Papers in Evolutionary Economic Geography

20.32

Explaining the dynamics of relatedness: the role of co-location and complexity
Sándor Juhász, Tom Broekel and Ron Boschma

Utrecht University   Human Geography and Planning

Explaining the dynamics of relatedness:
the role of co-location and complexity

Sándor Juhász,[1,2] Tom Broekel[3,4] and Ron Boschma[3,5]

[1] Agglomeration and Social Networks Lendület Research Group, Centre for Economic and Regional Studies, Hungarian Academy of Sciences, Budapest, Hungary
[2] Laboratory for Networks, Technology and Innovations, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Budapest, Hungary
[3] University of Stavanger Business School, Stavanger, Norway
[4] CRIE, Center of Regional and Innovation Economics, University of Bremen, Bremen, Germany
[5] Department of Human Geography and Planning, Faculty of Geosciences, Utrecht University, Utrecht, Netherlands

Abstract

Relatedness has become a key concept for studying the diversification of firms, regions, and countries. However, studies tend to treat relatedness as being time-invariant or, alternatively, consider its evolution as exogenously given. This study argues that relatedness is inherently dynamic and endogenous to technological and economic developments.

Using patent data, we test the extent to which relatedness between technologies developed along co-location and differences in technological complexity in the period 1980-2010. Our results show that co-located technologies are more likely to become related over time. Moreover, our results suggest that co-location and complexity of technologies are conducive to the intensification of relatedness over time.

Keywords: relatedness, geography, co-agglomeration, complexity
JEL codes: O33, R12

1. Introduction

The relatedness concept has drawn a lot of attention in Management Studies, Economic Geography, and Innovation Studies (Teece, Rumelt, Dosi, & Winter, 1994; Boschma, 2017; Hidalgo et al. 2018). Relatedness refers to the fact that two activities (technologies, industries, products, jobs) are not identical but share commonalities. Such commonalities may originate from two activities belonging to the same overarching technological or economic field, or they share complementarities and similarities (Broekel & Brachert, 2015; Nooteboom, 2000). There is ample evidence that relatedness of technologies conditions whether a technology will be invented or adopted by a firm, region or country. Moreover, relatedness has been shown to influence performance and diversification of firms (Fornahl, Broekel, & Boschma, 2011; Neffke, Henning, & Boschma, 2012), regions (Frenken, van Oort, & Verburg, 2007; Neffke, Henning, & Boschma, 2011), and nations (Hidalgo, Klinger, Barabási & Hausmann, 2007). Recently, relatedness has also found its way into policy as an evaluation and benchmarking dimension (Balland, Boschma, Crespo, & Rigby, 2019; Fitjar, Benneworth, & Asheim, 2019).

However, few studies have analyzed relatedness as dependent variable, and how relatedness changes and evolves over time. In fact, in most of the empirical studies, relatedness is either treated as time-invariant, or as an exogenously given component. While one might argue that relatedness is rather stable in the short run, this is less likely to be the case in the long-run (Broekel, 2015; Balland, Boschma, & Frenken, 2015; Boschma, 2017). Consequently, so far, we know little about how (technological) relatedness comes into existence, and how it develops over time (Menzel, 2015). This gap motivates the present paper studying the emergence and evolution of technological relatedness. More precisely, we evaluate two dimensions' relations with the emergence of technological relatedness and its intensification over time: co-location and complexity. Making use of patent data, our empirical findings suggest that co-location of technology pairs support the emergence of relatedness in Europe during the period 1980-2010. Moreover, both co-location and the level of complexity are correlated to the intensification of relatedness over time.

The paper is organized as follows. Section 2 provides a short introduction to the concept of relatedness and its possible interplay with co-location and complexity. Section 3 introduces the data, key variables, and modelling approaches. Section 4 presents and discusses the empirical findings, while Section 5 contains several robustness checks. The paper concludes with a discussion that outlines limitations, implications, and possibilities for future research.

2. Dynamics of relatedness

Relatedness has become a key input to outline possible technological and economic re-combination and diversification opportunities in particular. There is a consensus that the probabilities of firms, regions and countries to enter new and specific activities is a function of the number of the related activities they are specialized in (Neffke, Henning, & Boschma, 2011; Boschma, Balland, & Kogler, 2015; Hidalgo et al. 2018). Similarly, regions and countries with access to related variety tend to outperform economically those that are either highly specialized or overly diversified (Frenken, van Oort, & Verburg, 2007; Hidalgo, Klinger, Barabási, & Hausmann 2007; Fornahl, Broekel, & Boschma, 2011). Thereby, relatedness refers to two activities (e.g., technologies, industries, occupations, research fields) being based on similar underlying knowledge, skills or other inputs (Hidalgo et al. 2018; Boschma, 2017). The present paper focuses on technological relatedness, which is defined as the degree to which two technologies are proximate in a technological or cognitive dimension (Breschi, Lissoni, & Malerba, 2003; Kogler, Rigby, & Tucker, 2013; Boschma, Balland, & Kogler, 2015; Rigby, 2015).

