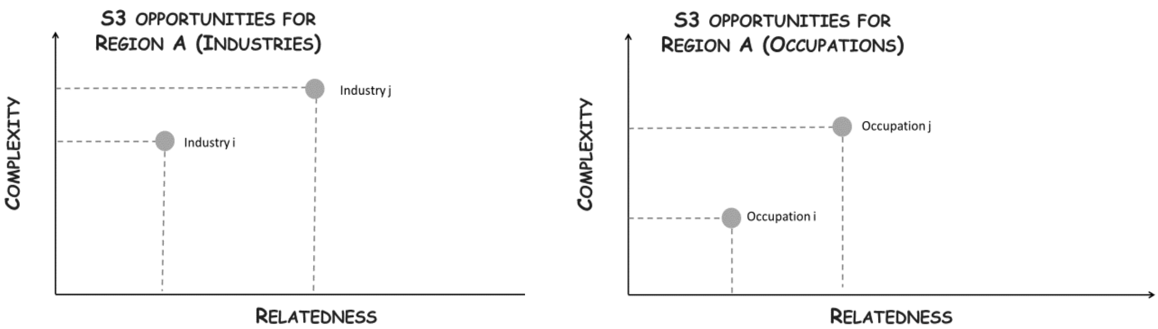
As explained before, a complex activity is what every region would like to master but that very few can: it gives monopoly rents, is hard to develop (Hidalgo and Hausmann 2009), and it is very concentrated in space (Balland and Rigby 2017). Complex activities rest on the recombination of many elements that are arranged in a specific way. We calculated a complexity measure that combines the method of reflection (Hidalgo and Hausmann 2009) with insights from the scaling literature (Balland et al. 2019). Sectors and occupations will rank high in terms of complexity when having a high level of relatedness in densely and largely populated regions. See Appendix 2 for technical details. In Appendix 3, we list the 10 most complex and the 10 least complex (2-digit) occupations and (2-digit) sectors. The most complex occupations turn out to be 'Information and Communication Technology Professionals' and 'Business and Administration Professionals', while the most complex sectors are 'Computer Programming, Consultancy and Related Activities' and 'Activities of Head Offices; Management Consultancy Activities'.

5. DIVERSIFICATION OPPORTUNITIES OF EUROPEAN REGIONS: OCCUPATIONS AND SECTORS

How can we determine the diversification opportunities of regions? Studies have used primarily patent data to map technological diversification opportunities of regions. However, patent data might not fully capture the whole range of capabilities and diversification opportunities available to regions. This is especially true for peripheral regions in the EU that tend to show low patent activity. Therefore, it is crucial for S3 policy to assess diversification options for regions that include other forms of capabilities that go beyond patents.

In this project, we use data on 40 occupations (2-digit) and 88 sectors (2-digit) that are derived from the EU Labour Force Survey Database. We assess the potentials of 292 NUTS-2 regions in Europe to develop new occupations and to develop new industries, following the S3 policy framework in Figure 1. The NUTS-2 regions include regions in all EU countries and the four EFTA-countries (Iceland, Liechtenstein, Norway and Switzerland). For each region in Europe, we map the relatedness to all occupations and sectors (x-axis) and the level of complexity of all these occupations and sectors (y-axis), as in Figure 3. We plot all missing industries and all missing occupations in a region and show their scores on relatedness and complexity. As shown in Figure 3, the rationale behind this S3 policy framework is that a region should aim to target and develop a new industry j (and not industry i) and a new occupation j (and not occupation i), as these would involve less risks (because of high relatedness with existing industries/occupations in the region) and would generate potentially higher economic benefits (due to higher complexity) than industry i and occupation i.

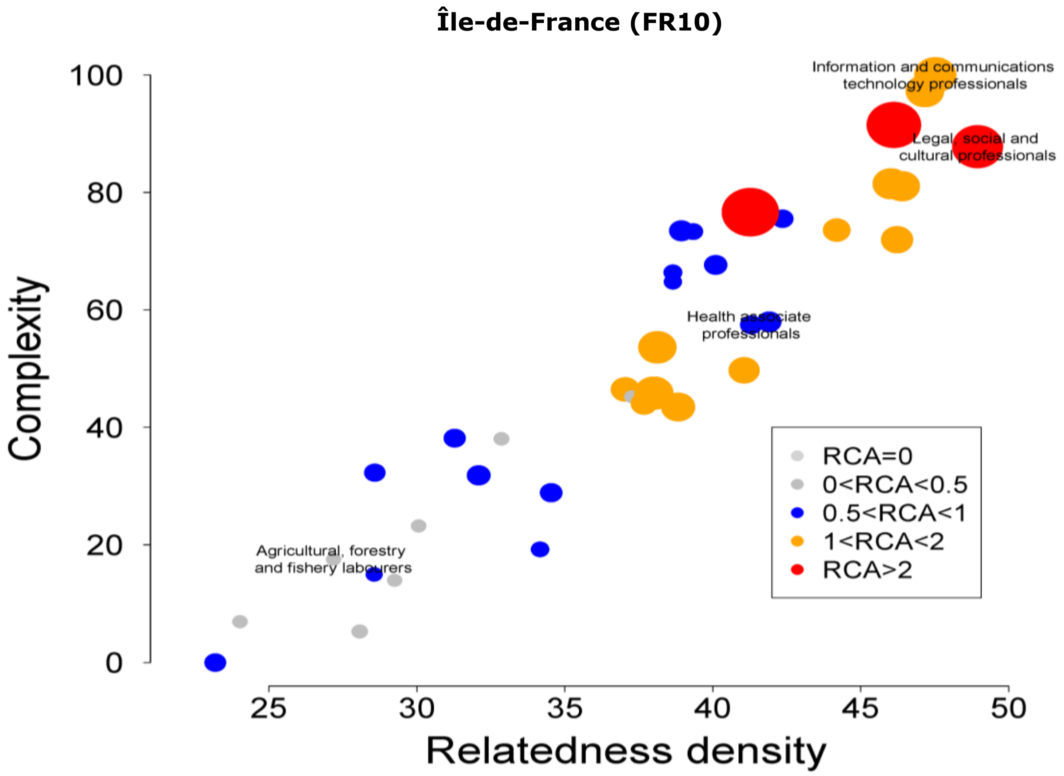
Figure 3. Smart Specialization Framework for industries and occupations (Balland et al. 2019)

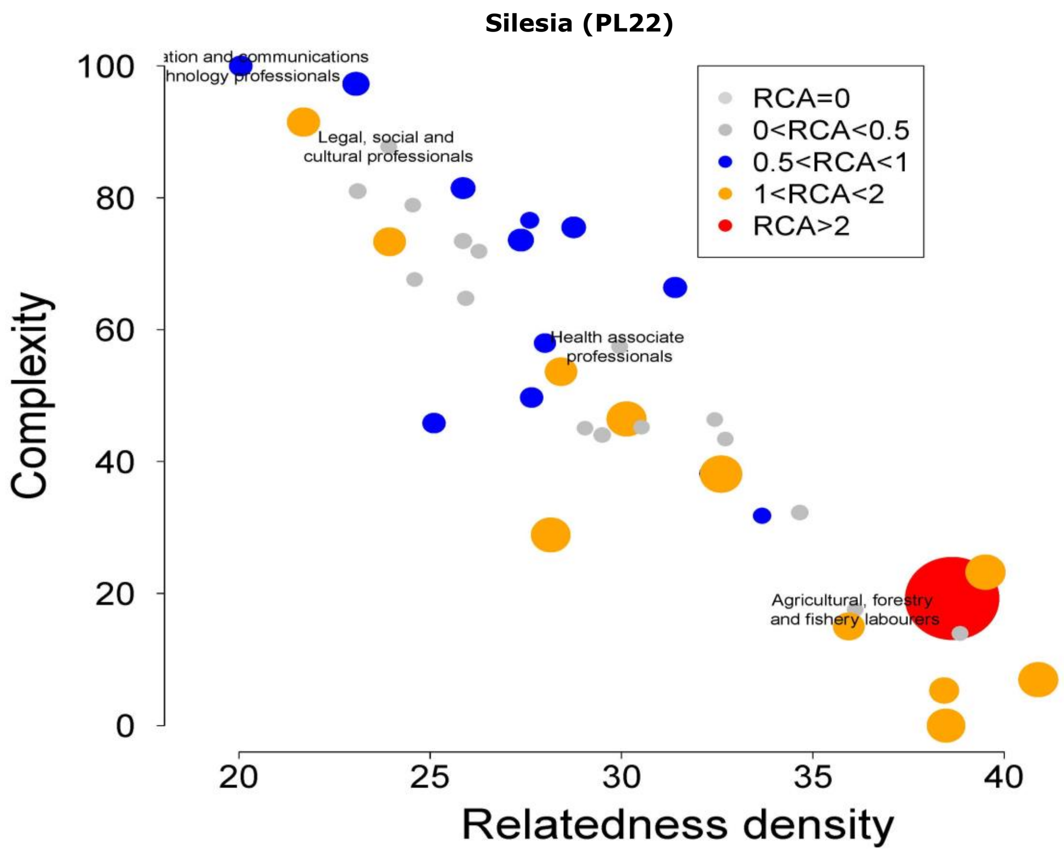


The first finding is that the 292 European regions differ tremendously in terms of diversification opportunities that can be potentially targeted in their S3 strategy. Some regions like the Île-de-France region have many diversification opportunities, while regions like the Extremadura region in Spain have few diversifications options. European regions also differ in the types of activities in which they show a high development potential. Moreover, we find regions differ widely in terms of capabilities that can be used as input for their S3 strategy. A region may score relatively high with respect to diversification potentials in new occupations, but low on diversification potentials in new industries, or new technologies, in contrast to other regions. This shows how crucial it is to account for different types of capabilities to assess diversification opportunities of regions, instead of relying on one type alone.

