INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

CERS-IE WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

Repeated collaboration of inventors across European regions

GERGŐ TÓTH – SÁNDOR JUHÁSZ – ZOLTÁN ELEKES– BALÁZS LENGYEL

CERS-IE WP – 2021/17
March 2021

https://www.mtakti.hu/wp-content/uploads/2021/03/CERSIEWP202117.pdf

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
BUDAPEST, 2021

ABSTRACT

This paper explores the spatial patterns and underlying determinants of repeated inventor collaboration across European NUTS 3 regions. It is found that only a small fraction of co-inventor linkages across regions are repeated, while community detection reveals that these collaborations are clustered in geographical space more intensively compared with collaboration in general. Additional results from gravity modelling indicate that links in the inter-regional co-patenting network emerge mainly through the triadic collaboration of regions, while geographical proximity becomes the most influential factor for repeating co-inventor ties. In addition to that, the combination of technological similarity and shared third partner regions offer a premium for the likelihood of repeating collaboration, but only when geographical proximity is present as an enabler.

JEL codes: D85, O31, O43, O52, R11, R58

Keywords: collaborative knowledge production; inter-regional collaboration; co-inventor network; repeated collaboration; European Research Area; gravity model

Gergő Tóth
Agglomeration and Social Networks Lendület Research Group, Centre for Economic- and Regional Studies, Budapest, Hungary
and
Spatial Dynamics Lab, University College Dublin, Dublin, Ireland
Email: toth.gergo@krtk.hu

Zoltán Elekes
Agglomeration and Social Networks Research Group, Centre for Economic and Regional Studies
and
Centre for Regional Science at Umeå University (CERUM), Umeå University, 90187 Umeå, Sweden
Email: elekes.zoltan@krtk.hu

Sándor Juhász
NETI Lab, Corvinus Institute for Advanced Studies, Budapest Corvinus University, Budapest, Hungary
Email: sandor.juhasz@uni-corvinus.hu

Balázs Lengyel
Agglomeration and Social Networks Lendület Research Group, Centre for Economic- and Regional Studies, Budapest, Hungary;
International Business School Budapest, Budapest, Hungary
and
NETI Lab, Corvinus Institute for Advanced Studies, Budapest Corvinus University, Budapest, Hungary
Email: lengyel.balazs@krtk.hu

A feltalálók ismételt együttműködése az európai régiókban

TÓTH GERGŐ – JUHÁSZ SÁNDOR – ELEKES ZOLTÁN – LENGYEL BALÁZS

ÖSSZEFOGLALÓ

A tanulmány a feltalálók ismételt együttműködésének térbeli mintáit és mögöttes meghatározó elemeit vizsgálja az európai NUTS 3 régiókban. A tanulmány rámutat, hogy a régiók közötti feltalálói kapcsolatoknak csak kis hányada ismétlődik meg, míg a közösség-keresési algoritmussal feltárt mintázatok azt mutatják, hogy ezek az együttműködések intenzívebben csoportosulnak a földrajzi térben, mint az újonnan létrejövő kapcsolatok. A gravitációs modellek további eredményei azt mutatják, hogy a régiók közötti szabadalmazási hálózatban a kapcsolatok főleg a régiók triádikus együttműködésén keresztül jönnek létre, míg a földrajzi közelség válik a legbefolyásosabb tényezővé a régiók közötti kollaboráció megismétlésében. Ezek mellett a technológiai hasonlóság és a közös harmadik féllel való együttműködés ugyan növeli az együttműködés megismétlődésének valószínűségét, de csak akkor, ha a földrajzi közelség mindezt lehetővé teszi.

JEL: D85, O31, O43, O52, R11, R58

Kulcsszavak: Kollaboráltív tudástermelésé; régiók közötti együttműködés; feltalálói hálózat; ismétlődő együttműködés; Európai kutatási térség; Gravitációs modell

Repeated collaboration of inventors across European regions

Gergő Tóth[1,2], Sándor Juhász[3], Zoltán Elekes[2,4,*], and Balázs Lengyel[2,3,5]

[1]Spatial Dynamics Lab, University College Dublin, Belfield, Dublin 4, Ireland

[2]Agglomeration and Social Networks Lendület Research Group, Centre for Economic and Regional Studies, 1097 Budapest, Hungary

[3]Laboratory for Networks, Technology & Innovation, Corvinus University of Budapest, 1093 Budapest, Hungary

[4]Centre for Regional Science at Umeå University (CERUM), Umeå University, 90187 Umeå, Sweden

[5]International Business School Budapest, 1037 Budapest, Hungary

[*]**Corresponding author:** elekes.zoltan@krtk.hu

Abstract:

This paper explores the spatial patterns and underlying determinants of repeated inventor collaboration across European NUTS 3 regions. It is found that only a small fraction of co-inventor linkages across regions are repeated, while community detection reveals that these collaborations are clustered in geographical space more intensively compared with collaboration in general. Additional results from gravity modelling indicate that links in the inter-regional co-patenting network emerge mainly through the triadic collaboration of regions, while geographical proximity becomes the most influential factor for repeating co-inventor ties. In addition to that, the combination of technological similarity and shared third partner regions offer a premium for the likelihood of repeating collaboration, but only when geographical proximity is present as an enabler.

Keywords: collaborative knowledge production; inter-regional collaboration; co-inventor network; repeated collaboration; European Research Area; gravity model

1. Introduction

In order to enhance the competitiveness of economic actors across the European Union, considerable effort has been pledged to boosting innovative capacity through research and development (R&D) investments (European Commission, 2010). However, the most promising innovative actors tend to be geographically concentrated in already more successful locations (Usai, 2011). EU15 countries outperform the later-joined EU13 countries in terms of research and innovation (R&I) activities (Delanghe et al., 2009; Quaglio et al., 2020), all the while within-country differences in innovative capacity are also substantial (Crescenzi et al., 2007). Efforts to increase public R&D expenditures were set back in the aftermath of the 2008 crisis (Rodríguez-Pose, 2020), while the impact of the current pandemic is likely to be considerable as well. Additionally, regions across Europe have demonstrated disparities in the ability to translate innovative activity to economic growth (Rodríguez-Pose, 2020). These processes amount to a lasting innovation divide amongst EU regions.

Integration and the promotion of knowledge sharing collaboration across regions is central to the strategy of the EU to alleviate this innovation divide (European Commission, 2020). Collaboration is expected to improve overall R&I performance, as knowledge production and innovation increasingly relies on combining the knowledge of multiple actors (Jones, 2009). Additionally, collaborative R&I activity holds the potential to connect lagging regions and territories of the EU13 countries more tightly to the core of European R&I activities (Lengyel & Leskó, 2016), to facilitate knowledge sharing across regions (Hoekman et al., 2013; Broekel, 2015; De Noni et al., 2018), and ultimately to decrease regional differences in R&I capacity. Hence, it is an explicit aim of initiatives like the European Research Area (ERA) to improve R&I collaboration between EU regions (Frenken et al., 2007; Council of the European Union, 2009; Harrap & Doussineau, 2017).

To explore potential obstacles and opportunities for inter-regional R&I related cooperation, scholars from different fields have studied the spatial networks of knowledge sharing and collaboration in the EU (e.g., Breschi & Cusmano, 2004; Roediger-Schluga & Barber, 2008; Maggioni et al., 2007; Maggioni & Uberti, 2009; Scherngell & Lata, 2013). This stream of studies reveals fragmentation for multiple forms of R&I-related collaboration between territories. Co-inventor ties of patent collaboration form spatial communities that resemble national borders or include groups of regions in neighbouring countries (Chessa et al., 2013). Under the EU

Framework Programme (FP), collaborative R&D activities between member states have become more integrated during the period of 2003-2007, but the EU13 countries are still dominantly engaged in technologies of lower complexity (Balland et al., 2019). Finally, while collaboration on scientific publications is less determined by territorial borders due to EU integration efforts, cooperation is still predominant between geographically proximate actors (Hoekman et al., 2010), and more complex scientific knowledge tends to cluster in regions in the North and West of Europe (Heimeriks et al., 2019).

However, to fully appreciate the relational structure of inventive collaboration across European regions, repeated collaborations also need to be explored. That is, we need to better understand how the repeating of collaboration to produce knowledge shapes cohesion in the ERA. Repeated collaboration may enhance the effectiveness of collaborative agreements (Zollo et al., 2002), and the performance of partners (Goerzen, 2007), and may facilitate the exchange of more complex knowledge (Reagans & McEvily, 2003; Sorenson et al., 2006), but may also decrease the quality of inventive outcome (Beaudry & Schiffauerova, 2011), and reinforce fragmentation through the formation of closed clubs (Balland et al., 2019). Therefore, the aim of this paper is to map the structure of European inter-regional co-inventor ties, and to test how multiple dimensions of similarity shape the spatial pattern of repeated inventive collaboration.

The analysis is based on patents filed at the European Patent Office from the EU28 and continental EFTA countries (Norway and Switzerland) between 2006 and 2010. To investigate the geographies of repeated collaboration, we construct an inter-regional co-inventor network on the level of NUTS 3 regions and identify repeated co-inventor relationships from before 2006. To characterize the spatial patterns of the full collaboration network and the network of repeated ties, we apply a community finding algorithm (Blondel et al., 2008) on both networks. We also investigate how repeated co-inventor ties, compared with co-inventorhsip in general, depend on geographical distance (Liben-Nowell et al., 2005; Lambiotte et al., 2008; Lengyel et al., 2015), the overlap between technological portfolios of regions (Maggioni et al., 2007; Maggioni & Uberti, 2009), and the number of third regions as common partners (Hazir & Autant-Bernard, 2014). We also apply a multivariate gravity equation approach (Maggioni et al., 2007; Broekel et al., 2014) to test these factors and their joint effect on the repeating of co-inventor ties between regions.