In the quickly growing literature using (technological) relatedness as key explanatory variable, most studies put little emphasis on relatedness as an evolving property. In fact, we know surprisingly little about how relatedness comes into existence, and how it develops over time (Balland, Boschma, & Frenken, 2015; Menzel, 2015). In most studies on (related) diversification, relatedness is being treated as time-invariant or exogenously given. For instance, relatedness is frequently added to the empirical models as independent variable that is constant over the observed time period (e.g. Breschi, Lissoni, & Malerba, 2003; Frenken, van Oort, & Verburg, 2007; Neffke, Henning, & Boschma, 2011). This assumption seems reasonable when considering short time-periods, but it is much less likely to be the case over the long run (Cowan, Jonard, & Zimmermann, 2007; Balland, Boschma, & Frenken, 2015; Boschma 2017), and in particular when technological paradigms shift (Dosi 1982) and thereby reshuffle the technology space (Rigby, 2015). This is exemplified by the rise of electronics and biotech industries in recent years (Krafft, Quatraro, & Saviotti, 2011; 2014).

Notably, in a number of studies, relatedness is modeled as being time-variant, i.e. it may change between time periods (e.g., Boschma, Balland, & Kogler, 2015; Rigby, 2015; Mewes & Broekel, 2020). However, none of these studies systematically discuss how relatedness emerges and evolves over time. Nevertheless, there are several ideas that give a hint about what factors and conditions might be important for the emergence and development of relatedness.

The relevance of geographical proximity is frequently highlighted to enhance knowledge spillovers between economic agents (Jaffe, Trajtenberg, & Henderson, 1993; Audretsch &

Feldman, 1996). Geographical co-location of actors facilitates localized learning processes and lays the foundation for technological (re-)combinations. Consequently, co-located actors with expertise in different technologies are more likely to engage in knowledge exchange. It increases the likelihood that they learn from each other, which may eventually (in some instances) even lead to the (re-)combination of distinct technological knowledge. In the latter case, their geographical proximity may have helped in growing the relatedness of their respective fields of expertise (technologies). These arguments fit well those of Jane Jacobs (1969), who argues that the diversity of ideas, knowledge, and technologies in cities (i.e., within the geographic vicinity) enables and triggers cross-fertilization and knowledge re-combination. Put differently, the co-location of experts with specific knowledge on technologies increases the chances that connections may be discovered between previously unrelated technologies and that complementarities are identified, which in turn create or intensify relatedness (Broekel & Brachert, 2015; Desrochers & Leppälä, 2011; Boschma, 2017). This is summarized in the following two hypotheses:

H1A: Co-location enhances the <u>emergence</u> of relatedness between technologies.
H1B: Co-location enhances the <u>intensification</u> of relatedness between technologies.

In addition to geographical proximity, we argue that another dimension which is of relevance in this context is complexity, as it is likely to influence knowledge combinatorial processes. The complexity of a technology is often understood as a function of the number of its (sub-)components and their interdependencies (Simon, 1962; Fleming & Sorenson, 2001). Combining two technologies requires a basic understanding and the mastering of both. Gaining such an understanding is more difficult the more complex technologies are (Hidalgo & Hausmann, 2009; Balland & Rigby, 2017; Fleming & Sorenson, 2001; Sorenson & Fleming, 2004). When combining multiple technologies, these difficulties will scale up with the number of technologies.

It can also be expected that it is more difficult to combine two complex technologies than it is to combine a simple technology with a complex one. One reason is that to combine multiple technologies, mastery in all involved is required, implying that the necessary learning efforts are cumulative. Another reason is that the result of the combinatorial activity, i.e. an innovation or new technology, is most probably at least as complex as the most complex technology it is based upon. Consequently, combining two technologies and mastering the result thereof, go beyond the sum of the efforts required to obtain mastery in its constitutive technological elements. Significantly, this will even hold when the combinatorial process is conducted by a team. Members of the team must have a basic understanding of the technologies involved, which, as argued above, is more difficult in the cases of more complex technologies. Moreover, complex technologies are made of heterogeneous and relatively unrelated (sub-)components implying they are characterized by a greater knowledge diversity

(Broekel, 2019). This implies that larger cognitive distances need to be bridged within the team. In turn, it makes coordination, communication and learning harder, which translates into greater efforts and higher costs of collaboration (Gross & McMullen, 1982; Lundvall, 1993; von Hippel, 1987). Consequently, due to its 'cost-and-efforts' driving nature, the complexity of technologies will have a negative impact on the likelihood and frequency of the technologies being combined.

Yet, this relation might be countered by economic reasons. Technologies have a value that reflects their supply and demand. Technologies that are simple to learn can diffuse easily between economic agents through re-invention, reverse engineering, or copying. Hence, they offer little potential for competitive advantage and consequently they have relatively little economic value (Maskell & Malmberg, 1999). Concomitantly, technologies that are more difficult to replicate and imitate – which particularly applies to complex technologies – will not diffuse easily. Complex technologies therefore offer a high potential for competitive advantage and thus, a higher level of socio-economic value (Fleming & Sorenson, 2001). In addition, Hidalgo and Hausmann (2009) argue that complex technologies are crucial stepping stones for acquiring additional technological competences in the future. This adds further to their value.