For illustrative purposes, we present some results for three regions, representing a major urban region (Île-de-France) in Western Europe, an old industrial region (Silesia) in Eastern Europe, and a peripheral region (Extremadura) in Southern Europe. In Figures 4 and 5, we present their diversification opportunities for occupations and sectors respectively in which the region is not yet specialized (meaning they have a Relative Comparative Advantage (RCA) <1), which are marked as blue and grey nodes. Within this broad category, we made three sub-categories: RCA=0, 0<RCA<0.5 and 0.5<RCA<1.0. We also included in these figures occupations and sectors the region is already specialized in (meaning they have a Relative Comparative Advantage RCA>1). These are marked as red and yellow nodes, and consist of two sub-categories: 1<RCA<2 and RCA>2.

Figure 4 Diversification opportunities in new occupations in Île-de-France (FR10), Silesia (PL22) and Extremadura (ES43)





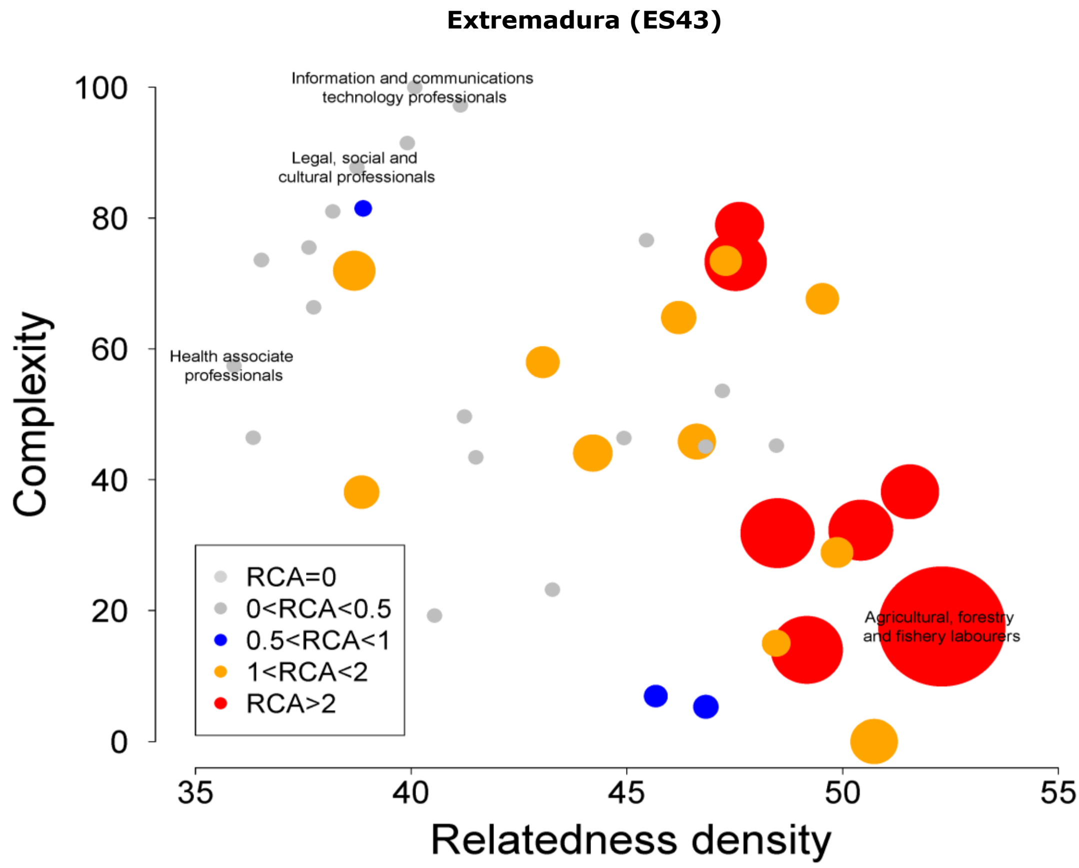
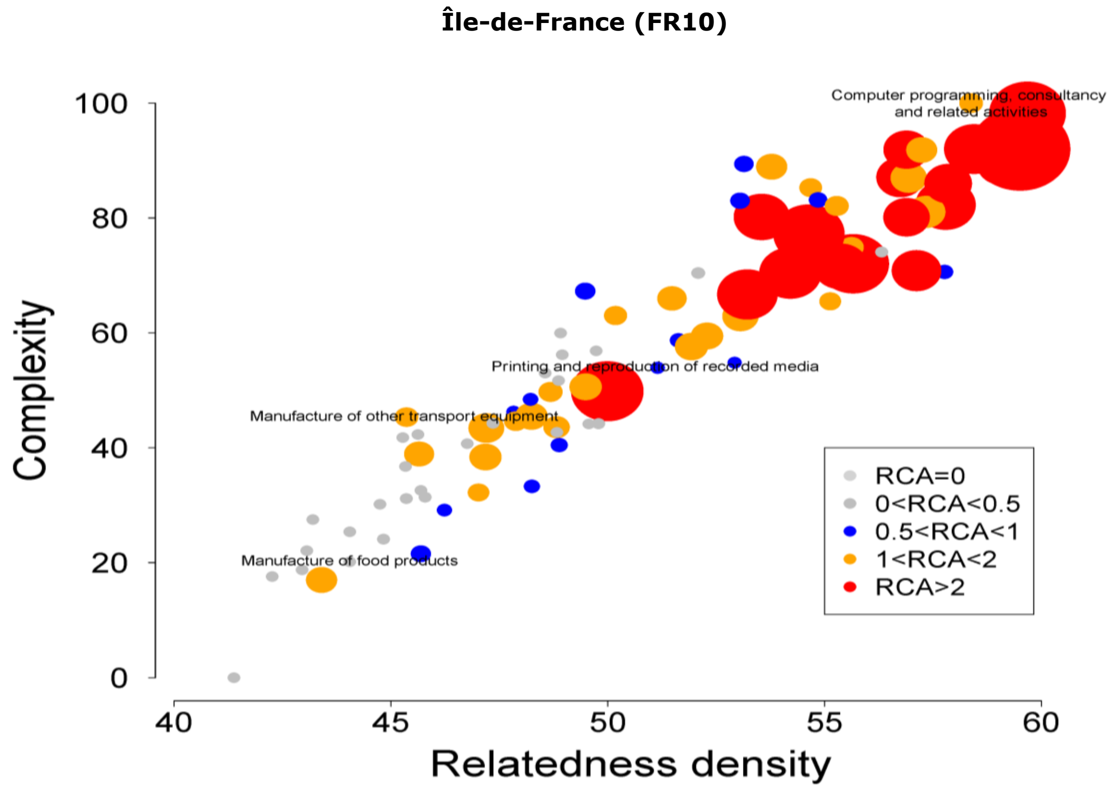
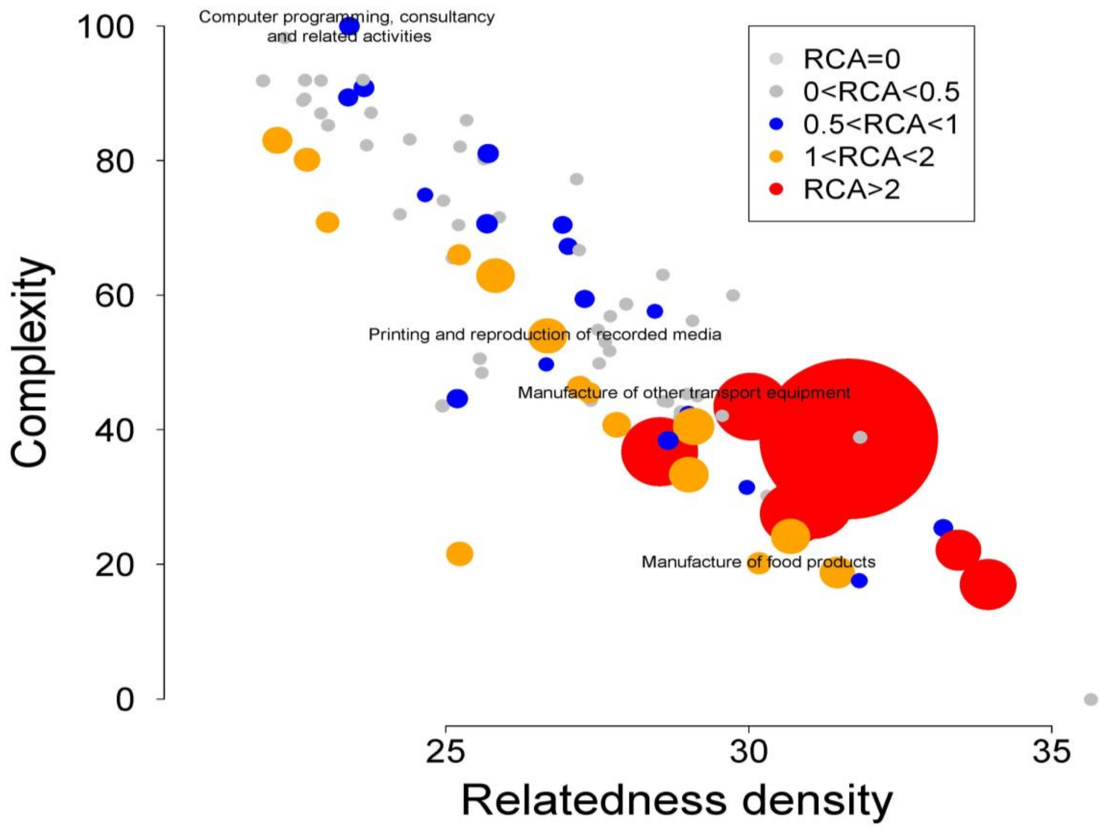