In short, our results indicate that inter-regional co-inventor ties are rarely repeated, but many regions are linked through these collaborations. Repeated co-inventor ties are concentrated on a smaller geographical scale compared with collaboration in general. While gravity modelling for the overall and the repeated collaboration networks show similar patterns, a feature unique to repeated collaboration is a very strong and positive correlation between the three-way interaction of similarities between regions and the number of such ties. These results contribute to the geography of innovation literature by showing how repeated co-inventor collaborations are structured.

In Section 2 we explain the relevance of repeated inter-regional co-inventor ties. Section 3 describes the data, the community detection method, the variables and the estimation strategy of our gravity approaches. Empirical results are detailed in Section 4, while Section 5 offers a discussion of the findings.

2. The relational structure of collaborative knowledge production across European regions

While knowledge is an increasingly important ingredient for securing competitive advantage and economic development (Foray & Lundvall, 1998), it's production is achieved more and more through the collaboration of knowledgeable actors of narrow expertise (Jones, 2009). As collaboration imposes co-ordination costs on the parties involved (Jackson, 2008), similarity among them can facilitate knowledge exchange. Indeed, it is well established by now that knowledge sharing is highly concentrated and mostly takes place in dense, local networks (Singh, 2005; Breschi & Lissoni, 2009). However, at a more systemic level this means that the resources necessary for a desired level of innovative activity are likely distributed amongst multiple agents and locations (Martin & Sunley, 2007). Hence, even though distant connections are relatively sparse compared with local interactions, inter-regional ties can bring novel ideas and new knowledge to regional economies, thus fostering internal innovation processes (Bathelt et al., 2004; Fitjar & Huber, 2014; De Noni et al., 2018). As the involved parties are from different locations and institutional settings, these linkages enable the combination of a greater variety of knowledge in the innovation process. Finally, evidence also indicates that both agglomeration and collaboration networks matter for R&D productivity at different stages of the knowledge production process (Varga et al., 2014).

Hence, increasing the performance and integration of collaborative knowledge production is a central pursuit for the European Union (European Commission, 2020), all the while considerable effort has been dedicated in academia to understand the relational structure and determinants of such collaborations across European regions. Prominent instantiations of inter-regional knowledge production studied in this literature include joint participation in FP projects (Breschi & Cusmano, 2004; Maggioni et al., 2007; Maggioni & Uberti, 2009; Scherngell & Barber, 2009; Hoekman et al., 2013; Maggioni et al., 2014; Hazir & Autant-Bernard, 2014; Wanzenböck et al., 2014; Balland et al., 2019), co-patenting and co-inventor collaborations (Maggioni et al., 2007; Hoekman et al., 2009; Maggioni & Uberti, 2009; Chessa et al., 2013; Wanzenböck et al., 2014) and scientific co-publication (Hoekman et al., 2009; Hoekman et al., 2010; Chessa et al., 2013; Hoekman et al., 2013; Wanzenböck et al., 2014). This research indicates that the ERA has a dense hierarchical core (Breschi & Cusmano, 2004; Hoekman et al., 2009), and while signs of integration are present with respect to FP participation of EU15 and EU13 countries (Balland et al., 2019), by and large Europe remains fragmented along national innovation systems (Chessa et al., 2013). In addition, while joining in on collaborative knowledge production for regions of CEE countries offers benefits in terms of patent quality, it also leads to polarization in the technological profiles of those that managed to connect to partners in other regions, and those that did not (Lengyel & Leskó, 2016).

Investigation into the determinants and constraints of collaborative knowledge production revealed that the geographical and relational proximity of the parties involved are key drivers in these networks (*e.g.* Fleming & Frenken, 2007; Balland, 2012; Broekel & Boschma, 2012; Brenner et al., 2013; Ter Wal, 2013; Cassi & Plunket, 2015; Crescenzi et al., 2016; Cantner et al., 2017). Relational proximity here means the degree to which collaborators are similar to one another along different dimensions, including domains of expertise, institutional and social norms and organizational context (*e.g.* Torre & Rallet, 2005; Boschma, 2005). This micro-level observation is reflected in the geography of inter-regional collaborations of knowledge production across Europe as well. First, *geographical proximity* remains important, as less distant regions tend to collaborate more in terms of patenting, scientific publication and FP projects (Maggioni et al., 2007; Hoekman et al., 2009; Maggioni & Uberti, 2009; Hoekman et al., 2010; Hazir & Autant-Bernard, 2014). Second, similarity with respect to *technological portfolios* of regions was found to increase the likelihood of inter-regional collaboration (Maggioni et al., 2007; Maggioni & Uberti, 2009), as such similarity may facilitate combining extra-regional knowledge for both regions involved. Third, collaboration tends to emerge between regions that have existing ties to common third regions, reflecting in space the development of new connections in the underlying R&D cooperation network (Hazir & Autant-Bernard, 2014). This indicates that *triadic closure*, *i.e.*

the partners of partners becoming partners, is an important factor in shaping the geography of collaborative knowledge production.

What is missing from these accounts is how repeated inter-regional collaboration is structured, and how the above drivers of spatial collaboration patterns affect these relations in particular. Exploring repeated interactions in collaborative knowledge production is important as it is unclear based on the literature how this type of inter-region collaborations affect innovative performance. The aim of this study is to provide insights on these collaborations in the context of inter-regional co-inventor networks across Europe.

On the one hand, repeated collaborations may be beneficial for knowledge production. Inventors represent the most immediate and influential social environment for the technical contents of an invention, they stay in touch for the duration of the collaboration and could get in touch again later (Breschi & Lissoni, 2005). Hence, past collaborators may represent know-who in terms of missing technical competences needed in a subsequent R&D project, and re-establishing co-operation among inventors who have worked together previously may be less costly due to past experience. Inventors regularly collaborating with one another may receive additional benefits by developing deeper understanding towards each other through the invention process. Such established relationships facilitate the development of more complex technological solutions (Reagans & McEvily, 2003; Sorenson et al., 2006; Aral, 2016), the repeating of a collaboration also signals stability, trust, mutual interest (Seabright et al., 1992), and represents long-term strategies for collaboration (Dahlander & McFarland, 2013; Goerzen, 2007). Therefore, in the context of an inter-regional co-inventor network, many repeated connections can indicate that regions are relative well connected by relationships that involve these benefits. Additionally, numerous repeated ties to distant places can act as established pipelines that bring new technological knowledge to the regional economy. This is important as recent studies indicate that channels to extra-regional knowledge open up the possibility for diversification into new technologies (Balland & Boschma, 2021; Whittle et al., 2020).

On the other hand, arguments can be put forward that the repeated collaboration of inventors may be detrimental to knowledge production. Repeated cooperation means more shared time and which implies heavy opportunity costs (Uzzi, 1997; Goerzen, 2007), especially when the repeating of collaboration is frequent. In such cases the growing overlap of knowledge bases also contribute to the circulation of redundant knowledge. For instance, case-study evidence indicates that the repeated collaboration of inventors has a negative impact on the resulting patent quality (Beaudry & Schiffauerova, 2011). This means that in an inter-regional co-inventor network, regions connected by many repeated collaborations can limit their access towards diverse knowledge sources and might face a lock-in situation in the long run (Malerba, 2009; Boschma & Frenken, 2010). A high overlap between inter-regional collaboration networks over time is indicative of the formation of closed clubs, that would go against the aim of an integrated research area across Europe. While such overlap was found to be relatively rare for the FP project network of EU countries (Balland et al., 2019), research into the geography of scientific co-publication and co-patenting indicates the presence of elite structures (Hoekman et al., 2009), meaning that regions hosting high quality scholars and funding tend to network among themselves.

In the following, we explore the spatial structure of co-inventor linkages and repeated co-inventor linkages between European regions, and model how geographical and relational proximity affects the formation of these collaboration networks.

3. Data and Methods

3.1. Data

For the analysis we use the full set of patents authored by European inventors, and registered by the European Patent Office (EPO). Patents became a widely used data source to construct such networks as they offer granular information on the location of inventors and on the invented technology for a relative long period of time (Griliches, 1990; Desrochers, 1998). Notwithstanding these merits, it is also widely acknowledged in the literature that patent data has its drawbacks, as some industries are more prone to patenting than others (*e.g.* Graf & Henning, 2009), and patents represent only a partial picture of inventors' social networks in particular, and cooperation for innovation more generally (Fritsch et al., 2020). Therefore, our analysis shows only a partial picture on inter-regional collaborative knowledge production.

The data is from the publicly available OECD REGPAT 2015 Database that contains the year of filing, the technological classes of patents, the unique identifier of inventors and the region of their residence. Figure 1 offers an overview of the data management process. For every patent the database includes a unique ID and location information at the NUTS3 level of the inventor(s), as well as different technology classes (4-digit IPC class in our case) that the patent was assigned to (Step 1).

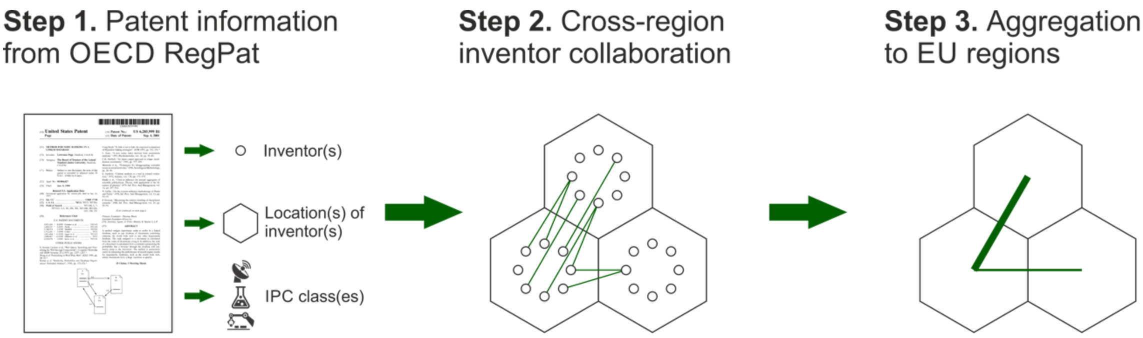


Figure 1. Schematic visualization of deriving the inter-regional co-inventor network. Source: Own illustration.