Accordingly, while complex technologies and their combinations are more likely to be associated with higher economic rewards (Dalmazzo, 2002; Hidalgo & Hausmann, 2009), they are also characterized by greater difficulties in being understood and combined. Considering this, we expect complexity to act as an (initial) barrier to technological combinations and thereby to the emergence of relatedness. When two technologies have not previously been combined, the (economic) potential of their combination is unknown, which will lower the actors' willingness and likelihood to explore it. This will change once this barrier has been overcome and the hidden potential becomes clearer. In this case, the potentially higher social-economic benefits associated with combinations of complex technologies should attract the interest of the actors and motivate them to invest. Consequently, there will be greater exploration of such combinations and quicker convergence of the two technologies in terms of their relatedness.

In sum, complexity represents a dimension of technologies that is likely to impact the likelihood and frequency of their (re-)combination and consequently to shape the evolution of their relatedness. Our second set of hypotheses summarizes this:

H2A: More complex technologies are less likely to become related.
H2B: Once they have become related, the relatedness of two complex technologies is more likely to intensify than that between two simple technologies

As pointed out above, despite the great interest in relatedness and complexity, little is known about their relation and their relation over time. The following empirical investigations seeks to shed some light on this for the first time.

3. Empirical approach
3.1 Dataset and relatedness of technology pairs

Following much of the literature, we rely on patent data as our primary data source (Breschi, Lissoni, & Malerba, 2003). More precisely, we use the OECD REGPAT dataset (version 2018) covering patents registered by the European Patent Office (EPO). It contains detailed information on patents' application year, technology classes, inventors and inventor locations since 1976. The pros and cons of patent data have been discussed extensively, so we refrain from this and refer to the relevant literature (Grilliches, 1990; Desrochers, 1998).

We focus on the period 1976-2010, as this period offers reliable patent information. Figure 1A illustrates the growth of the number of patent applications during this period. To track the evolution of technological relatedness, we constructed seven non-overlapping 5-year time periods from 1976 to 2010, with the first covering the years of 1976-1980, the second 1981-1985, and so on. By pooling the data of multiple periods, we increase the stability of our measures, as patent numbers tend to fluctuate strongly between years. This is particularly the case when they are aggregated at the regional or technological level (Buerger, Broekel, & Coad, 2012).

To empirically represent technological relatedness, we first need to define the technologies. We follow existing studies and consider technologies being represented by four-digit CPC classes (Balland, Boschma, Crespo, & Rigby, 2019; Broekel, 2019). We exclude technologies starting with the letter *Y,* indicating cross-technological patents[1], i.e. patents that could not be clearly assigned to other classes. This leaves 645 distinct technologies. Figure 1B illustrates that almost all of these technologies contributed patents to the data set since 1980. Accordingly, four-digit CPC classes provide a consistent representation of technologies over this period.

There are many ways in which technological proximity can be estimated. Many studies use the information on technologies jointly appearing on patents, so-called co-occurrences, as an indication of combinatorial innovation processes and technological distance (Breschi, Lissoni, & Malerba, 2003). We follow this literature and concentrate on the joint appearance of technologies (four-digit CPC) classes on patents.

---

[1] The 'Y' CPC class is only a technical category with significant overlap to other classes.

The 645 technologies translate into 207,690 potential technological combinations (excluding self-combinations), which serve as units of observation. For each of these pairs, we count the number of occurrences on patent documents within each of the seven time periods with no geographical restriction (including all the patents in OECD REGPAT from all over the world). More precisely, we count how frequently different four-digit CPC classes are combined on patents during the given period, thus abstracting from the same combinations of four-digit classes on the same patent for the purpose of simplification.

Figure 1C visualizes the number of distinct technological combinations for each five-year period. Their number is increasing linearly over time. This growth is both in absolute and relative terms. The share of realized combinations of technologies (appearing on at least one patent) on the potential number of combinations grows from around 5% to 18%. This can be explained by the strongly growing numbers of patents and the more or less constant number of technologies (four-digit CPC classes). Figure 1D highlights there are few technology combinations that appear in high frequencies (>100 times), while the vast majority are realized in significantly smaller numbers (<10).

In general, each patent represents a possibility of two technologies co-occurring in the data. Consequently, the likelihood of two technologies jointly appearing on patents is primarily determined by the size of the two technologies, i.e. by the number of patents that are assigned to these technologies (jointly and independently). In the literature, measures of relatedness are often constructed in such a way that they account for this size effect. For instance, some studies employ the Cosine index (Breschi, Lissoni, & Malerba, 2003; Ejermo 2003) or alternative measures of co-occurrence standardization (Kogler, Rigby, & Tucker, 2013; Mewes, 2019). However, specifying the correct standardization / normalization is far from straightforward (van Eck & Waltman, 2009).