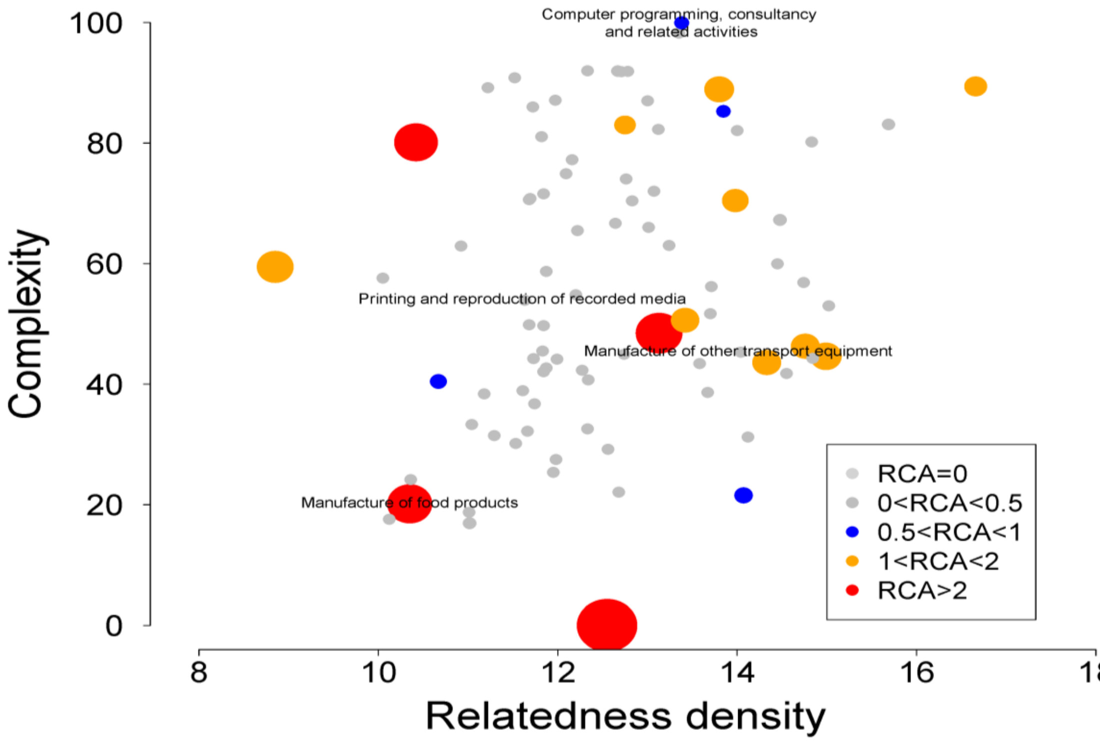
Figure 5. Diversification opportunities in new industries in Île-de-France (FR10), Silesia (PL22), and Extremadura (ES43)



Silesia (PL22)



Extremadura (ES43)



On average, existing specializations in a region (red and yellow nodes) tend to score higher on relatedness density than missing specializations in a region (blue and grey nodes). This does not come as a surprise: activities in a region are more likely to be underdeveloped when they cannot build on related specializations in a region, while existing specializations benefit from being co-located with other specializations with which they share similar capabilities. Moreover, when we look at the diversification opportunities for occupations and sectors in which the region is not yet specialized, one can observe big differences between the three regions. Île-de-France shows outstanding diversification opportunities in highly complex activities such as 'Legal, Social and Cultural Professionals' and 'Activities of Head Offices; Management Consultancy Activities'. The set of diversification opportunities of Silesia looks very different: these mainly concern manufacturing activities with low complexity. However, Silesia has some potential in some complex activities, like 'Science and Engineering Associate Professionals'. Extremadura shows little diversification opportunities in sectors more in general, scoring low on relatedness density. Nevertheless, it also shows some potentials in more complex occupations like 'Personal Care Workers'. To illustrate this further, we list the top 10 diversification opportunities in Île-de-France, Silesia and Extremadura in Appendix 4 for both occupations and sectors.

These findings bring to light a broader range of diversification potentials in a region than if one would have relied on one indicator only. We observe that a region may have high diversification potentials in occupations and low diversification potentials in sectors. A region may also show high diversification potentials in both occupations and sectors while the occupations and sectors concerned have little in common. This shows how crucial it is to account for different types of diversification opportunities of regions (as proxied by occupations and sectors), instead of relying on one type of diversification opportunities alone.

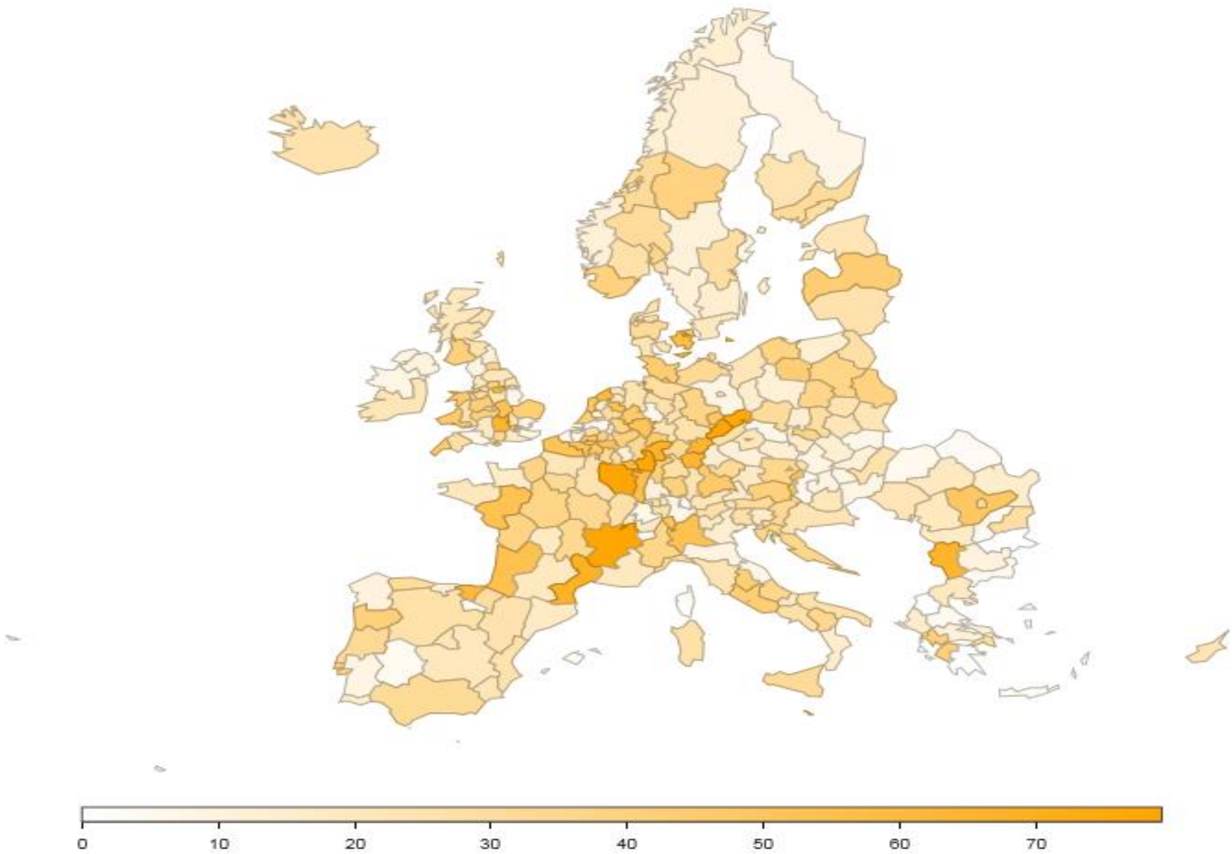
6. DIVERSIFICATION OPPORTUNITIES OF EUROPEAN REGIONS IN SOME KEY TECHNOLOGIES

While the main focus of this report is on occupations and sectors, patent data can be very useful for identifying diversification opportunities of regions in specific key technologies. Often these key technologies are not easy to link to particular occupations and sectors. To illustrate this, we have also analyzed, within the same S3 framework, the extent to which European regions have potentials to develop new technologies that are regarded as most promising. We present results for 7 key technologies: (1) clean, connected and autonomous vehicles; (2) cyber-security; (3) hydrogen technologies and systems; (4) micro-electronics; (5) batteries; (6) high performance computing; and (7) additive manufacturing.

