SCIENCE, TECHNOLOGY & INNOVATION INDICATORS

Open indicators: innovation, participation and actor-based STI indicators
PARIS 2017

Technological diversification of regions and public R&D funding: Evidence from the EU Framework Programmes[1]

Wolf-Hendrik Uhlbach[*], Pierre-Alexandre Balland[*] and Thomas Scherngell[**]

[*] w.p.uhlbach@students.uu.nl; p.balland@uu.nl
Department of Economic Geography, Utrecht University, Heidelberglaan 2, Utrecht, 3584CS (the Netherlands)

[**] Thomas.Scherngell@ait.ac.at
Center for Innovation Systems & Policy, AIT Austrian Institute of Technology GmbH, Donau-City-Straße 1, Vienna, 1220 (Austria)

Abstract

Over the last years the issue of technological diversification gained importance for STI policies. This is especially true in the context of regions, as an important unit for STI policies. Much research was therefore dedicated to explore the drivers of diversification. An increasing body of evidence suggests that diversification is a highly path dependent process in which regions tend to diversify into technologies that to a large extent draw on knowledge and capabilities that are already present in the region. This process is referred to related diversification. From a policy perspective the question arises which factors influence the capability for technological diversification and in particular whether and how public research and development (R&D) subsidies can be a positive impetus. Making use of regional participations in the EU Framework Programmes (FP) from the EUPRO database, it will therefore ask to what extent subsidization of certain technologies will promote diversification. Secondly it will investigate to what extent subsidization can allow regions to diversify into less related technologies. After establishing a convergence between FP projects and technology fields of patents, we explore the relationship between diversification and public funded projects by means of a fixed-effects linear probability model. Results indicate statistically positive effects of participation in FP projects and a decrease in the deferring importance of relatedness with increasing number of participations. Despite their statistical significance, the marginal effects are small.

Introduction

Today it is widely agreed in STI studies that knowledge producing and innovating actors need to constantly re-adapt their innovation strategies in order to react to the increasing pace of technological change and innovative activity (see, e.g., Patel and Pavitt 1997). Moreover, knowledge production processes have become multidimensional, requiring the combination of

[1] This work has been supported with the provision of original data via the RISIS (Research Infrastructure for and Innovation Policy Studies) infrastructure (risis.eu), using information from the EUPRO database on European Framework Programme participations.

increasingly specialised and manifold pieces of knowledge in the innovation process (see, e.g., Granstand 1998, Cowan 2004). The latter is also referred to as increasing 'complexity' of innovation (see, e.g. Andriani 2011), which forces innovating actors to build up capabilities in cognitive and relational dimensions, in particular through networks and collaborative knowledge production (see, e.g., Scherngell 2013). In this context, Fai and Tunzelman (2001) and Leten et al. (2007) underline the important role of technological diversification, in particular for the absorptive capacity of firms, enabling them to adapt more efficiently and faster to new technological trends.

The issue of technological diversification has also been taken up recently by STI policies at different levels. At the European level, the smart specialisation strategies of the European Commission (see, e.g., EC 2014) emphasises the importance of promoting knowledge spillovers to foster diversification capabilities. Such policy concepts mostly address the regional level, acknowledging their crucial role as spatial agglomerations of innovative activity. It is considered as important for regions to diversify into new technologies and, by this, come into a position to develop new economic activities (Boschma & Frenken, 2011). This however is not equally feasible for all regions, as it depends on their idiosyncratic technological make-up. In recent literature the argument that regions tend to diversify into activities that are closely related to their incumbent ones therefore became prevalent (Hidalgo et al., 2007; Rigby, 2015; Boschma et al., 2015). Relatedness in this context refers to the extent to which different technologies and industries draw on the same knowledge and capabilities.

This argument is based on the assumption that technological diversification is a process of knowledge transfer (Boschma et al., 2013). Thus, diversifying into more related technologies is more feasible, as most of the necessary knowledge and capabilities are already available within a region (Cohen & Levinthal, 1990). Empirical studies widely prove the validity and predictive power of this argument. Using export data on the country level, Hidalgo et al. (2007) find that countries tend to diversify into products closely related to their existing portfolio. Boschma et al. (2013) show in a study on the Spanish context, that this effect is even stronger on the regional level. As knowledge dynamics are more directly captured by patent data, a row of studies (see e.g. Ribgy, 2015; Essletzbichler, 2015; Boschma et al. 2015; Kogler et al., 2016) confirmed these results also for the technological diversification of European regions and US cities.