From this patent information, we derive the inventor collaboration network, in which two inventors are linked if they co-authored at least one patent over the 2006 to 2010 period (Step 2). Patents can be considered an outcome observed in the end of a successful collaboration, and while evidence indicates that R&D project collaboration tends to take up about 2-3 years, there is also considerable variation associated with this (Greve et al., 2009; Phelps, 2010; Ramlogan & Consoli, 2014). We opted for considering an inventor-inventor tie *repeated*, if the inventors had co-authored at least one patent in the 1991-2005 period as well. This implies that repeated collaboration in our analysis will represent a combination of frequently recurring co-operations and collaborations where inventors work together after a hiatus.

Finally, we aggregate the co-inventor network on the region-region level so that the tie weights between regions are proportional to the number of inventor collaborations between regions in the 2006 to 2010 period (Step 3). In a similar fashion, we aggregate repeated co-inventor links on the region-region level. Admittedly, a drawback of basing the inter-regional collaboration network on individuals as opposed to organisations is that boundaries of the firm are not considered (Ter Wal & Boschma, 2009). This decision implies that pieces of knowledge combined into a patent are embodied in individuals, rather than organizations they are affiliated with, and inventors have agency in creating their collaboration ties, which may be limited in a corporate setting in particular. An alternative would be deriving the inter-regional network from collaborative ties between organizations. However, co-patenting at the organizational level tends to be rare (Ter Wal & Boschma, 2009; Fritsch & Zoellner, 2020), likely due to the fact that cooperating companies prefer to avoid the legal complexity involved in co-patenting by dividing patents from joint R&D effort (Hagedoorn, 2003). Finally, critical to our investigation on spatial patterns, patents of multi-establishment companies tend to be assigned to the headquarters of these firms (Ter Wal & Boschma, 2009), which is amplified by the fact that bigger firms tend to patent more while they are also more likely to have multiple branches. For these reasons we opt for the co-inventor level when deriving the network. Additionally, the aggregation of ties also implies that our results should be less sensitive to the potential issues related to inventor identification over time and space (Raffo & Lhuillery, 2009; Schoen et al., 2014), and show general patterns at a large scale even when the collaboration between inventors cannot be observed directly.

3.2. Variables for modelling inter-regional co-inventor ties

Figure 2 illustrates the three key variables involved in the empirical analysis. First, *geographical proximity* P_{ij} between regions i and j is calculated as $max(D_{ij}) - D_{ij}$, where D_{ij} denotes the physical distance measured in kilometres between the centroids of each region. Second, a common way of quantifying technological similarity between two regions i and j is to calculate the *cosine similarity* (C_{ij}) of their patent portfolio vectors V_i and V_j. This variable is defined as $C_{ij} = \frac{V_i \cdot V_j}{|\sum_{i=1}^{M} V_i^2||\sum_{j=1}^{M} V_j^2|}$, where the numerator is the inner product of the regional patent portfolios. These portfolios contain the patents made between 2000 and 2005 for each 4-digit IPC class in the region. The denominator is the product of the Euclidean length of each patent portfolio vector (Schütze et al., 2008). In our case cosine similarity ranges from zero to one, where zero is the case of perfectly unrelated portfolios and one represents complete similarity. Finally, to capture the share of common third partners for each pair of regions, we use the *Jaccard index*, that measures the overlap between finite sample sets and is defined as the cardinality of the intersection divided by the cardinality of the union (Leydesdorff, 2008): $J_{ij} = \frac{|A_i \cap A_j|}{|A_i \cup A_j|}$, where A_i and A_j refer to the underlying collaboration vectors of regions i and j for the 2000-2005 period. This variable ranges from zero to one, where higher values indicate a higher share of common third partners.

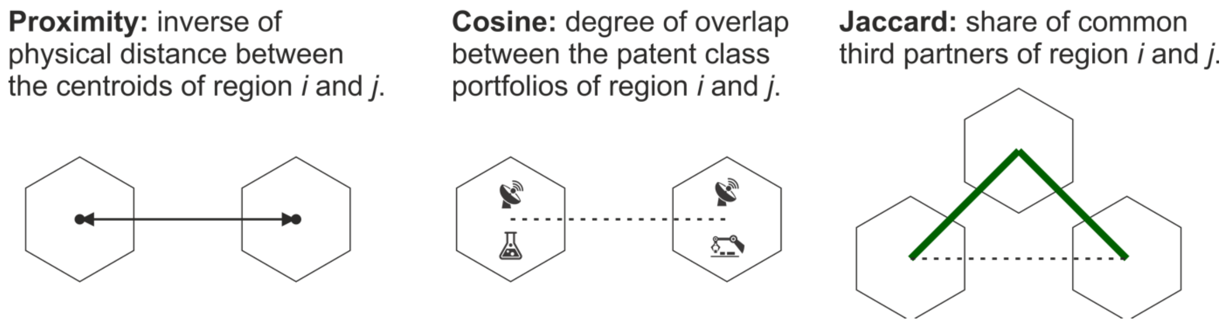


Figure 2. Schematic visualization of the network variables. Source: Own illustration.

To make the comparison of coefficients easier, the three main explanatory variables are rescaled to zero mean z-scores, where a unit change in z-score is equivalent to one standard deviation in the original variable. Table 1 presents the descriptive statistics for these three key variables.

Table1. Descriptive statistics and correlation of the standardized main explanatory variables.

	Mean	Min	Max	Cosine	Proximity	Jaccard
Cosine$_{ij}$	0	-1.564	3.376	1		
Proximity$_{ij}$	0	-2.749	8.364	0.258	1	
Jaccard$_{ij}$	0	-0.776	12.021	0.401	0.652	1

Source: Own calculation based on OECD REGPAT 2015.

3.3. Applied methods: community detection, gravity approach and regression design

As an initial step, we characterize the *spatial community structure* of (repeated) inter-regional co-inventor ties by using the Louvain algorithm (Blondel et al., 2008), a popular and effective method for partitioning a network through hierarchical clustering. This algorithm finds the best grouping of nodes - in our case a NUTS3 region - in a network by optimizing the modularity of the community structure, which is the measure of link concentration within groups of the partitioning, rather than between them (Newman, 2006). The modularity[1] Q of the network's partitioning can be written as $Q = \sum_{k=1}^{K}\left[\frac{L_k^W}{L} - \left(\frac{L_k}{L}\right)^2\right]$ where L is the total number of individual co-

[1] Because modularity is highly dependent on the size and density of the network, the mere comparison of the measure obtained from two networks - such as the full and repeated inter-regional collaboration networks - can be misleading. Hence, following Sah et al., (2017), we calculate the relative modularity of the two networks by dividing the modularity (Q) by the theoretical maximum (Q_{max}), that would be achieved if all links were within the communities. We report these tests in Table 1 and Table 2 of the Appendix.

inventor links in the network, L_k is the total number of links for members of group k, and L_k^w is the number of links within group k.

Next, we take a *gravity approach*, that is often applied in network science (Liben-Nowell et al., 2005; Lambiotte et al., 2008; Lengyel et al., 2015), to measure how distance decay (D), cosine similarity of technological portfolios (C) and collaborations with the same third regions (J) change the probability of collaboration and repeated collaboration between two regions. Distance decay influences network density (Liben-Nowell et al., 2005), and thus yields spatial communities in and of itself (Expert et al., 2011). Because technological similarity of regions and triadic network tie formation influence the establishment and repeating of collaboration (Maggioni et al., 2007; Ter Wal, 2013; Juhász & Lengyel, 2018), we consider probabilities as a function of these dimensions as well. We bin the distributions of region-region links into intervals for all three characteristics (denoted by b lower index). Then we calculate the probability of collaboration and repeated collaboration for every group such that

$$\Pr\left(\text{collaboration}_t \middle| D_b, C_{b,t-1}, J_{b,t-1}\right) = \frac{\sum\left(L_{ij} \middle| D_b, C_{b,t-1}, J_{b,t-1}\right)}{\left(N_{i,t} \times N_{j,t}\right)/2} \tag{1}$$

where L_{ij} is the number of observed individual co-inventor connections between regions i and j in the corresponding network, whereas N_i and N_j refer to the number of inventors in these regions who authored at least one patent in the 2006-2010 period, regardless whether they established a new collaboration or repeated an old one. To keep simplicity in the notation, we refer to time period 2006-2010 as t and to the previous period 2000-2005 as $t-1$.

Finally, we rely on a *regression framework* for disentangling the effect of multiple dimensions of similarity behind the spatial structure of the collaboration and repeated collaboration networks. Here the strength of region-region links is estimated in regression analyses with dyadic covariates and characteristics of both regions involved in the dyad. We look for differences in the correlation values of co-variates between the strength of collaboration and repeated collaboration ties on the same set of region-region links, for which the co-variates are identical. Therefore, a necessary condition for this analysis is the variation of collaboration versus repeated collaboration tie strengths. While his correlation is very strong (see Appendix Figure 1), it is paired with sufficient variance between these values.