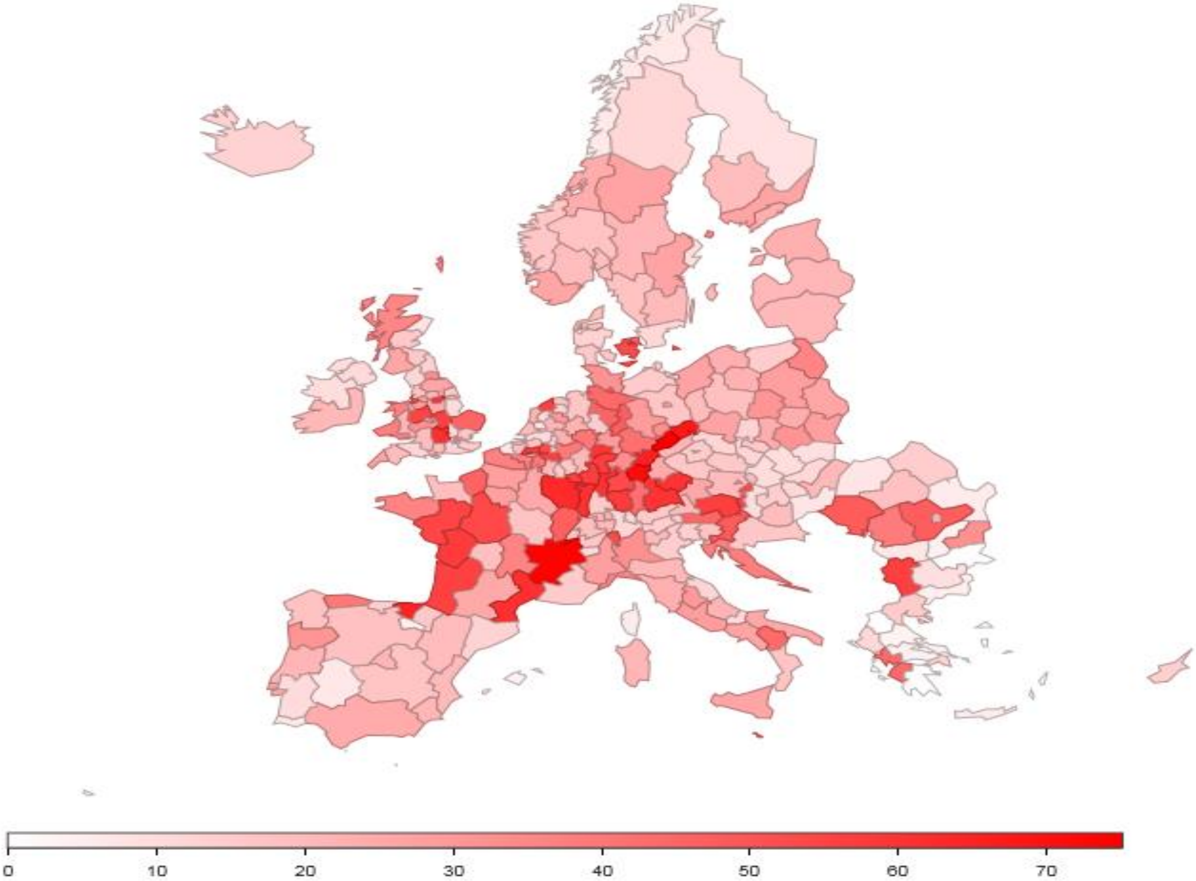
We used patent data (OECD REGPAT dataset 2018 version) and assigned them to specific 4-digit CPC patent classes. In Appendix 5, we present the list of CPC classes attached to each of those 7 key technologies. Then, we calculated the degree of relatedness between each key technology and all other technologies (total of 654 CPC technology classes) through co-occurrence analysis, following other studies (Boschma et al. 2015). When the same two technology classes are often combined on a patent document, this is interpreted as an indicator of relatedness between the two technologies. This information on relatedness is used to calculate a relatedness density measure, as in Balland et al. (2019), that is similar to our relatedness density measure for occupations and sectors (see Appendix 1). The higher relatedness density, the more related a key technology is to other technologies in a region, and the higher the potential of that region to develop that key technology.

To measure complexity of a key technology, we used the NK model proposed by Fleming and Sorenson (2001). It accounts for how many sub-classes are listed on a patent (N) and how often these sub-classes have been recombined in the past (K). This is based on the idea that the complexity of a system can be defined by the number of its elements (N) and the way that these elements interact (K). A patent that list many sub-classes will be considered as more complex than a patent that list few. Likewise, a patent that recombine classes that have rarely been combined in the past will be considered as more complex than a patent that makes typical re-combinations. We then average this score per technology category.

Map 1. Potentials of European regions in hydrogen technologies



Map 2. Potentials of European regions in battery technologies



The two maps show that European regions differ widely in terms of their potentials to contribute to these two key technologies. The top 5 regions in Europe for hydrogen are the Brussels region (BE10), Rhône-Alpes (FR71), Lorraine (FR41), Chemnitz (DED4) and Dresden (DED2). The top 5 regions in Europe for batteries are Chemnitz (DED4), Rhône-Alpes (FR71), Mittelfranken (DE25), Dresden (DED2) and Oberfranken (DE24).

We also show the potentials of Île-de-France, Silesia and Extremadura for all 7 key technologies in Figures 6, 7 and 8 respectively, for illustrative purposes. The size of the dots stands for RCA: the bigger the dot, the higher the RCA of that key technology in the region. What one can observe is that the Île-de-France region has high potentials in a number of key technologies like cyber-security and high performance computing, while the Silesia region has a high potential in additive manufacturing (3D-printing) in particular. By contrast, Figure 8 shows that the Extremadura region has no potential whatsoever in all 7 key technologies.

Figure 6. Potential of the Île-de-France region to develop 7 key technologies

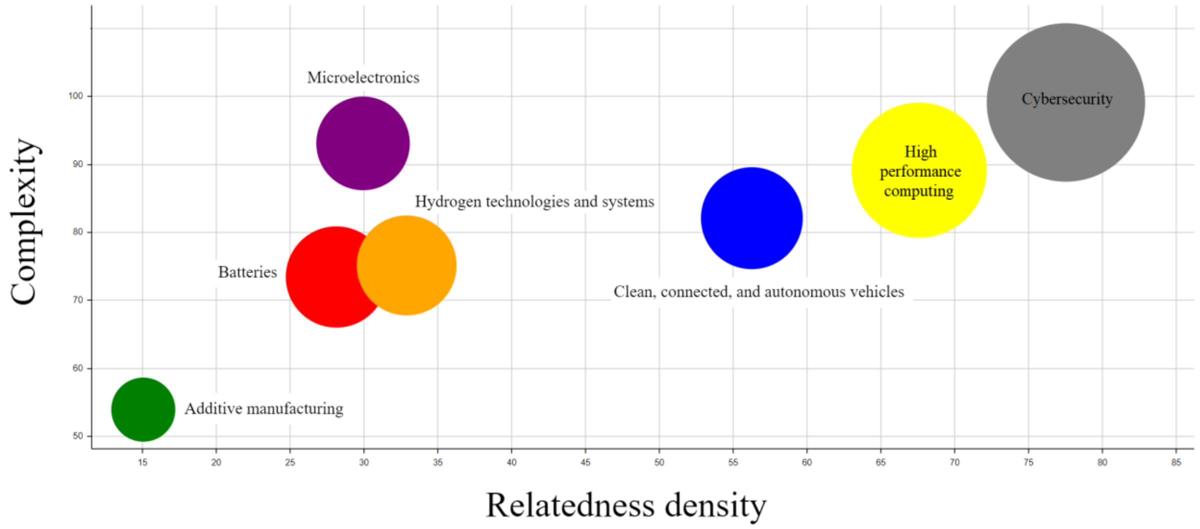


Figure 7. Potential of Silesia region to develop 7 key technologies

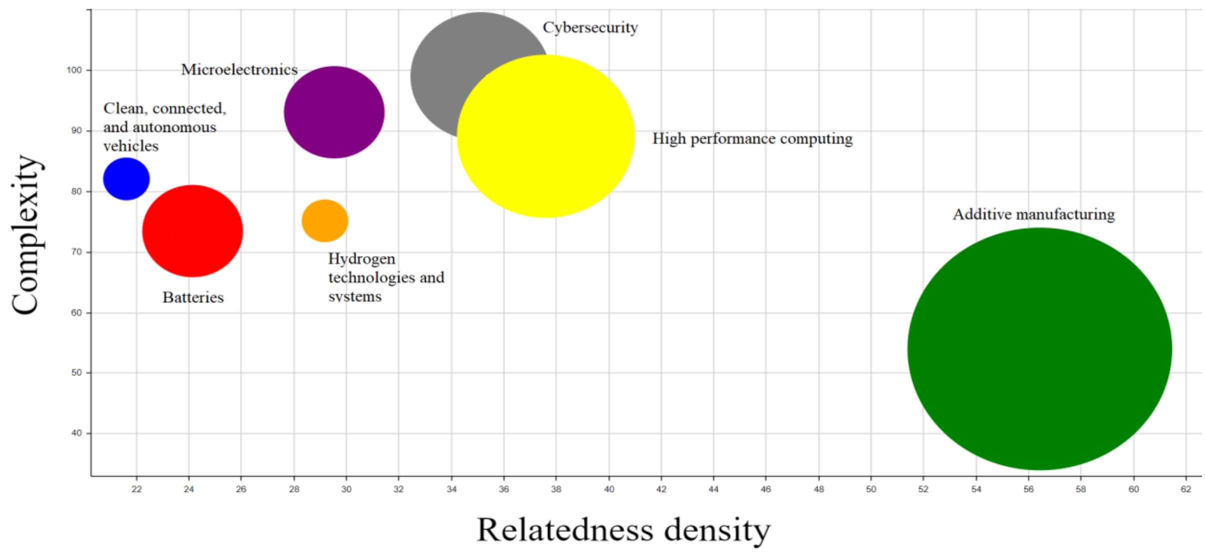
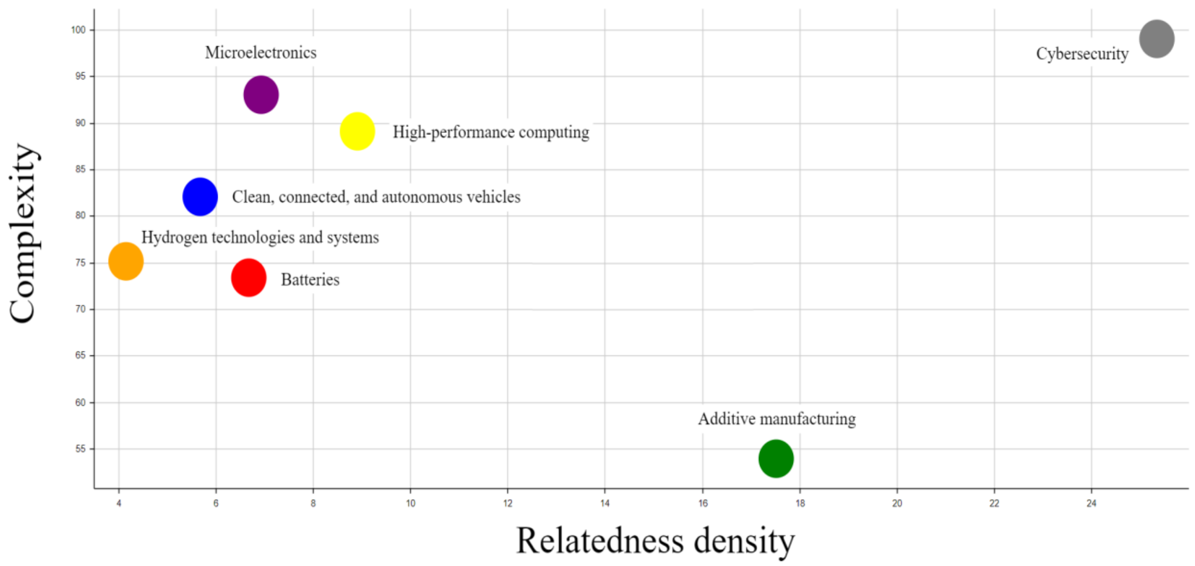