Thus, from a policy perspective the question arises which factors influence the capability for technological diversification, in particular whether and how public research and development (R&D) subsidies can be a positive impetus. Only a few studies identify factors beyond existing productive and technological capabilities (Boschma, 2017), such as Cortinovis et al. (2016) or Boschma and Capone (2015), who confirm the role of institutions and the form of capitalism in the diversification process. However, effects of public R&D subsidies are not systematically addressed in empirical works up to now. The rationale for subsidizing R&D activities and prioritizing certain fields is to actively shape the intensity and direction of inventive activity and, as a result, the technological diversification process.

Against this backdrop, the objective of this paper is to shed some light on the effects of collaborative R&D subsides on the diversification of European regions. It will more specifically ask to what extent subsidization of a certain technology will lead to the development of a comparative advantage in the respective technology. Secondly it will investigate to what extent subsidization can allow regions to diversify into less related technologies. We measure technological diversification, a region's changing patent portfolio,

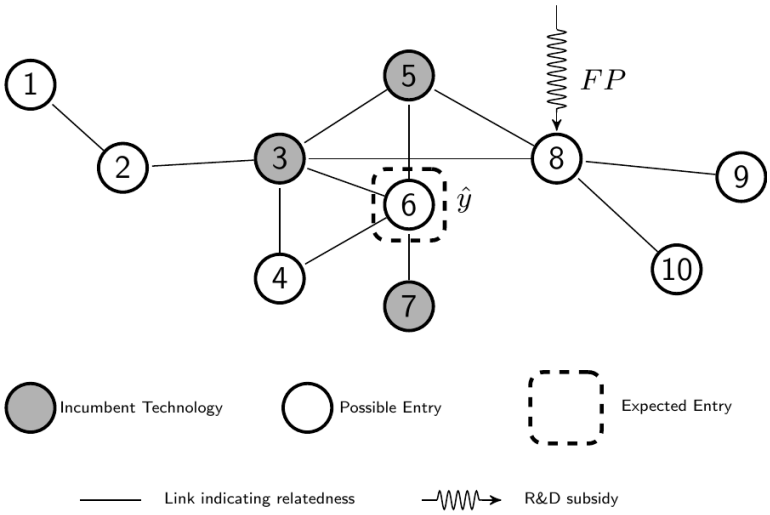
across technological domains in patents. R&D subsidies are measured by project participations in the European Framework Programme (FP). We explore the relationship between diversification and public subsidization by means of a fixed-effects linear probability model, controlling determinants identified by previous works, in particular the degree of existing technological relatedness of a region. The results are of high relevance for policy through estimating effects of past efforts and revealing opportunities on how funds may be allocated more efficiently in terms of technological diversification capabilities. Further the study points out interesting ways in the context of producing STI indicators by matching different R&D datasets across differing classification schemes for technological domains.

Background and hypotheses

Traditionally, innovation policy aims at fixing market and system failures, mainly by subsidizing R&D activities in order to make up for the appropriability problem or by facilitating knowledge transfers, by promoting R&D collaborations (Aghion et al., 2009). The first question that arises in this context is whether such subsidies can trigger regional innovative activities. Theoretically there are good arguments to believe so, because they give access to extra regional knowledge resources, which can also lead to intra-regional knowledge spillover (Wanzenböck et al., 2014). Hazir et al. (2016) confirm this for the ICT field and show that regions' patenting activity is also positively affected by knowledge flows, especially when partners are proximate. Broekel (2015), on the other hand, evaluates the effects of a German collaborative R&D program. His findings indicate no direct effects arising from the amount of funding. But he also contends that taking the number of projects is less biased, as the amount of subsidization can misestimate the actual project costs (Broekel, 2015). The following will therefore be hypothesized:

H1: *The participation intensity of a region in FP projects positively influences its technological diversification.*

Figure 1: Illustration of Research Questions



Notes: The figure depicts a technology space of a region in which each node represents a technology and the edges indicate whether two technologies are related to each other. In this example the region has a revealed technological advantage (RTA) in technologies 3, 5 and 7. The standard model would predict the entry of technology 6 as it is the most related to the regional portfolio (relatedness density (RD) = 75). Technology 8 however is less related (RD = 50) but the regions R&D activities in this field are supported by public funding.