The strength of collaboration and repeated collaboration between regions can be considered as count data of individual co-inventor links, in which most region-region links account for zero individual connections. Therefore, we apply a Zero-Inflated Negative Binomial (ZINB) regression, which consists of two parts (Greene, 1994). The first equation in the ZINB modelling is often referred to as regime selection, which is employed to deal with excessive zeros in the data and is formulated as

$$\Pr(Y_{ij,t}) = \gamma_0 + \gamma_1\theta_{ij} + \gamma_2(Z_{i,t-1}, Z_{j,t-1}) + \varepsilon_{ij,t} \tag{2}$$

where we estimate the probability that a connection may develop between two regions. In the equation θ_{ij} is a dummy variable that takes the value of 1 if region i and region j are in the same country, and Z_i, Z_j are a collection of region-level control variables that are commonly used in similar estimations.

The second equation refers to the count process, in which we estimate the number of co-inventor ties between regions by our three main variables:

$$log(Y_{ij,t} = y_{ij,t}) = \beta_0 + \beta_1 P_{ij} + \beta_2 C_{ij,t-1} + \beta_3 J_{ij,t-1} + \varepsilon_{ij,t} \tag{3}$$

where $Y_{ij,t}$ is the strength of collaboration or repeated collaboration between regions i and j in the period of 2006-2010, P_{ij} refers to geographical proximity, while $C_{ij,t-1}$ and $J_{ij,t-1}$ refer to variables explained above, based on data from the previous period 2000-2005. Besides transforming distance to proximity, which is important to have all correlations with identical signs, we standardize every variable so that coefficient sizes can be compared across varying scales. However, the correlation between geographical proximity, technological similarity and shared third connected regions call for better understanding. Therefore, in a further specification we include the interaction effect of our dichotomized main variables[2]. For a more detailed explanation on ZINB specifications see Greene, (1994) and Burger et al., (2009).

4. Results

[2] This transformation is neccesary because the three-way interactions of continous variables would be difficult to interpret otherwise.

We summarize the major characteristics of inter-regional co-inventor ties in Europe in Table 2. First, out of more than 772K inter-regional collaborations between inventors in 2006-2010, only every 18th tie was a repeated one. The low share of repeated collaboration signals that inventive knowledge is mostly produced through new collaborations. By aggregating these linkages to the level of European NUTS3 regions, we get a weighted network of regions in which the weight of a region-region tie is the number of individual collaborations connecting the two. After the aggregation, the collaboration network of regions consists of nearly 47K weighted ties, while a repeated collaboration exists only between one 7th of these region pairs. Consequently, the density of the repeated collaboration network is one magnitude lower than the density of the complete collaboration network. As repeated ties appear to be rare between both inventors and regions, we interpret the following results accordingly.

Table 2. Characteristics of the co-inventor network and the repeated collaboration network.

	Collaboration	**Repeated Collaboration**
Number of Individual Collaborations	772,378	41,883
Number of Region Ties	46,857	6,200
Density of the Region Network	0.05	0.006
Number of Communities in the Region Network*	7	23
Modularity of the Region Network	0.372	0.584
Relative Modularity of the Region Network	0.329	0.379

Source: Own calculation based on OECD REGPAT 2015.

Note: *Communities of size 1 are excluded.

4.1. Spatial communities of inventor collaboration

Our findings reported in Table 2 reveal a remarkably lower number of communities in the collaboration network, compared with the repeated collaboration network. These communities are groups of regions within which regions are relatively densely connected by collaboration ties, but regions across community borders are relatively loosely connected. The *modularity* measure further demonstrates that repeated collaboration tends to be concentrated to a higher extent within these communities than collaboration in general. However, this difference is due to the lower level of network density of repeated collaboration links, as both networks show similar levels of relative modularity. Taken together, these findings show that repeated collaboration tends to be fragmented into region groups, much like it is the case for co-inventor collaboration in general (Chessa et al., 2013).

In Figure 3 we map both networks and their community structures. Since the collaboration network has too many links to fit meaningfully on a map, and also because the tie weights disperse on a large scale, we simplify these networks into their maximum spanning tree for visualisation. In these, every region is connected to every other region by only one path such that the sum of the tie weights are maximized. Figure 3A reveals that most European countries have an outstanding innovation center, in which collaboration is concentrated, and these centers bridge the innovation system with other countries (*e.g.* in Spain, Sweden, Finland, Italy, Hungary and Romania). There is more than one center in Poland, France and especially in Germany, where regional centers emerge from the maximum spanning tree. Most of these spatial structures are present in the repeated collaboration network as well, in which all ties are depicted in Figure 3B. Repeated collaboration is concentrated in single innovation hubs in Sweden, Finland, Italy, whereas two centers emerge in France (Paris and Lyon) and multiple centers in Germany, where these centers are strongly connected to each other.

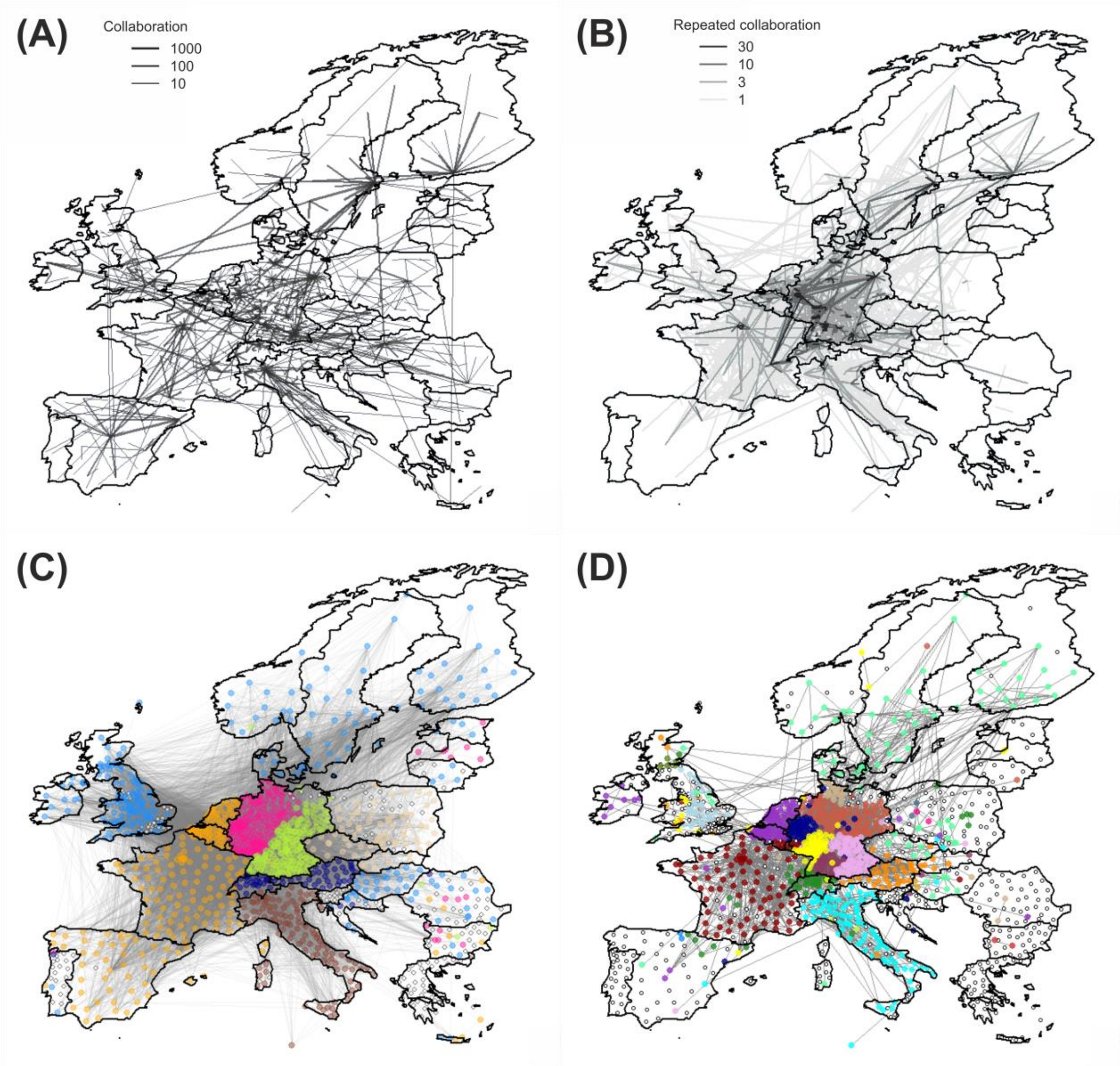


Figure 3. Spatial patterns of inventor collaboration and repeated collaboration networks of European regions. Source: Own calculation based on OECD REGPAT 2015.

Note: **(A)** The maximum spanning tree of the collaboration network across NUTS3 regions in Europe reveals the importance of national centers. **(B)** Most of the repeated collaborations remain within country borders, and strongest ties are concentrated within close proximity of innovative hubs. **(C)** The 7 communities of the collaboration network span across countries, with the exception of Germany that is divided into two communities and Italy, but are mostly concentrated in large regions. **(D)** Repeated collaboration is organized into 23 smaller-scale clusters.

Both networks are organized into spatially bounded communities (Figure 3C and Figure 3D), which is not surprising since spatial community structures have been repeatedly found in social and communication networks (Lambiotte et al., 2008; Sobolevsky et al., 2013; Lengyel et al., 2015). Chessa et al. (2013) also reported that communities of inventor collaboration are bounded by national borders in Europe. Our findings suggest that some of the communities expand countries: for instance, the community including the Benelux states, France and Spain. These communities contain neighbouring countries such as Switzerland and Austria, or the community of Poland, Czech Republic and Slovakia. However, the United Kingdom, Hungary and Croatia are found to be in one community with the Scandinavian countries. We find that inventor collaboration in Italy is a separate network community. Interestingly, Germany is organized into two large communities that do not follow perfectly the East-West divide, as particularly the southern part of the country belongs to the same community as the eastern part. This may be a sign of reorganization of inventive collaboration after the fall of the Iron Curtain (Jun et al., 2017).