Figure 8. Potential of Extremadura region to develop 7 key technologies



7. CONCLUSIONS AND POLICY IMPLICATIONS

In Smart Specialization (S3) policy, the role of regional capabilities is a key point of departure. Regions should build on existing capabilities to develop new activities and upgrade their local economies, because regional capabilities condition which new activities are feasible to develop. This begs the question how capabilities should be taken on board.

Studies have used product, industry, occupation or patent data to measure capabilities. In doing so, scholars have tried to assess the extent to which capabilities are shared across different activities, and which capabilities provide diversification opportunities to regions (Boschma 2017). In these studies, it is common to use one indicator only. This might lead to some bias, as each indicator tends to measure a particular type of capabilities, like patent data take up capabilities associated with high-tech and analytical knowledge, while ignoring other (non-tech or low-tech) capabilities that might still be relevant for identifying diversification opportunities in regions. For instance, it has been argued that a too narrow focus on technological capabilities through patent data might underestimate the diversification potentials of peripheral regions that in general score low on high-tech activities.

In this report, we made an attempt to account for different types of capabilities. We have used data on occupations and industries to identify diversification potentials of EU regions, and applied it to the S3 policy framework introduced by Balland et al. (2019) that is based on the concepts of relatedness and complexity. The use of different datasets turned out to be important, as the analyses show that EU regions differ tremendously with respect to the types of capabilities they have. This is not only the case with respect to capabilities based on one type of data (such as patents, industries or occupations). Regions show also different capabilities taken up by different data. A region may score generally low on diversification potentials in new technologies, while scoring relatively high with respect to diversification potentials in new occupations or new industries. In other words, our analyses bring to light a broader range of diversification potentials in a region than if one would have relied on one indicator only. This shows how crucial it is to account for different types of diversification opportunities of regions (as proxied by technologies, occupations and industries), instead of relying on one type of diversification opportunities alone.

This account of a range of diversification potentials of regions can be useful to assess the S3 strategies that are currently developed and implemented in the EU. Such an exercise could bring to light and measure the extent to which there is an overlap between diversification potentials in a region and the region's own prioritization, as embodied in its S3 strategy. If there is little overlap, there might a good reason to rethink and reconsider its S3 policy. However, while the statistical data analyses in this report are informative for identifying diversification potentials in a region, the final prioritization should be done through the entrepreneurial discovery process, in collaboration with local stakeholders. These two approaches need to go hand in hand, to construct an evidence-based S3 policy in Europe.

APPENDIX 1. MEASURES OF RELATEDNESS

We measure relatedness between sectors and between occupations from normalized employment co-location patterns. Two sectors/occupations will be considered as related if they tend to be simultaneously over-represented in the same regions. This measure comes close to the product space methodology in which two products are related if they are co-exported by the same countries (Hidalgo et al., 2007; Hidalgo et al., 2018). The main difference with the product space methodology, however, is that we do not dichotomize the index of Relative Comparative Advantage to compute relatedness. Using the continuous value (or employment shares directly), as noted by Davies and Maré (2019) allows to limit measurement error. Instead, relatedness is given by the correlation coefficient between continuous RCAs, where RCA is given by the ratio of the employment share of a given activity *i* in a given region *r* compared to the reference region (EU as a whole):

$$RCA_{r,i} = \frac{employment_{r,i} / \sum_i employment_{r,i}}{\sum_r employment_{r,i} / \sum_r \sum_i employment_{r,i}}$$

The relatedness between activities (industries/occupations) *i* and *j* is computed from the correlation between the vectors ($RCA_{1,i}$, $RCA_{2,i}$, … $RCA_{292,i}$) and ($RCA_{1,j}$, $RCA_{2,j}$, … $RCA_{292,j}$) and can be formalized as a network, the *occupation/industry space*, a *n*n* network where the individual nodes *i* (i =1,…, n) represent activities (88 2-digit NACE codes, 40 2-digit ISCO-codes), and the links between them indicate their degree of relatedness.

We also measure the degree of relatedness between an occupation/sector and other occupations/sectors for individual regions, expressed as relatedness density (Hidalgo et al., 2018; Balland et al., 2019). For each region *r*, we calculated the density of employment in the vicinity of individual occupations or sectors *i*. Following Hidalgo et al. (2007) and Balland et al. (2019), the density of employment around a given sector/occupation *i* in region *r* is derived from the relatedness $\varphi_{i,j,t}$ of sector/occupation *i* to all other sectors/occupations *j* in which the region has relative comparative advantage (RTA), divided by the sum of relatedness of sector/occupation *i* to all the other sectors/occupations *j* in the reference region (Europe):

$$\text{RELATEDNESS\_DENSITY}_{i,r} = \frac{\sum_{j \in r, j \neq i} \phi_{ij}}{\sum_{j \neq i} \phi_{ij}} * 100$$

APPENDIX 2. MEASURES OF COMPLEXITY

Hidalgo and Hausmann (2009) paved the way to the empirical analysis of economic complexity with their method of reflection. It is an outcome-based approach based on the idea that products that are not produced in many countries are complex as long as these countries also tend to be diverse. Although this method works well when there is a strong spatial structure in the data (trade data at the country level, patent data in metropolitan areas), it might produce unreliable rankings in other cases. As noted by Davies and Maré (2019), measurement error arises in particular from small perturbations in employment in regions with location quotients around 1. Using EconGeo (Balland, 2017), we computed Knowledge Complexity Index and confirmed this finding. Therefore, we propose an alternative solution by combining Hidalgo and Hausmann's seminal idea with the scaling literature that shows that complex economic activities concentrate in large cities (Balland et al., 2019).

Since our analyses are based at the regional level, using regional population directly as a proxy for urbanization economies is not ideal. For instance, one of the largest regions in terms of population, Andalusia (ES61), is composed of 8 cities. On the other hand, the Brussels region (BE10) has only around 1 million inhabitants but represents almost the entire Brussels metro area. In a first step, we compute urban scale by multiplying population by population density[3] which is a proxy for interaction potential. We then use urban scale as a weighting scheme to compute the relatedness density weighted average for each industry/occupation:

$$\text{COMPLEXITY}_{i,r} = \frac{\sum_c \left(\left(\frac{\sum_{j \in r, j \neq i} \phi_{ij}}{\sum_{j \neq i} \phi_{ij}} * 100 \right)_c * (Population_c * Density_c) \right)}{\sum_i c}$$

Industries/occupations with a high level of relatedness in densely and largely populated regions will rank higher in terms of complexity[4]. In the large majority of cases, this approach appears to produce reliable and robust results, but some oddities can sometimes appear by construction. Therefore, we recommend to systematically examine the rankings of complexity carefully before using them for research or for policy making decisions. As usual with data science analysis, humans are needed to provide context and fully make sense of them.

[3] Technically it also comes down to dividing the square terms of population by the land area. This is a way to take into account the network-based idea from the scaling literature that the potential for interactions (proxied by n2) rather than population itself (n) matters for agglomeration economies. We note that results are robust to using simple population density (pop/land).

[4] The results are consistent with other measures, such as average years of schooling by occupations and industries.