We are not so much interested in which technologies are significantly related in statistical terms, and which are not. We simply seek to approximate technological distance between two technologies in a continuous manner. Therefore, we will account for this size effect in the later regression analysis, following roughly the idea of Neffke (2009) and Neffke, Henning, and Boschma (2011).
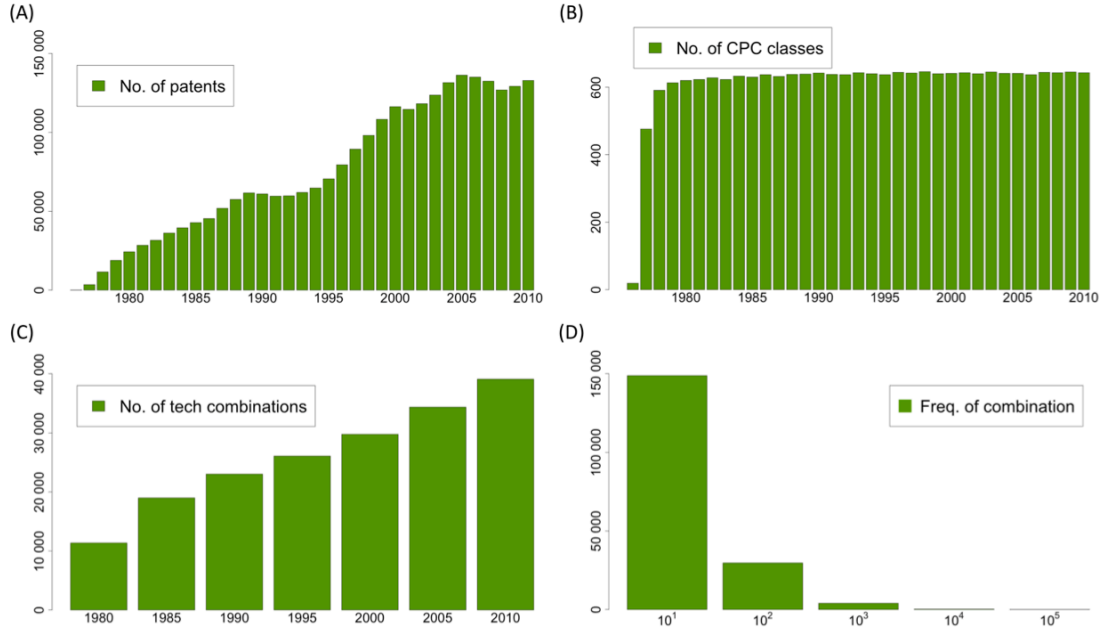
Fig1 (A) Number of patents in each year, 1976-2010

Fig1 (B) Number of different four-digit CPC technology classes that appear on patents in each year, 1976-2010

Fig1 (C) Number of observed distinct technological combinations. We measure technologies as four-digit CPC technology classes and by combination we count the co-occurrence of 2 technologies on patents in the selected 5-year long period.

Fig1 (D) Distribution of combination frequencies in the full sample over all the periods. For this chart, we excluded combinations with 0 frequency. Note that the x axis is on a log10 scale.

Source: Authors' own construction based on OECD REGPAT 2018

Note: Figures are based on the entire dataset with no geographical restrictions.

3.2 Co-location of technologies

As previously discussed, we expect technologies to become more related when they frequently co-locate. Empirically, this corresponds to the overlap of their spatial distributions. Empirically, we first assign patents to the NUTS2 regions of their inventors' addresses. Accordingly, patents with multiple inventors are assigned to multiple regions, without counting fractions. Next, we obtain the spatial distribution of a technology's patents, by counting the number of its patents assigned to each region. On this basis, we assess the overlap in two technologies' distributions with the technology pair level co-agglomeration measure proposed by Ellison, Glaeser and Kerr (2010). For two technologies *i* and *j*, the measure looks as follows:

$$COAGGLOM_{i,j} = \frac{\sum_{r=1}^{R}(s_{ir}-x_r)(s_{jr}-x_r)}{1-\sum_{r=1}^{R}x_r^2}.$$
(1)

whereby P_{ir} is the number of patents in technology i in region r and $s_{ir} = \frac{P_{ir}}{\sum_{r\prime=1}^{R} P_{ir\prime}}$ is the share of technology i in region r (and for j accordingly). x_r is the mean of these shares in region r across all technologies. The index has been used in multiple studies investigating industrial co-agglomeration (Ellison, Glaeser, & Kerr, 2010; Diodato, Neffke, & O'Clery, 2018). Its main advantage is that it is largely invariant to the distribution of inventors, applicants, and the type of spatial units considered (Diodato, Neffke, & O'Clery, 2018).

We calculate the co-location of the 207,690 technology pairs in all seven periods. To reduce the noise in the data, we set the patent counts to zero for all regions with less than ten patents in a particular technology. Moreover, we exclusively consider the 316 EU and EFTA NUTS-2 regions with sufficiently large patent numbers and the patents associated to these[2]. To make the co-location score comparable across technologies, we calculate its z-score. Detailed descriptive statistics of the *COAGGLOM* variable can be found in Table 1.