Repeated inter-regional collaborations are not organized by a universal pattern of spatial levels. In some cases these communities represent countries (e.g. France and Italy), groups of countries (e.g. Belgium and the Netherlands, or the Scandinavian countries), spatially clustered regions that span across countries (e.g. in the UK) and spatially concentrated communities with very few overlaps (in Germany). Nevertheless, all larger communities of inter-regional collaboration break into smaller ones of repeated collaboration. This finding suggests that collaboration is relatively more likely to be repeated between geographically proximate locations while collaboration in general is relatively more likely to bridge distant locations.

These results have implications with respect to the European innovation system, as one of the major R&I policy aims of the EU is to integrate disconnected national systems of innovation into the ERA in order to increase access to novelty and innovation output. Although inventor collaboration in general show signs of ongoing integration, it is still bounded by national systems of innovation. Repeated collaboration, however, seems to be even more difficult to integrate into a unified ERA. In the following subsections we aim to better understand the determinants behind this geography of co-inventor collaborations.

4.2. Gravity approach on inter-regional co-inventor ties

In the next stage of the analysis, we explore the probability of repeated collaboration, compared with collaboration in general, as a function of distance, technological similarity and shared third partner regions (Figure 4). Since we express these probabilities as observed over all possible ties for both networks, the smaller probability for repeated collaboration comes from the lower density of individual connections. However, these differences are surprisingly stable across all three distributions.

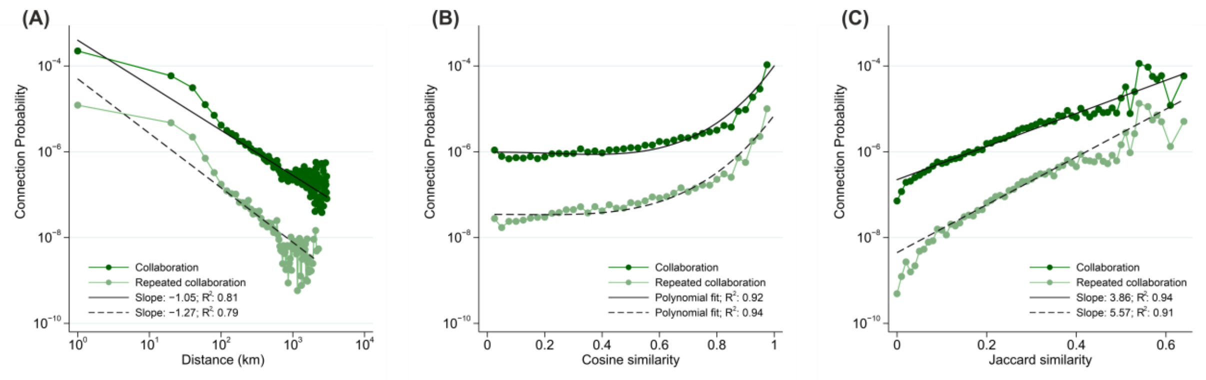


Figure 4. The probability of collaboration and repeated collaboration as a function of region-to-region characteristics. Source: Own calculation based on OECD REGPAT 2015.

Note: **(A)** Distance ($D_{b=20km}$) decay is smooth for geographically proximate collaboration and repeated collaboration, following a linear decay on log-log scale with the exponents -1.05 and -1.27 for distances larger than 100 km. **(B)** The overlap between technological profiles of regions, measured by cosine similarity ($C_{b=0.025}$), increases the probability of collaboration with a growing intensity as similarity rises. **(C)** The probability of collaboration grows linearly on a logarithmic scale as the share of common third connected regions, measured by Jaccard similarity ($J_{b=0.01}$), increases. The exponent is 3.86 for collaboration and 5.67 for repeated collaboration.

The exponent of *distance decay* in the repeated collaboration network is somewhat higher than in the complete collaboration network. This signals that the role of geographical proximity is more important for repeated collaborations because the probability of co-inventing decays faster as distance grows. However, these power laws can be fitted to the middle of the distance distribution only, and curves are remarkably similar for distances smaller than 100 km (Figure 4A). Regarding *technological similarity*, we find that a polynomial fit captures the relationship between cosine similarity and link probability very well (Figure 4B). This suggests that a wide overlap of technological profiles of regions increases the probability of collaboration and repeated collaboration ties in a similar manner. Moreover, above a threshold of extensive technological similarity, these probabilities increase exponentially, revealing that collaboration is more likely between regions that have strongly similar technological profiles. In terms of *shared third partners*, we find that a higher value of Jaccard similarity is associated with a higher link probability. Linear fits on the semi-logarithmic graph indicate an exponential relationship, where the exponent is higher for repeated collaboration (Figure 4C). This indicates that tie-dependence is an important driver of repeated co-inventor interactions between regions.

Taken together these findings suggest that while the association between various dimensions of similarity and the likelihood of inter-regional collaboration show similar patterns for repeated collaboration and for collaboration in general, the former is subject to a more pronounced distance decay and is influenced more by the existence of common third partners. In the next section we complement this exploration with regression analysis to investigate the interaction effect of these dimensions.

4.3. Regression analysis on inter-regional co-inventor ties

Multivariate gravity models enable us to compare coefficients of the three determinants of inter-regional collaboration and also informs us whether these coefficients differ for repeated collaboration and collaboration in general. Estimations of the regression model are illustrated in Figure 5 where point estimates are presented with their 95% confidence intervals. Figure 5A depicts coefficients estimated from Equation 3. Because all three explanatory variables are standardized, coefficients show the expected change in the log number of collaboration and repeated collaboration in case there is one standard deviation change in the independent variable.

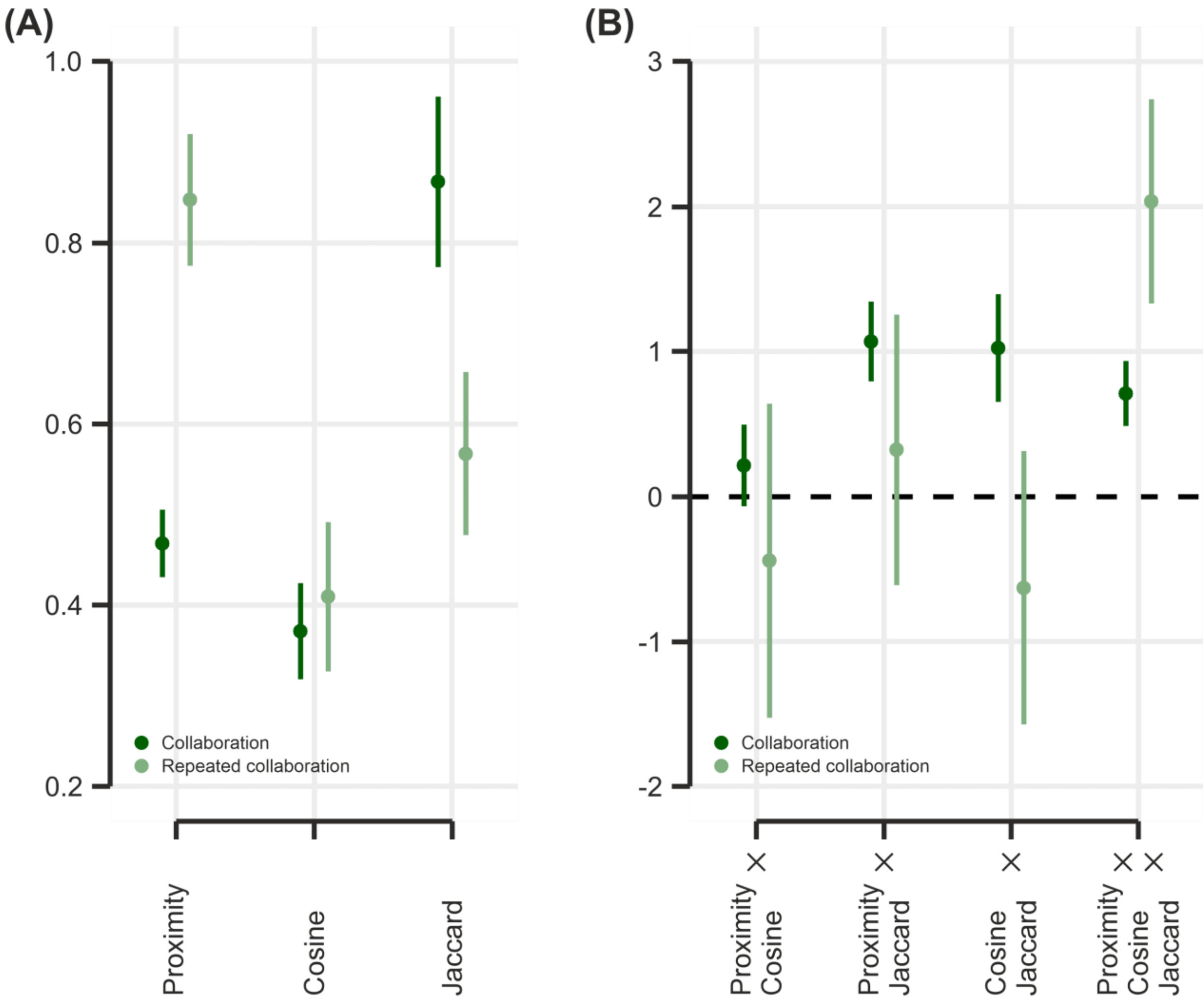


Figure 5. Results of the multivariate estimations. Source: Own calculation based on OECD REGPAT 2015.