APPENDIX 3. COMPLEXITY OF OCCUPATIONS AND SECTORS

Table A1: Top 10 occupations in terms of their degree of complexity

Rank	ISCO code	Complexity	Description of Occupation
1	250	100	Information and communications technology professionals
2	240	95.8	Business and administration professionals
3	260	85.9	Legal, social and cultural professionals
4	120	84.4	Administrative and commercial managers
5	340	77.5	Legal, social, cultural and related associate professionals
6	210	76.1	Science and engineering professionals
7	330	74.6	Business and administration associate professionals
8	350	70.2	Information and communications technicians
9	940	69.0	Food preparation assistants
10	220	68.4	Health professionals

Table A2: Bottom 10 occupations in terms of their degree of complexity

Rank	ISCO code	Complexity	Description of Occupation
40	830	0.0	Drivers and mobile plant operators
39	610	2.8	Market-oriented skilled agricultural workers
38	750	3.1	Food processing, wood working, garment and other craft and related trades workers
37	720	3.6	Metal, machinery and related trades workers
36	920	7.0	Agricultural, forestry and fishery labourers
35	710	11.0	Building and related trades workers, excluding electricians
34	810	11.3	Stationary plant and machine operators
33	820	13.4	Assemblers
32	520	26.3	Sales workers
31	630	26.9	Subsistence farmers, fishers, hunters and gatherers

Table A3: Top 10 industries in terms of their degree of complexity

Rank	NACE code	Complexity	Description of Industry
1	J62	100	Computer programming, consultancy and related activities
2	M70	95.5	Activities of head offices; management consultancy activities
3	Q86	95.4	Human health activities
4	M71	92.9	Architectural and engineering activities; technical testing and analysis
5	Q88	91.2	Social work activities without accommodation
6	K66	88.5	Activities auxiliary to financial services and insurance activities
7	M73	86.7	Advertising and market research
8	K64	86.5	Financial service activities, except insurance and pension funding
9	M74	85.3	Other professional, scientific and technical activities
10	P85	84.7	Education

Table A4: Bottom 10 industries in terms of their degree of complexity

Rank	NACE code	Complexity	Description of Industry
88	A1	0.0	Crop and animal production, hunting and related service activities
87	C25	18.5	Manufacture of fabricated metal products, except machinery and equipment
86	C22	20.2	Manufacture of rubber and plastic products
85	C31	20.7	Manufacture of furniture
84	C10	21.8	Manufacture of food products
83	C16	22.3	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
82	C27	22.8	Manufacture of electrical equipment
81	T98	23.4	Undifferentiated goods- and services-producing activities of private households for own use
80	H49	26.4	Land transport and transport via pipelines
79	E36	26.6	Water collection, treatment and supply

APPENDIX 4. TOP 10 DIVERSIFICATION OPPORTUNITIES IN THE ÎLE-DE-FRANCE, THE SILESIA REGION, AND THE EXTREMADURA REGION IN TERMS OF OCCUPATIONS AND INDUSTRIES

Last column: in brackets, ranking in complexity

Table A5: **Île-de-France (FR10)**: top 10 potentials in occupations and industries

Top 10 potential **occupations** in Île-de-France (FR10)

ISCO	Name (2-digit)	Relatedness	Complexity
260	Legal, social and cultural professionals	52	86 (3)
250	Information and communications technology professionals	51	100 (1)
240	Business and administration professionals	51	96 (2)
340	Legal, social, cultural and related associate professionals	50	78 (5)
120	Administrative and commercial managers	49	84 (4)
350	Information and communications technicians	49	70 (8)
210	Science and engineering professionals	49	76 (6)
330	Business and administration associate professionals	48	75 (7)
130	Production and specialized services managers	45	68 (12)
410	General and keyboard clerks	45	52 (18)

Top 10 potential **industries** in Île-de-France (FR10)

NACE	Name (2-digit)	Relatedness	Complexity
70	Activities of head offices; management consultancy activities	61	95 (2)
59	Motion picture, video and television programme production, sound recording and music publishing activities	61	85 (11)
73	Advertising and market research	61	87 (7)
74	Other professional, scientific and technical activities	60	85 (9)
64	Financial service activities, except insurance and pension funding	60	86 (8)
62	Computer programming, consultancy and related activities	60	100 (1)
51	Air transport	59	76 (19)
72	Scientific research and development	59	81 (16)
71	Architectural and engineering activities; technical testing and analysis	59	92 (4)
90	Creative, arts and entertainment activities	59	82 (15)

Table A6: **Silesia (PL22)**: top 10 potentials in occupations and industries

Top 10 potential **occupations** in Silesia (PL22)

ISCO	Name (2-digit)	Relatedness	Complexity
720	Metal, machinery and related trades workers	44	4 (37)
820	Assemblers	43	13 (33)
810	Stationary plant and machine operators	42	11 (34)
830	Drivers and mobile plant operators	42	0 (40)
750	Food processing, wood working, garment and other craft and related trades workers	42	3 (38)
710	Building and related trades workers, excluding electricians	40	11 (35)
740	Electrical and electronic trades workers	36	35 (25)
610	Market-oriented skilled agricultural workers	36	3 (39)
310	Science and engineering associate professionals	34	37 (23)
920	Agricultural, forestry and fishery labourers	34	7 (36)

Top 10 potential **industries** in Silesia (PL22)

NACE	Name (2-digit)	Relatedness	Complexity
49	Land transport and transport via pipeline	36	26 (80)
31	Manufacture of furniture	36	21 (85)
14	Manufacture of wearing apparel	35	27 (77)
5	Mining of coal and lignite	35	29 (73)
22	Manufacture of rubber and plastic products	34	20 (86)
80	Security and investigation activities	34	38 (61)
36	Water collection, treatment and supply	34	27 (79)
1	Crop and animal production, hunting and related service activities	34	0 (88)
29	Manufacture of motor vehicles, trailers and semi	34	28 (75)
27	Manufacture of electrical equipment	33	23 (82)

Table A7: **Extremadura (ES43)**: top 10 potentials in occupations and industries

Top 10 potential **occupations** in Extremadura (ES43)

ISCO	Name (2-digit)	Relatedness	Complexity
920	Agricultural, forestry and fishery labourers	55	7 (36)
540	Protective services workers	54	35 (26)
830	Drivers and mobile plant operators	54	0 (40)
930	Labourers in mining, construction, manufacturing and transport	53	30 (27)
520	Sales workers	53	26 (32)
530	Personal care workers	53	59 (15)
610	Market-oriented skilled agricultural workers	53	3 (39)
710	Building and related trades workers, excluding electricians	52	11 (35)
960	Refuse workers and other elementary workers	52	28 (30)
420	Customer services clerks	51	68 (13)

Top 10 potential **industries** in Extremadura (ES43)

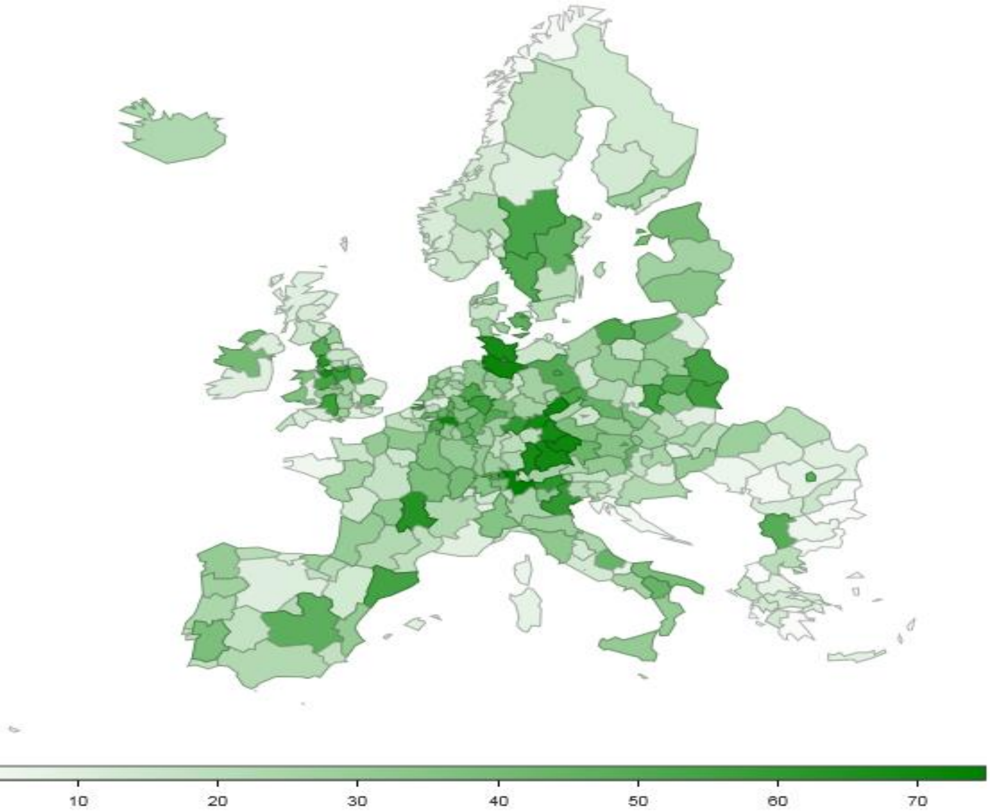
NACE	Name (2-digit)	Relatedness	Complexity
86	Human health activities	21	95 (3)
84	Public administration and defence; compulsory social security	20	61 (35)
41	Construction of buildings	19	50 (43)
56	Food and beverage service activities	19	59 (38)
96	Other personal service activities	18	73 (22)
88	Social work activities without accommodation	18	91 (5)
55	Accommodation	17	50 (44)
1	Crop and animal production, hunting and related service activities	17	0 (88)
85	Education	17	85 (10)
97	Activities of households as employers of domestic personnel	17	43 (53)