Further, not all kinds of diversification are qualitatively the same. As the analogy of the knowledge space reveals, some technologies are easier to reach, as the region already has related knowledge and know-how available, while for others it lacks many. This case is commonly referred to as *unrelated diversification*. This kind of diversification is rather an exception than the rule (Boschma, 2017). For policy makers on the other hand, this will be the more interesting case, as it comes less natural and is connected to a series of additional system failures, so called transformational system failures (Weber & Rohracher, 2012). Further it is important for regions to diversify into unrelated activities every once in a while, to avoid lock-ins. The role of R&D subsidies in this context is more ambiguous. As Frenken (2017) lays out, generic R&D subsidization strategies will lead to more related diversification, whereas clearly directed and non-neutral subsidization has the potential to lead to unrelated diversification. Directedness in this context can be understood as demand-led. Latter is regarded as an important characteristic, as it actively steers research efforts towards particular fields and away from others. This idea is depicted by figure 1, in which technology 6 is closest related to the region's portfolio, but technology 8 gets subsidized. The question will therefore be to what extent research efforts are directed towards field 8 and how this will change the entry probabilities of the respective technologies. As policy on EU level, even when called technology-neutral has been technology-specific in their implementation (Azar & Sanden, 2011), it will be hypothesized:

H2: *The participation intensity of a region in FP projects aiming at a certain technology, decreases the deferring influence of relatedness on a region's capability for technological diversification*

Data and Methods

The empirical analysis is based on two different data sources. Firstly, the patent data set from OECD REGPAT was used to map the technological diversification of European regions. Secondly, the data on FP participations was extracted from the EUPRO data base. The EUPRO data base contains information about funded projects and their participants on the NUTS2 level and covers all FPs, i.e. 1984-2013. Yet, the information on the latest available programme, FP7, is not completely in the data base yet, and reliable information is available until 2012 (Heller-Schuh et al., 2015). The study area covers collaborative research projects from FP5 to FP7. Since not all subprogram areas from the FPs are relevant for technological diversification, only projects from the *cooperation* programme from FP7 (excluding "Socio-economic Sciences and Humanities) were selected. Relevant sub-programmes from FP5 and FP6 were selected based on the concordance table provided by Rietschel (2010). This resulted in a total of 15,983 projects covering the years 1999-2011.

Data Preparation

The big challenge for the empirical analysis in this study is to combine two data sets that draw on fundamentally different classification logics; in essence, FP projects have to be assigned to technology domains appearing in patents. This is not feasible based on the existing classifications in the two datasets. Information of projects recorded in 'EUPRO', however, offer textual descriptions, such as project titles, project abstracts, objectives and achievements as well as a list of resulting documents and their titles. Therefore, the initial idea has been to use this information for establishing a convergence based on text classification procedures. However, the amount and quality of the textual information available is not sufficient to

delineate more than 600 IPC classes in a reliable manner. Therefore, the classification scheme used instead is the scheme of technology fields (TF) as proposed by Schmoch (2008). It summarizes the IPC classes to 35 distinct and balanced technology fields[2]. For the matching purposes this suits well, as it allows for more accurate predictions.

In order to link the FP projects to the technology fields, a machine learning technique, namely a maximum entropy classifier, as provided by the "maxent" package in R (Jurka, 2012) is applied. In order to be able to make predictions, this technique requires a large number of project descriptions that can be used as training documents. To achieve this, out of the 15,983 selected projects a random sample is retrieved and uniquely classifiable projects are assigned to one TF manually. To improve the assignment Google Patent Search was used to look up the technological affiliation of certain terms. In so doing, 1,001 projects were assigned manually. After pre-processing the textual information, which entails removing stop words and non-alphanumerical symbols as well as weighting terms, the classifier is fitted and applied to the remaining unclassified projects. As a result, a project-technology matrix is returned, where the elements indicate the probability with which a project belongs to a certain TF. As FP projects are also generic and can address different technological fields, for each project a vector of up to five TFs with the highest probability and the corresponding probabilities is retrieved. Based on this, projects are assigned fractionally to the TFs[3]. In order to reduce the noise and wrong classifications, classifications with less than 5% probability are not assigned. To evaluate the classification quality, two approaches were taken. Firstly, the training dataset was split and 101 classified abstracts were reclassified. Secondly, an inventory of 295 patents, resulting from 161 FP projects from the FP7 ICT programme was used (Jacob et al., 2015). These tests reveal, that between 51% (for ICT projects) and 63% (project abstracts) of the fields are predicted correctly[4]. Overall, regarding the amount of available information, and the number of classes, as well as the bias of the validation data, towards ICT sector, these results will be considered as sufficient.