Note: **(A)** Considering single effects only, we observe that the probability of collaboration is mostly increased by common third partners, while repeated ties gain probability if regions are geographically proximate. **(B)** Interaction effects reveal that repeated ties gain extra probability if regions are geographically proximate, technologically similar and connected to the same set of regions at the same time.

The findings suggest that geographical proximity is more important for repeated collaboration than what is observed in the full collaboration network. Technological similarity has a relatively small correlation with both values of collaboration and repeated collaboration across European regions. Interestingly, high shares of common third connected regions favour collaboration more but not so much the repeating of collaborations. These aggregate level findings are in line with previous results of Ter Wal, (2013) on co-inventor networks, who argues that cohesive network formation decreases costs and uncertainties for new link formation, but once the collaboration is established, geographical proximity is more important in decreasing the costs of maintaining and thus further strengthening these relations. The importance of geographical proximity gives a hint already at why spatial communities of repeated networks are of smaller spatial scale than the ones in the general collaboration network. Additional details of the parameter estimations are presented in Table 3 of the Appendix.

We applied a variety of robustness checks to validate our results. Collaboration ties can only be repeated in case they existed previously. This means that the number of repeated collaborations highly depends on the number of collaborations in the past. To take this deterministic process into account, as a first robustness check, we include the number of existing co-inventor ties between regions i and j, based on the 1991-2005 period. However, one can argue that in order to correctly estimate the repeating of collaborations, the sample must be narrowed down to regions with non-zero collaborations in the past. To do so, in another robustness check, we run our original model on repeated ties based only on the sub-sample of previously connected regions. These results are similar to the original estimation, but the different sample sizes would make the comparison of betas precarious. Additionally, repeated collaboration can be driven by intra-firm collaboration. As a further robustness check, we identified firms as assignees of patents and estimated regressions by keeping only those repeated ties where the collaboration happens across different firms. For this, we exclude those repeated collaborations where we found repeating sequences in the assignees' names. As a theoretical example, patents made by "Best Toy Factory, Budapest" and "Best Toy Factory, Paris" are considered as intra-firm collaboration. These latter tests did not change our results substantially. Robustness checks are reported in Appendix Table 3 Model 3, 4 and 5.

In Figure 5B, we illustrate the coefficients and 95% confidence interval of interaction terms of our main explanatory variables. Additional details of the estimations are presented in Table 4 of the Appendix. The interaction coefficients of ties in the full collaboration network are positive and significant in all cases when the Jaccard index is involved. This means that connections towards the same set of regions increases the effect of both geographical proximity and technological similarity. More importantly, we find that repeated collaboration is increased by interaction terms only if all three variables are included in the joint effect. This finding suggests that the repeating of inter-regional inventor collaboration in Europe gains extra likelihood when regions are geographically proximate, are similar in their technological profile and connected to the same set of other regions. Consequently, the communities of repeated collaboration are of smaller spatial scale not only because the effect of geographical proximity is at work, but also because this is coupled with technological similarity and shared third regional connections.

5. Conclusion

Collaborative knowledge production in an integrated ERA is a lasting pursuit of the European Union (Council of the European Union, 2009; European Commission, 2020), however a series of studies document the illusiveness of this aim (*e.g.* Hoekman et al., 2009; Hoekman et al., 2010; Chessa et al., 2013), as well as that collaboration is subject to various dimensions of geographical and relational proximity (*e.g.* Breschi & Cusmano, 2004; Roediger-Schluga & Barber, 2008; Maggioni et al., 2007; Maggioni & Uberti, 2009; Scherngell & Lata, 2013). The aim of this paper has been to explore an aspect of such collaborations hitherto neglected in the related literature, namely the relational structure of those collaborations that were repeated over time. Deploying multiple tools of analysis in the context of inventor collaborations across European NUTS3 regions, we found that only a small fraction of co-inventor linkages across regions are repeated, and they are clustered in geographical space more intensively compared with the complete collaboration network. Our results show that collaborations in the inter-regional co-inventor network emerge mainly through the triadic collaboration of regions, while geographical proximity becomes the most influential factor for the repeating of co-inventor collaboration. In addition to that, the combination of technological similarity and shared third partner regions offer a premium for the likelihood of repeated collaboration, but only when geographical proximity is present as an enabler.

Hence, these findings indicate that despite establishing the ERA, inventive knowledge production, as partially proxied in this paper by co-inventorship, is still fragmented into spatial formations resembling national systems of innovation. The results on repeated inter-regional collaboration add that collaborations involving inventors who have already worked together before are likely to revert to spatial clustering of co-inventorship between similar technological profiles, thus contributing to fragmentation.

What is curious is the relatively low share of repeated collaboration in inter-regional co-inventor activities. As argued earlier, such collaborations involve accumulated experience in working together, and consequently one could expect the emergence of more robust spatial and relational structures on the back of these connections. However, the fact that this type of collaboration in inter-regional co-inventorship is relatively rare points to a very dynamic system of knowledge production across Europe. According to the findings presented in this paper, higher order tie-dependence in the form of shared third partners brings more stability to these collaborations. As a result our study does not find evidence on the formation of closed clubs in inter-regional co-patenting on the back of repeated collaboration. This way it complements the recent findings of Balland et al., (2019) on a relatively low level of overlap between FP project collaboration networks over time, and evidence by Fritsch and Zoellner (2020), showing that there is a high level of fluidity in German inter-regional co-inventor networks where only a small share of links between inventors carry over from a 3 year period to the next. These accounts are prompting the question why dyadic connections from various collaborative knowledge production efforts in the European spatial structure tend to not get carried over to future collaboration.

Taking up on this question goes beyond the confines and limitations of this paper and should be taken up in future research. First, as discussed earlier, it is unclear based on the literature whether repeated ties of inter-regional collaboration represent opportunities or threats with respect to the quality of knowledge produced. More broadly, it is unclear whether patents created in repeated collaborations are different from patents created in new collaborations. Due to lack of information, we could not engage with these questions in this paper. Second, we identified collaboration network dynamics in a simple way by looking at repeated ties that could be modeled using a gravity approach. However, the formation and repeating of collaboration ties are strongly influenced by the previous features of the network (Broekel et al., 2014). Hence, one might investigate network dynamics more in detail with more nuanced definitions of a tie and applying models that take network topology into account in a more sophisticated way. By doing so, one could reveal how network cohesion favours the repeating of individual ties and how technological profiles of inventors contribute to this process in geographical space.

Acknowledgements

The authors are thankful for the services of the SZTAKI Cloud group at the Hungarian Academy of Sciences. The authors acknowledge comments received at the workshop of MIT Media Lab and The Institute for New Economic Thinking Young Scholars Initiative on ”Innovation, Complexity and Economic Geography”, the ”Innova 4” workshop in the Joint Research Center of the European Commission in Ispra, the ”4th RA X Networks” workshop at Andrássy University Budapest, the 2018 Conference of the Hungarian Regional Science Association in Kecskemét and ICUBERD workshop at Pécs University. G.T. would like to acknowledge funding from the Science Foundation Ireland (SFI) under the SFI Science Policy Research Programme (grant agreement No 17/SPR/5324, SciTechSpace).

References

Aral, S. (2016). The Future of Weak Ties. *American Journal of Sociology, 121*(6), 1931–1939.

Balland, P-A. (2012). Proximity and the evolution of collaboration networks: evidence from research and development projects within the global navigation satellite system (GNSS) industry. *Regional Studies, 46*(6), 741–756.

Balland, P-A., & Boschma, R. A. (2021). Complementary interregional linkages and Smart Specialisation: an empirical study on European regions. *Regional Studies*, published online.

Balland, P-A., Boschma, R., & Ravet, R. (2019). Network dynamics in collaborative research in the EU, 2003–2017. *European Planning Studies, 27*(9), 1811-1837.

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography, 28*(1), 31–56.

Beaudry, C., & Schiffauerova, A. (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *European Management Journal, 29*(5), 362–376.

Blondel, V. D., Guillaume, J-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional Studies, 39*(1), 61-74.

Boschma, R., & Frenken, K. (2010). The spatial evolution of innovation networks. A proximity perspective. In Boschma, R., & Frenken, K. (eds.): *The Handbook of Evolutionary Economic Geography*. Cheltenham – Northampton: Edward Elgar, 120–135.

Brenner, T., Cantner, U., & Graf, H. (2013). Introduction: Structure and dynamics of innovation networks. *Regional Studies, 47*(5), 647–650.

Breschi, S., & Cusmano, L. (2004). Unveiling the texture of a European Research Area: emergence of oligarchic networks under EU Framework Programmes. *International Journal of Technology Management, 27*(8), 747-772.

Breschi, S., & Lissoni, F. (2005): Knowledge Networks from Patent Data. In Moed, H. F., Glänzel, W., & Schmoch, U. (eds.): *Handbook of Quantitative Science and Technology Research*. Kluwer: Dordrecht, 613-643.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography, 9*(4), 439–468.

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

Broekel, T., Balland, P-A., Burger, M., & van Oort, F. (2014). Modeling knowledge networks in economic geography: a discussion of four methods. *The Annals of Regional Science, 53*(2), 423–452.

Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography, 12*(2), 409–433.

Burger, M., van Oort, F., & Linders, G. J. (2009). On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis, 4*(2), 167-190.

Cantner, U., Hinzmann, S., & Wolf, T. (2017). The Coevolution of Innovative Ties, Proximity, and Competencies: Toward a Dynamic Approach to Innovation Cooperation. In Glückler, J., Lazega, E. & Hammer, I. (eds.): *Knowledge and Networks*. Cham: Springer, 337–372.

Cassi, L., & Plunket, A. (2015). Research collaboration in co-inventor networks: combining closure, bridging and proximities. *Regional Studies, 49*(6), 936–954.