3.3 Technological complexity

Empirically approximating technological complexity is a challenge, as the literature offering few approaches that are applicable to the data at hand. Balland and Rigby (2017) transferred the idea of the Economic Complexity index of Hidalgo and Hausmann (2009) to patent data and calculated an index of technological complexity. In essence, the method utilizes information on the extent to which technologies are ubiquitous, since few regions specialize in them, and how diversified regions are; how many technologies do they specialize in? By means of an iterative procedure, the index sets these two aspects into relation and measures the complexity of technologies by their tendency of co-concentration with other ubiquitous technologies.

Earlier, Fleming and Sorenson (2001) employed an N/K model to derive a technological complexity index based on the frequencies with which patent subclasses co-occur on patents. The underlying argument is that more complex combinations are harder to realize and thereby less likely to be observed. The difficulty of (re-)combining two classes is empirically assessed by evaluating the currently observed co-occurrence of patent subclasses on patents in relation to the cumulative frequency of their co-occurrence in previous years.

While both approaches have their merits, Broekel (2019) argues that some of the core assumptions underlying these measures are problematic. Moreover, he shows that both

---

[2] Regions belonging to the following countries are considered: Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

indices fail in resembling several stylized facts commonly associated with technological complexity. However, more importantly, in our analyses both indices carry the danger of being very closely related to key variables in the model. The complexity index of Balland and Rigby (2017) is based on spatial co-location of technologies, which is at the heart of the measure of co-agglomeration. Fleming and Sorenson's (2001) approach utilizes the frequency of patent class co-occurrences on patents as a central ingredient, which is the basis of our technological relatedness measure. To avoid the potential of (artificial) empirical overlap, we turn to an alternative measure of complexity that does not raise these concerns.

We adopt the structural complexity index developed by Broekel (2019). It measures technological complexity by modeling technologies (four-digit CPC classes) as combinatorial networks of (knowledge) components. While in this, co-occurrence information of patent classes on patents is used as well, in this case, these networks are dichotomized eliminating large portions of the frequency information, which is central in our measure of technological distance. Broekel (2019) argues that the combinatorial networks of complex technologies are characterized by a greater diversity of (sub-)network topologies. Network science shows that a greater diversity implies a larger information content of these networks. Consequently, it can be argued that this demands greater R&D efforts for the invention of such technologies and also represents greater obstacles to them being learned or copied. Both aspects make complex technologies more exclusive and therefore potentially more economically valuable (Broekel 2019).

To capture this diversity, Broekel (2019) proposes the *network diversity score* of Emmert-Streib and Dehmer (2012), as empirical approximation of *structural diversity* of technologies i.e. as empirical measure of technological proximity. This score is estimated as follows. For each technology *T* (four-digit CPC class), its combinatorial network is constructed in period *t*, by extracting all patents with at least one patent subclass (ten-digit CPC subclass) belonging to technology *T*. Next, the co-occurrence matrix of all patent subclasses appearing on these patents is created and dichotomized with all positive entries being one, and all others remaining zero. This matrix represents the combinatorial network of technology *T* in which complexity is assessed by means of the *network diversity score* (NDS). In the estimation of the NDS, multiple subsamples are drawn from the networks' main component. For each subsample *i*, the share of modules (α_{module}), the variability of module sizes (v_{module}), the variability of the Laplacian matrix ($V_{Laplacian}$), and the relation of graphlets of size three and four ($r_{graphlets}$) are calculated and the individual *NDS* score is estimated as:

$$iNDS_i = \frac{\alpha_{module} * r_{graphlets}}{V_{Laplacian} * V_{module}}$$

10

The *iNDS* is subsequently averaged over sample networks giving the NDS for this technology's combinatorial network. It is log-transformed and multiplied by -1 to obtain the final complexity measure of *structural diversity*, which signals higher complexity with larger values. For a more detailed introduction to the measure see Broekel (2019).

In practice, we use the values of *structural complexity* given in Broekel (2019).[3] These are annual complexity values for each of the 645 four-digit CPC classes. As structural complexity scores are relatively stable over short time (Broekel 2019), we used the complexity values of technologies given at the end of each period.

As discussed in Section 2, we expect complexity to influence the likelihood and the frequency of two technologies being combined, i.e. co-occurring on patents. Crucially, this influence may be negative, as it is more difficult to combine complex technologies. However, complexity may also have a positive impact by signaling the potential of greater rewards associated to the combination of two complex technologies. In any case, it is less the individual complexity values of technologies which matter most, but rather their joint configuration. We model this with a variable representing the sum of structural complexity values of two technologies (*COMPLEX_SUM*). Figure 2A gives a first impression on this by visualizing how, potentially, all technology pairs' complexity values potentially align. Figure 2B then shows how these values are actually distributed across the observed (co-occurring) technology pairs. The figures suggest that most of the realized combinations are pairs with mid- or higher-value of complexity, which supports our theoretical argumentation of complexity positively correlating to the (economic and technological) potential of combinations of technologies. To control for the possibility that the sum of two technologies' complexity may be relatively high as a result of a combination of a highly complex and a rather simple technology, we also include the absolute difference of two technologies' complexity values (*COMPLEX_ABSDIFF*) in the models. Descriptive statistics on both complexity variables can be found in Table 1.