Empirical Approach and Model

In order to estimate the respective effects of subsidization on the technological diversification of European regions, we employ a fixed effects panel regression approach. In so doing 282 NUTS2 regions (EU27, Switzerland, Norway and Iceland) for two four year periods (2003-2006; 2007-2010) are used, in which the independent variables were lagged by one period (1999-2002; 2003-2006). In this approach however, relatedness and entry were computed on the level of 613 IPC classes, while funding related variables where computed on the level on 33 technology fields[5] (Schmoch, 2008).

Given our empirical setting, we employ – as in previous related works (see e.g. Boschma et al., 2015) – a fixed effects linear probability panel regression approach. This allows testing for omitted variables on the regional and technology level that would otherwise be neglected. The model specification in scalar notation is as follows:

[2] The 35 original fields were reduced to 33. Fields 10 and 11 as well as fields 14 and 16 were merged.
[3] An example to illustrate: Assume that a project was assigned to TF 6 with a probability of 0.70, to TF 7 with 0.10, to TF 8 with 0.05 and to TF 4 with 0.02, etc. This would lead to a fractionalized assignment to TF 6 with a share of 0.80, to TF 7 with a share of 0.11, to TF 8 with a share of 0.06 and to TF 4 with a share of 0.03. An overview over the result of the assignment can be found in the following figures.
[4] A more detailed classification evaluation can be found in the appendix
[5] This approach may result in methodological shortcomings, such as the multilevel problem, which have to be dealt with in a further progression of the research.

$$ENTRY_{i,r,t} = \beta_1 FP_{i,r,t-1} + \beta_2 RD_{i,r,t-1} + \beta_3 FP_{i,r,t-1} * RD_{i,r,t-1} + \theta_i + \varphi_r + \omega_t + \varepsilon_{i,r,t}$$

In this model, the dependent variable $ENTRY$ is a binary variable taking the value 1 if a region r gains a specialisation (RTA[6]) in a certain technology i, at time t, it was not specialised in at period $t-1$ and the value 0 otherwise. FP is a proxy for subsidisation intensity:

$$FP_{i,r,t} = \sum_{p \in i} \frac{x_{p,r} * w_{p,i}}{k_{p,t}}$$

It is measured by the number of participations x of a region r in a project p weighted by the field weight w and divided by the number of periods it covers k. The variable RD denotes the relatedness density around a certain technology. This measure is widely used and describes how close a technology is to the existing technological portfolio of a region. In order to compute it, it is first necessary to construct a so called technology or knowledge space (see Hidalgo et al., 2007). This network representation indicates which technologies are related to each other. For its calculation the method of Rigby (2015) is applied. In this methodology two technology classes are thought to be related when they are assigned to the same patents, above a certain threshold. Thus, in a first step for each period a co-occurrence matrix between technologies i and j, in which $i = 1, \dots, m$ and $j = 1, \dots, m$ is constructed. This results in a $m \times m$ matrix with the dimensions 613×613 IPC classes. To further get the degree of relatedness between each pair of technologies, this matrix normalized, using the association measure, by Eck and Waltman (2009) in the version of Steijn (forthcoming). In a final step, the matrix is dichotomized, such that value 1 indicates that two technologies are related, whereas value 0 indicates no such relatedness. This results in a complex network, in which some technologies are related to many others, while others are only related to a few (Frenken 2017). Based on this, the relatedness density is computed as follows, where σ indicates whether technologies i and j are related to each other:

$$RD_{i,r,t} = \frac{\sum_{j \in r, j \neq i} \sigma_{ij}}{\sum_{j \neq i} \sigma_{ij}} \times 100$$

The values of this variable range between 0 and 100, where a value of 0 indicates, that no related technologies are existing in a region and a value of 100 indicates, that all related technologies around technology i, are incumbent in region r. Next to FP and RD, an interaction effect $FP*RD$ between the participations and relatedness is included to test the second hypothesis. This will inform about the conditional importance of relatedness at different levels of subsidisation and vice versa. In a final step, observations that already have an RTA in period $t-1$ were deleted from the panel. Descriptive statistics are presented in Table 1.