Appendix 5. List of CPC classes attached to each of the 7 key technologies

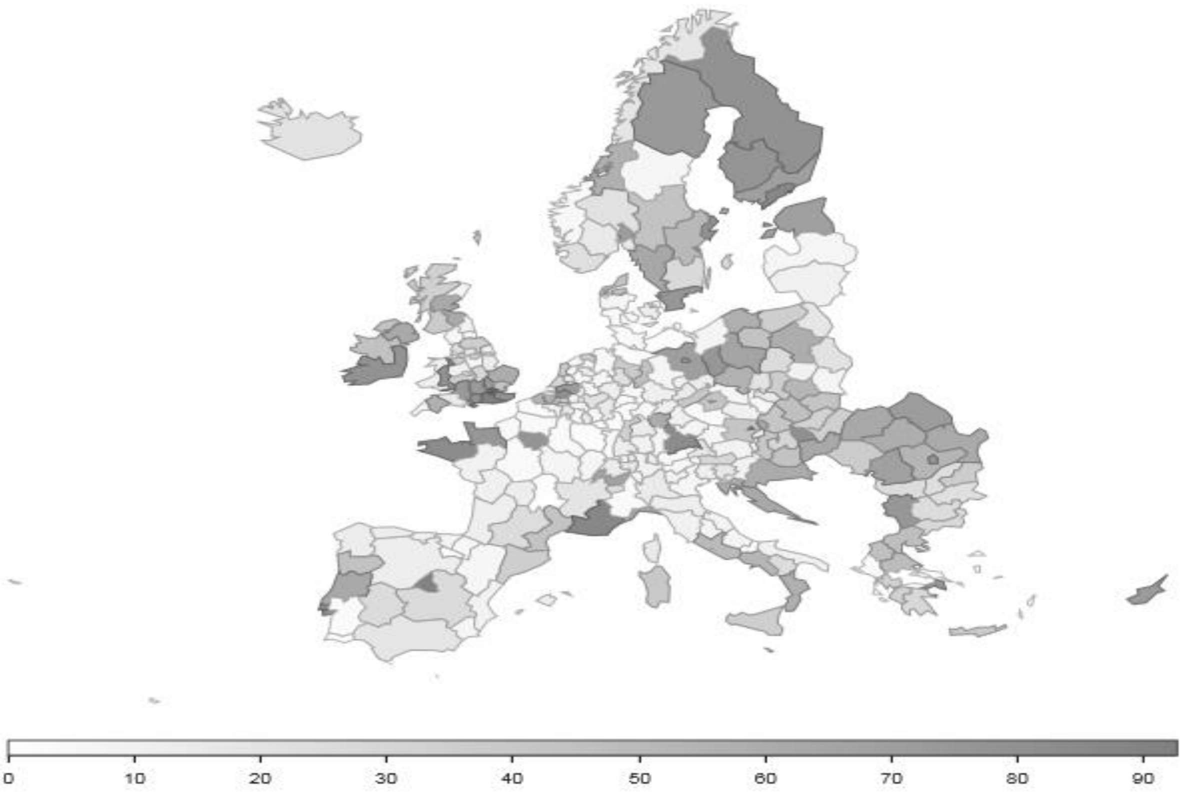
Key technology	Name	CPC	Description
Additive manufacturing	Additive manufacturing	Y02P10/29	Additive manufacturing [climate change mitigation technologies]
Additive manufacturing	Additive manufacturing	B29C64	Additive manufacturing by 3D printing, stereo-lithography or selective laser sintering
Additive manufacturing	Additive manufacturing	B33Y	Additive manufacturing, i.e. 3D printing
Batteries	Batteries	H01M	Processes Or Means, E.g. Batteries, For The Direct Conversion Of Chemical Energy Into Electrical Energy
Clean, connected, autonomous vehicles	Autonomous vehicles	G05D1	Control of position, course, altitude, or attitude of land, water, air, or space vehicles, e.g. automatic pilot
Clean, connected, autonomous vehicles	Autonomous vehicles	G01S17	Systems using the reflection or re-radiation of electromagnetic waves other than radio waves, e.g. lidar systems
Clean, connected, autonomous vehicles	Autonomous vehicles	H04W4/40	Wireless communication networks for vehicles, e.g. vehicle-to-pedestrians [V2P]
Clean, connected, autonomous vehicles	Autonomous vehicles	B60W30/14	Cruise control
Clean, connected, autonomous vehicles	Autonomous vehicles	B60T2201/08	Lane monitoring; Lane Keeping Systems
Cybersecurity	Cybersecurity	G06F21	Security arrangements for protecting computers or computer systems against unauthorised activity
Cybersecurity	Cybersecurity	H04W12	Network architectures or network communication protocols for wireless network security
Cybersecurity	Cybersecurity	H04L9	Cryptographic mechanisms or cryptographic arrangements for secret or secure communication
Cybersecurity	Cybersecurity	H04L63	Network architectures or network communication protocols for network security
Cybersecurity	Cybersecurity	Y04S40/24	Arrangements for network security, i.e. cybersecurity
High performance computing	Artificial intelligence	G06N3	Computer systems based on biological models
High performance computing	Artificial intelligence	G06N5	Computer systems utilizing knowledge-based models
High performance computing	Artificial intelligence	G06N7	Computer systems based on specific mathematical models
High performance computing	Artificial intelligence	G06N99	Subject matter not provided for in other groups of this subclass
High performance computing	Cloud computing	G06F9/5072	Grid computing
High performance computing	Quantum computers	B82Y10	Nanotechnology for information processing, storage or transmission, e.g. quantum computing or single electron logic
High performance computing	Quantum computers	H01L27/18	Devices consisting of a plurality of semiconductor or other solid-state components formed in or on a common substrate including components exhibiting superconductivity
High performance computing	Quantum computers	G06N10	Quantum computers, i.e. computer systems based on quantum-mechanical phenomena
High performance computing	Data technology	G06F	Electric digital data processing
High performance computing	Data technology	G06Q	Data processing systems or methods
High performance computing	Data technology	G06K	Recognition of data
High performance computing	Data technology	G06T	Image data processing or generation
Hydrogen technologies and systems	Hydrogen technology	Y02E60/3	Hydrogen technology
Hydrogen technologies and systems	Hydrogen technology	Y02E60/5	Fuel cells
Microelectronics	Semiconductors	H01L	Semiconductor devices; electric solid state device

APPENDIX 6. POTENTIALS OF EUROPEAN REGIONS IN KEY TECHNOLOGIES

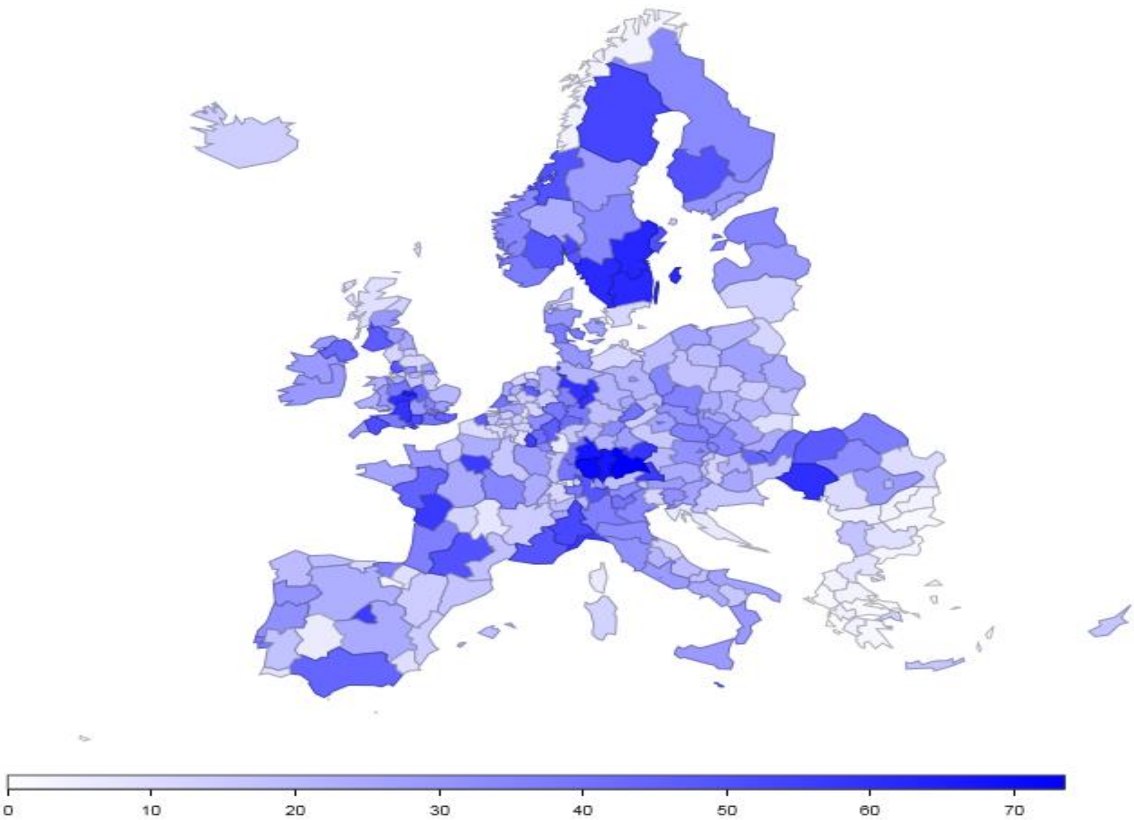
Map A1. Potentials of European regions in additive manufacturing technologies