Chessa, A., Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., & Riccaboni, M. (2013). Is Europe evolving toward an integrated research area? *Science, 339*(6120), 650-651.

Council of the European Union (2009). Conclusions of the Council on the definition of a '2020 vision for the European research area'. *Official Journal of the European Union,* 2009/C 25/01.

Crescenzi, R., Rodriguez-Pose, A., & Storper, M. (2007). The territorial dynamics of innovation: a Europe–United States comparative analysis. *Journal of Economic Geography, 7*(6), 673-709.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? On proximity and collaborative knowledge creation. *Research Policy, 45*(1), 177–194.

Dahlander, L., & McFarland, D. A. (2013). Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly, 58*(1), 69–110.

De Noni, I., Orsi, L., & Belussi, F. (2018). The role of collaborative networks in supporting the innovation performances of lagging-behind European regions. *Research Policy, 47*(1), 1–13.

Delanghe, H., Muldur, U., & Soete, L. (2009). *European science and technology policy: Towards integration or fragementation?* Cheltenham – Northampton: Edward Elgar.

Desrochers, P. (1998). On the abuse of patents as economic indicators. *The Quarterly Journal of Austrian Economics, 1*(4), 51–74.

European Commission (2010). *Europe 2020: A European Strategy for Smart, Sustainable and Inclusive Growth.* Brussels: European Commission.

European Commission (2020). *Science, Research and Innovation Performance of the EU 2020. A Fair Green and Digital Europe*. Brussels: European Commission.

Expert, P., Evans, T. S., Blondel, V. D., & Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences, 108*(19), 7663-7668.

Fitjar, R. D., & Huber, F. (2014). Global pipelines for innovation: insights from the case of Norway. *Journal of Economic Geography, 15*(3), 561–583.

Fleming, L., & Frenken, K. (2007). The evolution of inventor networks in the Silicon Valley and Boston regions. *Advances in Complex Systems, 10*(01), 53–71.

Foray, D., & Lundvall, B-Å. (1998). The knowledge-based economy: from the economics of knowledge to the learning economy. In Neef, D., Siesfeld, G. A., & Cefola, J. (eds.) *The Economic Impact of Knowledge*. Woburn, MA: Butterworth-Heinemann, 115-121.

Frenken, K., Hoekman, J., & van Oort, F. (2007). *Towards a European Research Area*. Rotterdam: Nai Publishers.

Fritsch, M., Titze, M., & Piontek, M. (2020). Identifying cooperation for innovation—a comparison of data sources. *Industry and Innovation, 27*(6), 630-659.

Fritsch, M., & Zoellner, M. (2020). The fluidity of inventor networks. *The Journal of Technology Transfer, 45*, 1063-1087.

Goerzen, A. (2007). Alliance networks and firm performance: the impact of repeated partnerships. *Strategic Management Journal, 28*(5), 487-509.

Graf, H., & Henning, T. (2009): Public Research in Regional Networks of Innovators: A Comparative Study of Four East German Regions. *Regional Studies, 43*(10), 1349-1368.

Greene, W. H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. *NYU Working Paper*, No. EC-94-10.

Greve, H. R., Baum, J. A., Mitsuhashi, H., & Rowley, T. J. (2010). Built to last but falling apart: Cohesion, friction, and withdrawal from interfirm alliances. *Academy of Management Journal, 53*(2), 302-322.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, *28*(4), 1661–1701.

Hagedoorn, J. (2003). Sharing intellectual property rights—an exploratory study of joint patenting amongst companies. *Industrial and Corporate Change, 12*(5), 1035-1050.

Harrap, N., & Doussineau, M. (2017). Collaboration and networks: EU13 participation in international science. *JRC Policy Insights, Stairway to Excellence Brief Series*, Issues 2.

Hazir, C. S., & Autant-Bernard, C. (2014). Determinants of cross-regional R&D collaboration: some empirical evidence from Europe in biotechnology. *The Annals of Regional Science, 53*(2), 369-393.

Heimeriks, G., Li, D., Lamers, W., Meijer, I., & Yegros, A. (2019). Scientific knowledge production in European regions: patterns of growth, diversity and complexity. *European Planning Studies, 27*(11), 2123-2143.

Hoekman, J., Frenken, K., & Van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science, 43*(3), 721-738.

Hoekman, J., Frenken, K., & Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy, 39*(5), 662-673.

Hoekman, J., Scherngell, T., Frenken, K., & Tijssen, R. (2013). Acquisition of European research funds and its effect on international scientific collaboration. *Journal of Economic Geography, 13*(1), 23–52.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *The Review of Economic Studies, 76*(1), 283–317.

Juhász, S., & Lengyel, B. (2018). Creation and persistence of ties in cluster knowledge networks. *Journal of Economic Geography, 18*(6), 1203–1226.

Jun, B., Pinheiro, F., Buchmann, T., Yi, S., & Hidalgo, C. (2017). Meet me in the middle: The reunification of Germany's research network. *arXiv preprint*, arXiv:1704.08426.

Lambiotte, R., Blondel, V. D., De Kerchove, C., Huens, E., Prieur, C., Smoreda, Z. & Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications, 387*(21), 5317–5325.

Lengyel, B., & Leskó, M. (2016). International Collaboration and Spatial Dynamics of US Patenting in Central and Eastern Europe 1981-2010. *PLOS One, 11*(11), e0166034.

Lengyel, B., Varga, A., Ságvári, B., Jakobi, Á., & Kertész, J. (2015). Geographies of an online social network. *PLOS One, 10*(9), e0137248.

Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology, 59*(1), 77–85.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., & Tomkins, A. (2005). Geographic routing in social networks. *Proceedings of the National Academy of Sciences, 102*(33), 11623–11628.

Maggioni, M. A., Nosvelli, M., & Uberti, T. E., (2007). Space versus networks in the geography of innovation: A European analysis. *Papers in Regional Science, 86*(3), 471–493.

Maggioni, M. A., & Uberti, T. E. (2009). Knowledge networks across Europe: which distance matters? *The Annals of Regional Science, 43*(3), 691-720.

Maggioni, M. A., Uberti, T. E., & Nosvelli, M. (2014). Does intentional mean hierarchical? Knowledge flows and innovative performance of European regions. *The Annals of Regional Science, 53*(2), 453-485.

Malerba, F. (2009). Increase learning, break knowledge lock-ins and foster dynamic complementarities: evolutionary and system perspectives on technology policy in industrial dynamics. In Foray, D. (ed.) *The New Economics of Technology Policy*. Cheltenham – Northampton: Edward Elgar, 33-45.

Martin, R., & Sunley, P. (2007). Complexity thinking and evolutionary economic geography. *Journal of Economic Geography, 7*(5), 573-601.

Phelps, C. C. (2010). A longitudinal study of the influence of alliance network structure and composition on firm exploratory innovation. *Academy of Management Journal, 53*(4), 890-913.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577–8582.

Quaglio, G., Millar, S., Pazour, M., Albrecht, V., Vondrak, T., Kwiek, M., & Schuch, K. (2020). *Exploring the performance gap in EU Framework Programmes between EU13 and EU15 Member States*. European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 641.542 – June 2020

Raffo, J., & Lhuillery, S. (2009). How to play the "Names Game": Patent retrieval comparing different heuristics. *Research Policy, 38*(10), 1617-1627.

Ramlogan, R., & Consoli, D. (2014). Dynamics of collaborative research medicine: the case of glaucoma. *The Journal of Technology Transfer, 39*(4), 544-566.

Reagans, R., & McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly, 48*(2), 240–267.

Rodríguez-Pose, A. (2020). The Research and Innovation Divide in the EU and its Economic Consequences. In European Commission: *Science, Research and Innovation Performance of the EU 2020. A Fair Green and Digital Europe*. Brussels: European Commission, 676-707.

Roediger-Schluga, T., & Barber, M. J. (2008). R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results. *International Journal of Foresight and Innovation Policy, 4*(3-4), 321-347.

Sah, P., Leu, S. T., Cross, P. C., Hudson, P. J., & Bansal, S. (2017). Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proceedings of the National Academy of Sciences, 114*(16) 4165-4170.

Scherngell, T., & Barber, M. J. (2009). Spatial interaction modelling of cross-region RD collaborations: empirical evidence from the 5th EU framework programme. *Papers in Regional Science, 88*(3), 531–546.

Scherngell, T., & Lata, R. (2013). Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. *Papers in Regional Science, 92*(3), 555-577.

Schoen, A., Heinisch, D., & Buenstorf, G. (2014). Playing the Name Game to identify academic patents in Germany. *Scientometrics, 101*(1), 527–545.

Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval. Vol. 39*. New York, NY: Cambridge University Press.

Seabright, M. A., Levinthal, D. A., & Fichman, M. (1992). Role of individual attachments in the dissolution of interorganizational relationships. *Academy of Management Journal, 35*(1), 122-160.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science, 51*(5), 756-770.

Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., & Ratti, C. (2013). Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLOS One, 8*(12), e81707.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy, 35*(7), 994–1017.

Ter Wal, A. L., & Boschma, R. A. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science, 43*(3), 739-756.

Ter Wal, A. L. J. (2013). The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. *Journal of Economic Geography, 14*(3), 589–620.

Torre, A., & Rallet, A. (2005). Proximity and localization. *Regional Studies, 39*(1), 47-59.

Usai, S. (2011). The geography of inventive activity in OECD regions. *Regional Studies, 45*(6), 711-731.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly, 42*(1), 35–67.

Varga, A., Pontikakis, D., & Chorafakis, G. (2014). Metropolitan Edison and cosmopolitan Pasteur? Agglomeration and interregional research network effects on European R&D productivity. *Journal of Economic Geography, 14*(2), 229–263.