[6] The RTA (Revealed technological advantage) is a measure indicating the relative specialisation of a region in a technonology. Analogously to the revealed comparative advantage (RCA) it is calculated as follows. $RTA_i = 1$ if $\frac{patents_{i,r,t}/\sum_i patents_{i,r,t}}{\sum_r patents_{i,r,t}/\sum_r \sum_i patents_{i,r,t}} > 1$

Table 1: Descriptive Statistics and Correlation Matrix

	N	*Mean*	*Min*	*Max*	*SD*	*(1)*	*(2)*
ENTRY(1)	284,508	0.11	0	1	0.31		
FP (2)	345,732	2.87	0	485	11.47	0.03	
RD (3)	345,732	17.79	0	100	12.94	0.14	0.14

Results

Table 2 presents the results for our model as presented in the previous section. We estimate different model versions, adding variables step-by-step to check robustness and highlight changes when another variable is added.

Table 2: Linear Probability Model; entry and relatedness variables on IPC and funding variables on 35 TF level

ENTRY	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Full Model**
Constant	0.1080 *** (0.0005850)	0.1115 *** (0.0006005)	0.1118 *** (0.0006020)	0.1150 *** (0.0006373)	
Log(FP)	0.0242 *** (0.0007309)		0.0118 *** (0.0007677)	0.0159 *** (0.0007625)	0.0167 *** (0.0017133)
RD		0.0033 *** (0.0000434)	0.0031 *** (0.0000458)	0.0030 *** (0.0000466)	0.0003 *** (0.0000804)
Log(FP)×RD				− 0.0010 *** (0.0000529)	− 0.0004 *** (0.0000565)
Time F.E.	No	No	No	No	Yes
Region F.E.	No	No	No	No	Yes
Technology F.E.	No	No	No	No	Yes
N	284,508	284,508	284,508	284,508	284,508
Adj. R^2	0.004661	0.0197	0.02073	0.02196	0.06323

Notes: The independent variables in models 1 to 4 are mean centred. Heteroscedasticity-robust standard errors are reported in parentheses. All variables are statistically significant at the *** $\boldsymbol{p < 0.01}$ level.

As model 1 reveals, funding indeed seems to have a positive and significant impact on the entry of a new technology, which confirms hypothesis 1. When looking at the marginal effect however, it becomes obvious that the impact is small. Thus, a 10% increase of project participations will only be associated with a 2% increase of the mean probability of a technology to enter the region's portfolio. The effect of relatedness density is as expected, significantly positive (model 2). Also the scope of its effect is within the range of previous findings (e.g. Boschma et al., 2015). According to model 2, a 10% increase in relatedness density would lead to a 29% increase in the mean entry probability. Model 3 shows that also together, the effects remain the same. Model 4 adds the interaction effect, which turns out to be negative and significant. Thus, in line with hypothesis 2, this would suggest that higher

R&D subsidization can to some extent compensate for a lack of relatedness. These results also stay robust after adding time, technology and regional fixed effects[7].

This will have important implications for policy makers, as it suggests that by subsidizing certain technologies, they will also allow diversification into less related technologies. But, this does not mean, that R&D subsidization will make relatedness completely unimportant. While the coefficient is small an ad hoc interpretation is less intuitive. This may be related to the linearity assumption in our model not sufficiently approximating the relationship between the dependent and the independent variables. An intuitive approach in interpretive terms is to discriminate effects between different groups of observations in terms of their magnitude. We do so for the models presented in Table 3. In the Model 6, FP is held continuous, while RD is transformed into a dummy variable in which 1 denotes a relatedness density higher than the mean. In Model 7, FP was analogously transformed into a dummy variable.