---

[3] The data can be downloaded via: https://doi.org/10.1371/journal.pone.0216856.s007.
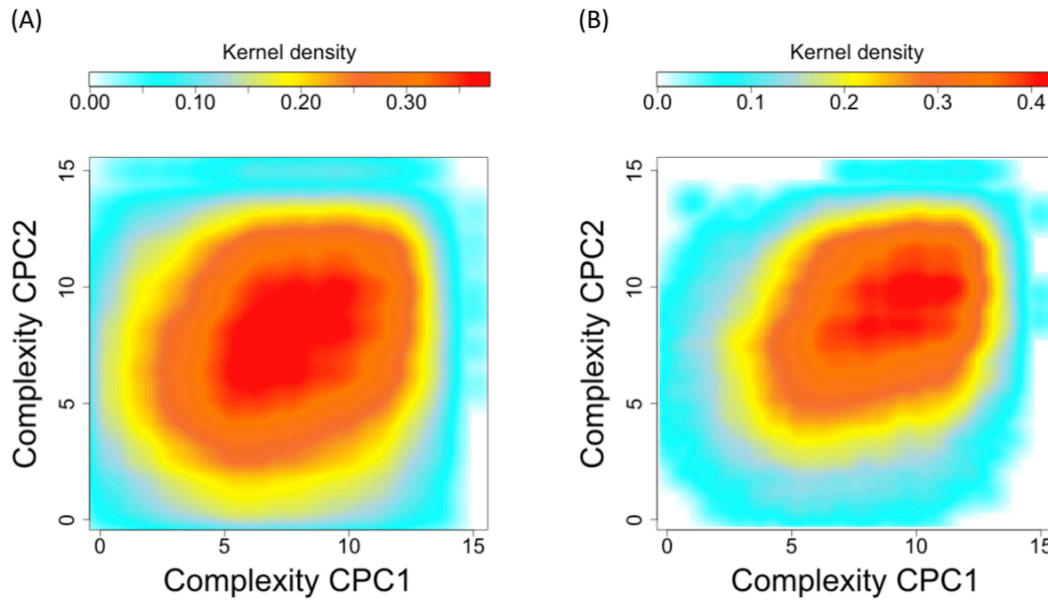
Fig2 (A) The density of all the possible technological combinations on a complexity-complexity plot (1980-2010)
Fig2 (B) The density of all observed technological combinations on a complexity-complexity plot (1980-2010)
Source: Authors' own construction based on OECD REGPAT 2018
Note: For the construction of these charts we do not split the dataset into periods.

3.5 Empirical model

We seek to explain the level (and change) of technological relatedness in our baseline models. Accordingly, the units of analysis are technology pairs and their numbers of co-occurrences are the dependent variable.[4] To account for the fact that large technologies have more possibilities of co-occurring on patents, we follow the idea of Neffke (2009) and Neffke, Henning, and Boschma (2011) by estimating a model for (revealed) relatedness and include the absolute numbers of patents that are assigned to either of the two technologies. In other words, the number of co-occurrences that are predicted by the sizes of the two technologies can be seen as the baseline number of expected co-occurrences given the size of two technologies. Further explanatory variables are added to the model that are either technology specific (at the level of technologies) or technology pair specific (i.e. at the level of technology relations). These variables seek to explain the observed deviations from this baseline number, i.e. they give insights into why some technological combinations are observed more or less frequently than what can be expected by just considering the numbers of patents of two technologies. This particularly refers to the degree of geographical co-location and complexity.

---

[4] The approach is very similar to the frequently employed gravity regression model (Broekel, Balland, Burger, & van Oort, 2014; Gómez-Herrera, 2012; Scherngell & Barber, 2009).

As we are interested in both the emergence and intensification of relatedness between technology pairs, we use two different estimation strategies. First, to observe the relationship between the first appearance of a technological combination and the level of co-location and the level of technological complexity, we rely on logit models similar to the entry models frequently used in economic geography and related fields to study diversification processes (Boschma, Balland, & Kogler, 2015; Balland, Boschma, Crespo, & Rigby, 2019). For this exercise we filter the dataset to those technology pairs that are not present in every period and become related at some point during the observed time. In other words, we restricted observations with zero relatedness in every period and higher than zero relatedness in every period, which results an unbalanced panel setting.

Our model specification for the logit regression is the following:

$$\Pr\left(y_{ijt} = 1\right) = \beta_0 + \beta_1 COAGGLOM_{ijt} + \beta_2 COMPLEX\_SUM_{ijt} +$$
$$\beta_3 COMPLEX\_ABSDIFF_{ijt} + \beta_4 \log(N_{it}) + \beta_5 \log(N_{jt}) + \lambda_t + \eta_t \tag{3},$$

where our dependent variable y_{ijt} takes the value of zero when there is no observed co-occurrence of technology *i* and *j* in period *t* and in *t+1*, and one in case there is no observed combination in *t* but there is one in *t+1*. *COAGGLOM, COMPLEX_SUM* and *COMPLEX_ABSDIFF* denote the characteristics of the technology pair *i* and *j* in period *t*. N_{it} and N_{jt} are the number of patents in technology *i* and *j* at time *t* as in gravity model specifications. λ_t is a technology fields effects constructed by the concatenation of the first letter of the two technology codes combined and η_t is a period fixed effect. Detailed descriptive statistics regarding this specific model setting in comparison to the entire dataset is represented in Table 1. Correlation between the main variables are presented in the Appendix (Table 4).