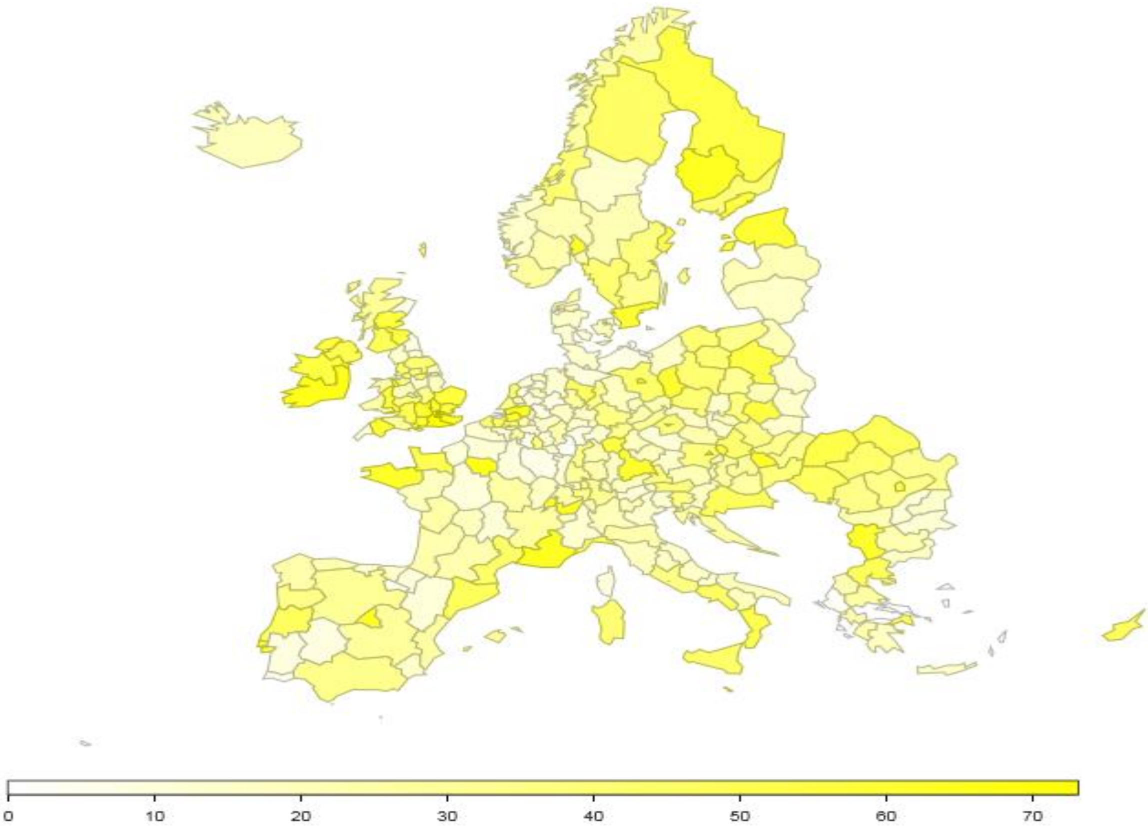
Map A2. Potentials of European regions in cyber-security technologies



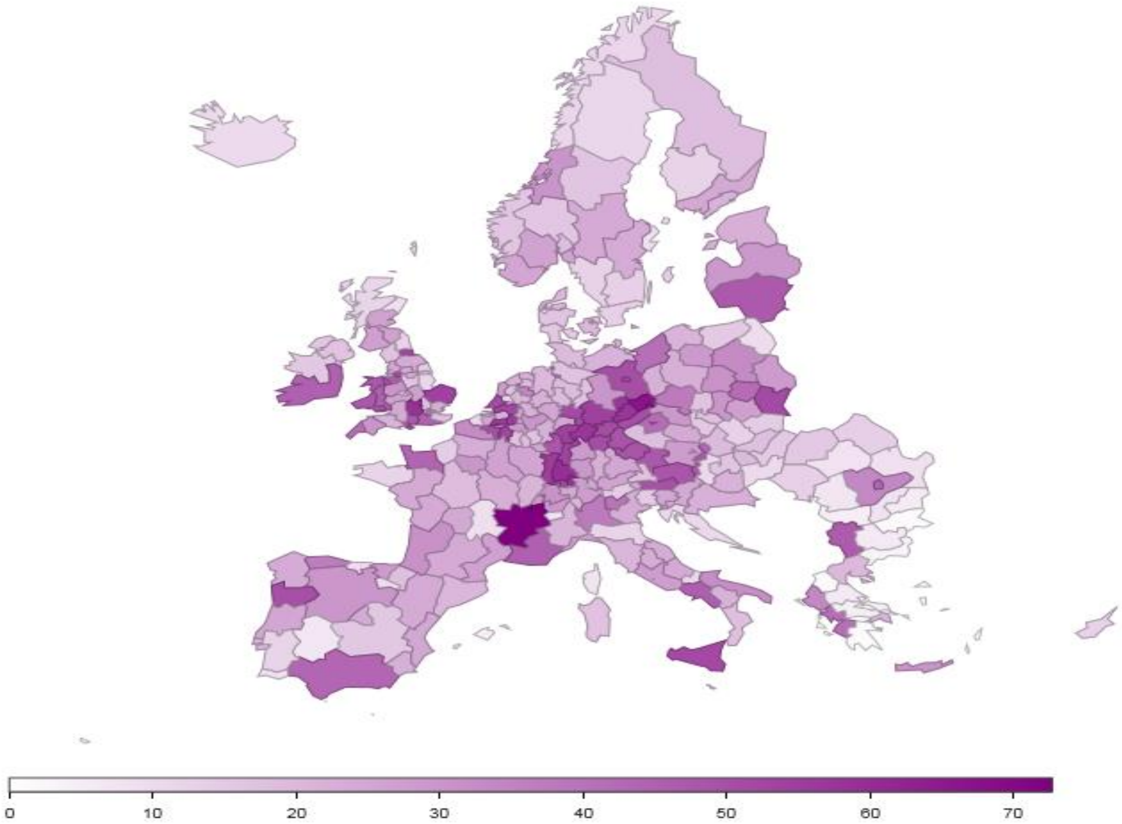
Map A3. Potentials of European regions in clean, connected and autonomous vehicles technologies



Map A4. Potentials of European regions in high performance computing technologies



Map A5. Potentials of European regions in micro-electronics technologies



REFERENCES

Balland, P.A., R. Boschma, J. Crespo and D. Rigby (2019). Smart specialization policy in the EU: Relatedness, knowledge complexity and regional diversification, *Regional Studies* 53(9), 1252-1268.

Balland, P.-A. and Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography 93*(1), 1–23.

Boschma, R. (2014). Constructing regional advantage and smart specialization: Comparisons of two European policy concepts, *Italian Journal of Regional Science* 13 (1), 51-68.

Boschma, R. (2017). Relatedness as driver behind regional diversification: a research agenda**,** *Regional Studies*, 51 (3), 351-364.

Boschma, R., P.A. Balland and D.F. Kogler (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in U.S. metropolitan areas from 1981 to 2010, *Industrial and Corporate Change* 24 (1), 223-250.

Boschma, R. and G. Capone (2016). Relatedness and Diversification in the European Union (EU-27) and European Neigbourhood Policy countries, *Environment and Planning C: Government and Policy* 34, 617-637.

Boschma, R., A. Minondo and M. Navarro (2013). The emergence of new industries at the regional level in Spain: a proximity approach based on product-relatedness, *Economic Geography* 89(1), 29–51.

Colombelli, A., J. Krafft and F. Quatraro (2014). The emergence of new technology-based sectors in European regions: a proximity-based analysis of nanotechnology, *Research Policy* 43, 1681-1696.

Coniglio, N. D., R. Lagravinese, D. Vurchio and M. Armenise (2018). The pattern of structural change: testing the product space framework. *Industrial and Corporate Change* 27 (4), 763-785.

Davies, B. and D.C. Mare (2019). Relatedness, complexity and local growth, Institute of Labor Economics DP, no. 12223.

Essleztbichler, J. (2015). Relatedness, industrial branching and technological cohesion in US metropolitan areas, *Regional Studies* 49 (5), 752–766.

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

Foray, D. (2015). *Smart Specialisation: Opportunities and Challenges for Regional Innovation Policy*. London: Routledge.

Guo, Q. and C. He (2017), Production Space and Regional Industrial Evolution in China. *Geojournal* 82, 379-396.

Hausmann, R. and C.A. Hidalgo (2011). The network structure of economic output, *Journal of Economic Growth* 16 (4), 309-342.

He, C., Y. Yan and D. Rigby (2018). Regional Industrial Evolution in China. *Papers in Regional Science* 97 (2), 173-198.

He, C. and S. Zhu (2019). *Evolutionary economic geography in China*, Springer.

Hidalgo, C., Balland, P.A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., and Zhu, S. (2018). The Principle of Relatedness, *Springer Proceedings in Complexity,* 451-457.

Hidalgo, C. and R. Hausmann (2009). The building blocks of economic complexity, *Proceedings of the National Academy of Sciences* 106: 10570-10575.

Hidalgo, C.A., Klinger, B., Barabasi, A.L. and Hausmann, R. (2007). The product space and its consequences for economic growth, *Science* 317, 482–487.

Kogler, D.F., D.L. Rigby and I. Tucker (2013). Mapping Knowledge Space and Technological Relatedness in US Cities, *European Planning Studies* 21(9), 1374-1391.

McCann P., Ortega-Argilés R. (2015). Smart Specialisation, Regional Growth and Applications to EU Cohesion Policy, *Regional Studies*, 49 (8), 1291-1302.

Mewes, L. and T. Broekel (2019) .Technological complexity and economic growth of regions, working paper, Utrecht University, Utrecht.

Montresor, S. and F. Quatraro (2017). Regional branching and key enabling technologies. Evidence from European patent data, *Economic Geography* 93(4), 367-396.

Morgan, K. (2015). Smart Specialisation: Opportunities and Challenges for Regional Innovation Policy, *Regional Studies* 49(3), 480-482.

Muneepeerakul, R., J. Lobo, S.T. Shutters, A. Gomez-Lievano and M.R. Qubbaj (2013). Urban economies and occupation space: can they get "there" from "here"?, *PLoS ONE* 8(9): e73676. doi: 10.1371/journal.pone.0073676.

Neffke F., Henning M. and Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions, *Economic Geography* 87(3), 237–265.

Rigby, D. (2015). Technological relatedness and knowledge space. Entry and exit of US cities from patent classes, *Regional Studies* 49 (11), 1922-1937.

Shutters, S., Muneepeerakul, R., and Lobo, J. (2016). Constrained pathways to a creative urban economy, *Urban Studies* 53(16), 3439-3454.

Tanner, A.N. (2016). The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases, *Journal of Economic Geography* 16 (3), 611-635.

Zhu, S., C. He and Y. Zhou (2017). How to Jump Further and Catch Up? Path-Breaking in an Uneven Industry Space. *Journal of Economic Geography*, doi:10.1093/jeg/lbw04.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

[Catalogue number]

Publications Office