Table 3: Linear probability models discriminating effects across groups in terms of their relatedness and FP participation intensity

ENTRY	Model 6	Model 7
Constant	0.0759 *** (0.0007539)	0.1057 *** (0.0007479)
FP	0.0290 *** (0.0011297)	0.0263 *** (0.0013850)
RD	0.0753 *** (0.0012592)	0.0036 *** (0.0000531)
FP×RD	− 0.0241 *** (0.0015099)	− 0.0017 *** (0.0001038)
N	284,508	284,508
Adj. R^2	0.02038	0.02178

Notes: In model 6 FP is continuous while RD is a dummy variable. In model 7 RD is continuous, while FP is a dummy variable. The continuous independent variables are mean centred. Heteroscedasticity-robust standard errors are reported in parentheses. All variables are statistically significant at the *** $\boldsymbol{p < 0.01}$ level.

Now the results can be interpreted more straightforwardly. In model 6 the interaction term represents the difference of the slope coefficient for the FP variable for observations with high and low relatedness density. Thus, for observations with a high relatedness density the coefficient would decrease by 0.0241, resulting in a slope coefficient of 0.0049. While in this specification a 10% increase in participations for observations with a low relatedness density is associated with a 3.6% increase in the mean entry probability, for observations with a high relatedness density this would only result in a 0.6% increase of the entry probability. Model 7 compares observations with low and high numbers of participations. For observations with low numbers of participations a 10% increase of RD would be associated with an increase of the entry probability by 34%. For the group with high number of participations the same increase of RD would result in only an 18% increase of the mean entry probability.

[7] The model has also been estimated with all variables on the level of technology fields. This however led to different results. The FP variable turned out not to be significant until regional and industry fixed effects were added.

This crude additional analysis added more understanding of the effects. They show that if technologies in regions are funded that are less related to the region's portfolio the additional effect of funding will be bigger than the funding of technologies for which the regions already has many related capabilities. For highly funded technologies on the other hand, the effect of relatedness decreases. For further interpretations it is important to have in mind that high and low values in this case always refer to values above and below average. In further analysis it will therefore be important to add more nuance to this analysis.

Discussion and Conclusion

The aim of this research was to assess to what extent public R&D subsidization contributes to the technological diversification of European regions. In so doing it is novel in two ways. Firstly, it establishes a convergence of funding data, on project level and secondly it investigates the effects of subsidization policies on the European level. Overall, the results confirm our two hypotheses. Thus, we find that a region's participation intensity in FP projects indeed increases its likelihood to diversify. However, the marginal effect of this is small. In line with our second hypothesis, we find that a high participation intensity can to some extent compensate for a lack of relatedness.

These results, however, should be interpreted with caution as the study is not free of limitations. First and foremost, the established convergence introduces a set of problems. Due to the imperfection of the text classification, the subsidization of some technology fields may be underestimated, while on the other hand noise can be introduced, when wrong technology fields are predicted. Secondly, only a basic model was estimated, which is trying to capture a complex relationship. A key problem here can be seen in an allocation bias of the funds. This means that funds are a priori allocated in regions that already have an advanced R&D sector, and that are therefore already more capable of diversifying into less related technologies. Finally, by looking at a collaborative funding scheme, more effects can be thought to arise from the collaborations. It would therefore be adequate to also take into account spatial effects, as for instance done by Hazir et al. (2016).

Before useful policy implications can be deducted about the effects of R&D subsidization, more research is needed. It is important to understand in which regions subsidization delivers most additional benefits. Further, it is important to shed more light on the extent of unrelated diversification it actually enables. It should therefore be investigated, whether subsidization can enable the entry of technologies when there are no related technologies incumbent in the region, or whether its effect strongest at medium levels of relatedness density.

Acknowledgements
The authors would also like to thank, David Rigby and Tom Broekel for their insightful comments.