Table 1. Descriptive statistics behind the entire dataset

Variable	Number of observations	Mean	Min	Max	Standard Deviation
	Entire dataset				
RELATEDNESS	1,453,830	1.912	0	13,373	39.450
COAGGLOMERATION	1,453,830	-0.006	-3.874	35.943	0.952
COMPLEXITY_SUM	1,453,830	14.275	0	28.158	4.659
COMPLEXITY_ABSDIFF	1,453,830	2.973	0	14.983	2.390
	Logit dataset				
RELATEDNESS (0/1)	206,976	0.255	0	1	0.436
COAGGLOMERATION	206,976	0.046	-3.874	26.422	1.010
COMPLEXITY_SUM	206,976	14.344	0	28.118	4.082
COMPLEXITY_ABSDIFF	206,976	2.537	0	13.642	1.973
	Growth models				
RELATEDNESS (growth)	143,206	13.542	-4.143	3.922	0.750
COAGGLOMERATION	143,206	0.168	-3.874	32.148	0.938
COMPLEXITY_SUM	143,206	16.623	0	28.118	3.791
COMPLEXITY_ABSDIFF	143,206	2.164	0	13.406	1.685

Source: Authors' own construction based on OECD REGPAT 2018
Note: The entire dataset contains all the possible CPC technology combinations for all the periods. For a few technology classes the complexity measure is not available.

Second, we explore the relationship between the changes in the strength of technological relatedness and the co-location and complexity of technologies. To do so we use panel OLS models on the growth rates of technological relatedness. For this second exercise our focus is only on technology pairs with higher than zero combinations at time *t*.

For the growth models, we use the corresponding specification:

$$Growth\ RELATEDNESS\ (y_{ijt}) = \gamma_0 + \gamma_1 \log(y_{ijt}) + \gamma_2 COAGGLOM_{ijt} + \gamma_3 COMPLEX\_SUM_{ijt} + \beta_4 COMPLEX\_ABSDIFF_{ijt} + \gamma_5 \log(N_{it}) + \gamma_6 \log(N_{jt}) + \tau_t + \delta_t$$

(4),

where y_{ijt} is the rate of relatedness growth from *t* to *t+1*. $\log(y_{ijt})$ is the level of relatedness at time t, *COAGGLOM, COMPLEX_SUM* and *COMPLEX_ABSDIFF* denote the characteristics of the technology pair *i* and *j* in period *t*. N_{it} and N_{jt} are the number of patents in technology *i* and *j* at time *t* as in gravity model specifications. τ_t is a technology fields effects constructed by the concatenation of the first letter of the two technology codes combined and δ_t is a period fixed effect. We applied several further estimations to improve the robustness of our results. We summarize these robustness checks in Section 5.

4. Results

Table 2 presents the coefficients of four logit regression models described in Equation (3). Control variables are included in all models and the central explanatory variables are introduced in a stepwise manner. The two size variables, the number of patents of the considered technologies (No. Patents CPC1, CPC2) are significantly positive and their coefficients are relatively large in all four models. The finding reflects that technologies characterized by many patents have more opportunities to co-occur on patents.

Table 2. Models on the entry of new relatedness ties

	ENTRY MODELS			
	(1)	(2)	(3)	(4)
COAGGLOM	0.038***			0.037***
	(0.006)			(0.006)
COMPLEX_SUM		-0.001		-0.003
		(0.002)		(0.002)
COMPLEX_ABSDIFF			-0.034***	-0.034***
			(0.004)	(0.005)
No. Patents CPC1	0.165***	0.166***	0.161***	0.163***
	(0.005)	(0.005)	(0.005)	(0.005)
No. Patents CPC2	0.169***	0.170***	0.165***	0.167***
	(0.005)	(0.005)	(0.005)	(0.005)
Constant	-3.698***	-3.699***	-3.557***	-3.552***
	(0.051)	(0.051)	(0.052)	(0.052)
Technology fields FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Observations	206 976	206 976	206 976	206 976
Log Likelihood	-99 051	-99 074	-99 009	-98 987
AIC	198 193	198 239	198 109	198 069

Note: The regressions are based on all the combinations that appear at least once during the 1980-2010 period and not observed in each period. 'Technology fields FE' refers to a control variable constructed by the concatenation of the first letter of the two technology codes combined. We also clustered standard errors by these groups. Standard errors in parentheses, *p<0.1; **p<0.05; ***p<0.01