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.

Whittle, A., Lengyel, B., & Kogler, D. F. (2020). Understanding Regional Branching – Knowledge Diversification via Inventor Collaboration Networks. *Papers in Evolutionary Economic Geography*, No. 2006.

Zollo, M., Reuer, J. J., & Singh, H. (2002). Interorganizational Routines and Performance in Strategic Alliances. *Organization Science, 13*(6), 701-713.

Appendix

Appendix Table 1. Rand index between alternative community structures of the full co-inventor network.

	2	3	4	5
1	0.703	0.722	0.737	0.741
2		0.658	0.671	0.673
3			0.685	0.687
4				0.702

Source: Own calculation based on OECD REGPAT 2015.

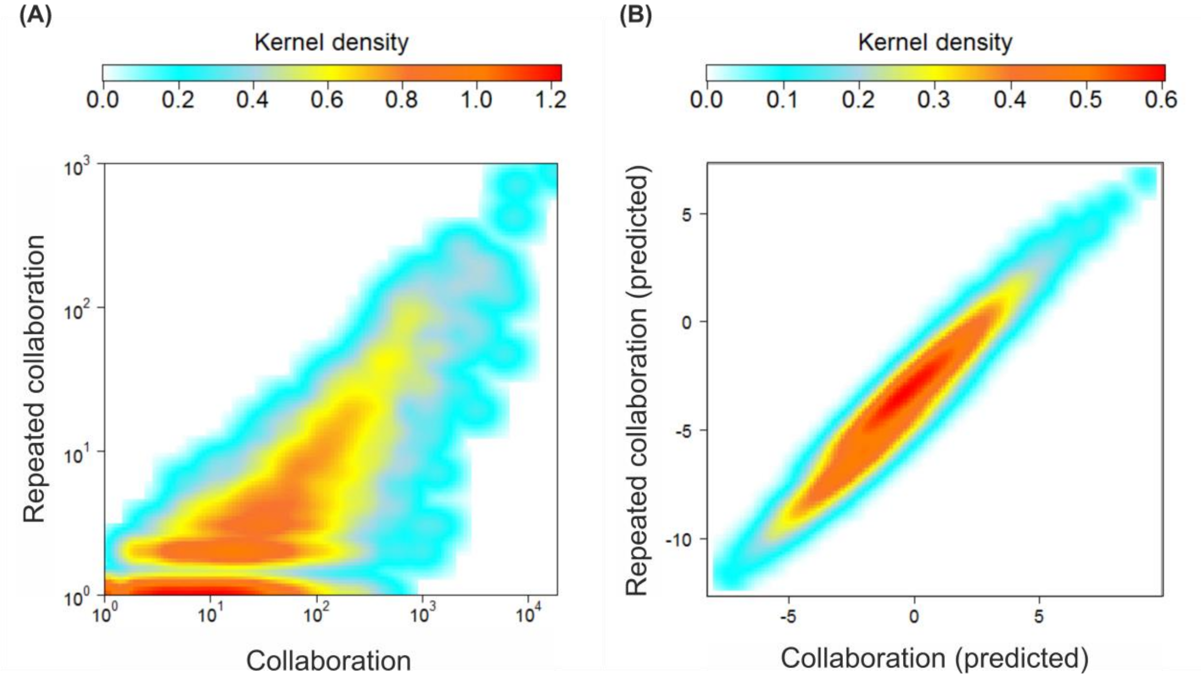
Appendix Table 2. Rand index between alternative community structures of the repeated co-inventor network.

	2	3	4	5
1	0.811	0.808	0.769	0.779
2		0.724	0.693	0.702
3			0.692	0.700
4				0.674

Source: Own calculation based on OECD REGPAT 2015.

Similar values of relative modularity would indicate that differences in community structures in the full- and in the repeated inter-regional co-inventor networks are not the artefact of differing network sizes and densities. The Louvain method includes heuristics in defining groups to speed up calculations; therefore, the algorithm needs to be run on the same network with randomly shuffled matrix representations. The community structure can then be considered correct if there is a strong overlap across identified communities in the randomly re-shuffled matrices.

Appendix Figure 1. Correlations of collaboration and repeated collaboration.



Source: Own calculation based on OECD REGPAT 2015.

Note: **(A)** Tie strength correlation between total collaboration and repeated collaboration ties. Despite the evident correlation, there is a relatively wide distribution of collaboration strength at certain levels of repeated tie strength. For example, for those region-region links that have 101 repeated ties: the minimum, mean and maximum values are around 101, 102 and 103. **(B)** Marginals estimated from the gravity equation of the strength of collaboration and repeated collaboration across regions of the ZINB regressions. The skewed distribution suggests that the gravity estimation narrows down the variance we observe in SI Fig1A. The two separate estimations, in which only the dependent variables and the Jaccard indexes differ, yield very similar predicted values of collaboration and repeated collaboration for region pairs. Therefore, if coefficients suggest different effects, one can argue for diverging mechanisms of establishing new inventor collaboration versus repeating old collaborations.

Appendix Table 3. Multivariate gravity models, zero-inflated negative binomial regression.

	(1) Collaboration	(2) Repeated Collaboration	(3) Repeated Collaboration (existed)	(4) Repeated Collaboration (subsample)	(5) Repeated Collaboration (inter-firm)
Main effects					
Cosine$_{ij}$	0.371***	0.409***	0.312***	0.349***	0.358***
	(0.027)	(0.042)	(0.034)	(0.045)	(0.033)
Proximity$_{ij}$	0.468***	0.847***	0.722***	0.704***	0.937***
	(0.019)	(0.037)	(0.052)	(0.038)	(0.046)
Jaccard$_{ij}$	0.867***	0.567***	0.556***	0.353***	0.470***
	(0.048)	(0.046)	(0.026)	(0.041)	(0.055)
Existed connections$_{ij}$			0.315***		
(1991-2005)			(0.058)		
Constant	-2.261***	-5.271***	-5.229***	-4.021***	-2.337***
	(0.141)	(0.250)	(0.200)	(0.173)	(0.076)
Zero-inflation					
Same country	-6.986***	-3.174***	-3.595***	-3.159***	-2.673***
(dummy variable)	(0.260)	(0.151)	(0.170)	(0.269)	(0.215)
log connections region *i*	-11.960***	-0.522**	-0.640***	-0.770**	-0.880**
	(1.180)	(0.204)	(0.217)	(0.322)	(0.133)
log connections region *j*	-13.890***	-0.801***	-0.981***	-1.112***	-0.072
	(0.750)	(0.133)	(0.145)	(0.206)	(0.089)
log inventors region *i*	3.877***	-3.371***	-3.606***	-4.380***	-2.384***
	(0.683)	(0.381)	(0.376)	(0.868)	(0.280)
log inventors region *j*	11.120***	1.100***	1.152***	2.394***	-1.368***
	(1.146)	(0.269)	(0.249)	(0.629)	(0.231)
log pop density region *i*	0.370***	0.045	0.066	0.137	0.070
	(0.105)	(0.066)	(0.039)	(0.116)	(0.082)
log pop density region *j*	0.710***	0.227***	0.223***	0.170*	0.060
	(0.078)	(0.041)	(0.044)	(0.091)	(0.068)
log GVA region *i*	-2.040***	0.345***	0.403***	0.962***	0.509***
	(0.119)	(0.131)	(0.131)	(0.270)	(0.140)
log GVA region *j*	-5.137***	-0.657***	-0.682***	-0.822***	-0.096
	(0.340)	(0.095)	(0.098)	(0.218)	(0.115)
Constant	61.010	12.320***	12.870***	7.248***	7.325***
	(.)	(0.718)	(0.753)	(0.912)	(0.748)
log Alpha	2.220***	2.160***	2.208***	2.028***	1.423***
Constant	(0.056)	(0.059)	(0.060)	(0.051)	(0.047)
P	0	0	0	0	0
log likelihood	-207048.8	-32905.8	-32595.4	-25100.9	-21389.6
N	872,235	872,235	872,235	64,055	59,958

Source: Own calculation based on OECD REGPAT 2015.

Note: that there is a missing standard error for constant term in the first model due to we set the maximum number of iteration to 100 for the estimation to avoid infinite convergence. Standard errors in parentheses; *p<0.10, **p<0.05, ***p<0.01.

Appendix Table 4. Gravity models with interaction terms, zero-inflated negative binomial regression.

	Collaboration	Repeated Collaboration	Collaboration	Repeated Collaboration
Binary and interaction effects				
$Cosine_{ij}$ (reference category: low)	1.420***	1.933***	0.940	1.075***
	(0.049)	(0.070)	(0.098)	(0.424)
$Proximity_{ij}$ (reference category: low)	1.574***	1.686***	0.394***	0.433
	(0.066)	(0.156)	(0.109)	(0.434)
$Jaccard_{ij}$ (reference category: low)	2.924***	2.931***	1.340***	2.342***
	(0.114)	(0.191)	(0.119)	(0.383)
$Proximity_{ij} \times Jaccard_{ij}$			1.067***	0.321
			(0.141)	(0.476)
$Proximity_{ij} \times Cosine_{ij}$			0.212	-0.444
			(0.144)	(0.553)
$Jaccard_{ij} \times Cosine_{ij}$			1.023***	-0.631
			(0.190)	(0.481)
$Jaccard_{ij} \times Cosine_{ij} \times Proximity_{ij}$			0.709***	2.034***
			(0.115)	(0.359)
P	0	0	0	0
log likelihood	-225777.2	-37693.3	-226004.8	-37645.1
N	872,235	872,235	872,235	872,235

Source: Own calculation based on OECD REGPAT 2015.

Note: Standard errors in parentheses; *p<0.10, **p<0.05, ***p<0.01.