References
Aghion, P., David, P. A., & Foray, D. (2009). Science, technology and innovation for economic growth: linking policy research and practice in STIG Systems. *Research Policy, 38(4), 681–693.*

Andriani, P. (2011). *Complexity and innovation*. The SAGE handbook of complexity and management, 454-470

Azar, C., & Sandn, B. A. (2011). The elusive quest for technology-neutral policies. *Environmental Innovation and Societal Transitions, 1(1), 135 - 139.*

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

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

Boschma, R., & Capone, G. (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy, 44(10), 1902–1914.*

Boschma, R., & Frenken, K. (2011). The emerging empirics of evolutionary economic geography. *Journal of Economic Geography, 11(2), 295–307.*

Boschma, R., Minondo, A., & Navarro, M. (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.*

Broekel, T. (2015). Do cooperative research and development (R&D) subsidies stimulate regional innovation efficiency? Evidence from Germany. *Regional Studies, 49(7), 1087–1110.*

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly, 128–152.*

Cortinovis, N., Xiao, J., Boschma, R., & van Oort, F. (2016). Quality of government and social capital as drivers of regional diversification in Europe (*Papers in Evolutionary Economic Geography No. 16.10)*. Utrecht University, Section of Economic Geography.

Cowan, R. (2004). Network Models of Innovation and Diffusion. *Research Memorandum, MERIT, 10.*

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

European Commission (2014) National/Regional innovation strategies for smart specialization (RIS3), http://ec.europa.eu/regional_policy/sources/docgener/informat/2014/smart_specialisation_en.pdf

Fai, F., Von Tunzelmann, N., 2001. Industry-specific competencies and converging technological systems: evidence from patents. *Structural change and economic dynamics 12, 141-170.*

Frenken, K. (2017). A complexity-theoretic perspective on innovation policy. *Complexity, Governance & Networks(1), 35–47.*

Granstand O (1998) Towards a theory of the technology-based Firm, *Research Policy 27, 465-489*

Hazir, C. S., LeSage, J., & Autant-Bernard, C. (2016). The role of R&D collaboration networks on regional knowledge creation: Evidence from information and communication technologies. *Papers in Regional Science.*

Heller-Schuh, B., Barber, M., Züger, M., & Scherngell, T. (2015*). Report on the content and technical structure of EUPRO v7.4.11* (task 1 of wp6).AIT Austrian Institute of Technology.

Hidalgo, C., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317(5837), 482–487.*

Jacob, J., Sanditov, B., Smirnov, E., Wintjes, R., Surpatean, A., Notten, A., & Sasso, S. (2015, 2). Inventory of the patents and publications of FP7 projects in the field of ICT. European Commission.

Jurka, T. P. (2012). Maxent: An r package for low-memory multinomial logistic regression with support for semi-automated text classification. *The R Journal, 4(1), 56–59***.**

Kogler, D. F., Essletzbichler, J., & Rigby, D. L. (2017). The evolution of specialization in the EU15 knowledge space. *Journal of Economic Geography, 17(2), 345*.

Leten, B., Belderbos, R. and Van Looy, B. (2007), Technological Diversification, Coherence, and Performance of Firms. *Journal of Product Innovation Management, 24: 567–579.*

Neffke, F., Henning, M., & 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.*

Patel, P., Pavitt, K., 1997. The Technological Competencies of the World's Largest Firms: Complex and Path-dependent, but Not Much Variety. *Research Policy 26, 141-156.*

Rietschel, E. T. (2010). Evaluation of the sixth framework programmes for research and technological development 2002-2006: Report of the expert group. EUR-OP.

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

Scherngell T. (ed) (2013) *The Geography of Networks and R&D Collaborations.* Springer-Physica Verlag, Berlin-Heidelberg-New York

Wanzenböck, I., Scherngell, T., & Brenner, T. (2014). Embeddedness of regions in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications. *The Annals of Regional Science, 53(2), 337-368.*

Weber, K. M., & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive failures framework. *Research Policy, 41(6), 1037 - 1047.*

Appendix

Appendix 2: Classification Statistics

	Precision	Recall	F1 Score	n
Avg. ICT Patents	0.32	0.51	0.37	161
Avg. Abstracts	0.44	0.63	0.50	101

Notes: This table shows the evaluation of the classification results. The first row presents the results, based on an external inventory of 295 patents resulting from ICT projects (Jacob et al., 2015). In the second row 101 manually classified project abstracts were used as test data. In order to control the quality of multiclass multilabel classifications, three measures are commonly employed. Precision, which is defined as the number of correct predictions divided by the number of all predictions, recall, which is defined as the ratio of correct predictions, divided by the number of predictions that should have been returned and an overall measure that combines the previous two measures, by taking the harmonic mean, the F1 score.