The first key variable is *COAGGLOM* approximating the extent to which concentrations of technology 1 tend to appear in the same regions as technology 2. The positive and significant coefficient for *COAGGLOM* in model (1) suggests that the more two technologies are co-located, the more likely they are related. In model (2) *COMPLEX_SUM* is introduced and its coefficient is insignificant. Consequently, the complexity of two technologies does not have explanatory power of whether they will become related over time. In model (3) *COMPLEX_ABSDIFF* is added. In contrast to COMPLEX_SUM its coefficient is significant, but negative. Accordingly, the more two technologies differ in their levels of complexity, the less likely it is that they will become related. Model (4) contains all of our main explanatory variables. It shows that their signs and significances do not change when they are simultaneously included. These findings support our first hypothesis *H1A*: geographical co-location of technologies facilitates the emergence of relatedness. In contrast, Model (4)

suggests rejecting our hypothesis *H2A*: complexity does not seem to be crucial for the emergence of technological relatedness. What is striking about the results is that the two significant coefficients COAGGLOM and COMPLEX_ABSDIFF are of almost similar magnitude, implying that both dimensions are of somewhat equal empirical importance in this respect, albeit co-agglomeration being of somewhat higher relevance.

The second set of models focuses on the explanation of the intensification of relatedness, i.e. its development after the first connection between technologies has been made. Table 3 represents the results of the growth model specified in Equation (4). Again, we include our main variables step-by-step from model (5) to model (8). Generally, our (few) control variables show the expected signs. The two size variables, the coefficients of the number of patents of the considered technologies (No. Patents CPC1, CPC2) are significantly positive and are relatively large. Hence, technologies characterized by many and increasing patent numbers do occur more frequently on patents.

Table 3. Models on the intensification of relatedness

	Growth of RELATEDNESS			
	(5)	(6)	(7)	(8)
RELATEDNESS	0.044***	0.041***	0.040***	0.038***
	(0.002)	(0.002)	(0.002)	(0.002)
COAGGLOM		0.031***	0.030***	0.030***
		(0.002)	(0.002)	(0.002)
COMPLEX_SUM			0.008***	0.008***
			(0.001)	(0.001)
COMPLEX_ABSDIFF				-0.012***
				(0.001)
No. Patents CPC1	0.069***	0.070***	0.063***	0.062***
	(0.002)	(0.002)	(0.002)	(0.002)
No. Patents CPC2	0.066***	0.068***	0.061***	0.060***
	(0.002)	(0.002)	(0.002)	(0.002)
Technology fields FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Observations	143 206	143 206	143 206	143 206
R^2	0.047	0.049	0.049	0.050

Note: The dependent variables are growth rates of relatedness. Hausman tests supported the use of fixed effect models. 'Technology fields FE' refers to a control variable constructed by the concatenation of the first letter of the two technology codes combined. We also clustered standard errors by these groups. Standard errors in parentheses, * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Model (5) confirms that the intensification of relatedness is positively linked to the previous level of technological relatedness, which is also the strongest predictor in all model settings. Model (6) adds *COAGGLOM* and demonstrates that the co-location of technologies correlates to growing relatedness. This finding adds empirical support to the hypothesis *H1B* that the relatedness between technologies becomes stronger in case they are strongly co-located. In Model (7), *COMPLEX_SUM* is added and model (8) *COMPLEX_ABSDIFF*. The significantly positive coefficient of *COMPLEX_SUM* suggests that the complexity of two technologies

significantly correlates to the growth of their relatedness. This holds even after controlling for their differences in terms of their complexity (*COMPELX_ABSDIFF*). It thereby shows that technologies at the upper end of the complexity spectrum are more likely to intensify their relatedness than those at the lower end. The finding backs hypothesis *H2B* and provides some support for our argument that the larger the economic rewards associated to the combination of complex technologies, in contrast to less complex ones, lead to more activities being (successfully) undertaken in that respect.

5. Robustness checks

We applied a variety of robustness checks to validate our results. As the emergence of relations between technologies is a binary variable, logit regression with a large number of fixed effects with few time periods can lead to biased results because of the prominent incidental parameters problem (Neyman & Scott, 1948; Mewes & Broekel, 2020). Therefore, we re-estimate the binary (entry) regression models as linear probability model (LPM). The results are shown in Table 5 in the Appendix. By and large, they support our findings.

Our logistic models only focus on the emergence of relatedness (first co-occurrence). To validate the findings, we also estimate models for the disappearance of technological relatedness, i.e. no more co-occurrences (see Appendix Table 6). To identify this disappearance event, we exclude all technology pairs with strictly positive co-occurrence counts in all periods (*RELATEDNESS* > 0) and those that never occur (*RELATEDNESS* = 0) during the examined periods. The disappearing dummy is then constructed such that it is one if in any of the periods we observe a positive number of co-occurrences and simultaneously, in a subsequent period, none. In these models, we find the coefficient of *COAGGLOM* to be significantly positive. However, its level of significant is smaller than in the case of the first-co-occurrence model. More importantly, the coefficient of *COMPLEX_SUM* is significantly negative. This suggests that the more complex two technologies are, the less likely it is that their combinations will disappear. This is in line with our arguments that once complex technologies are combined and their relatedness is discovered, they represent economically valuable technology combinations that tend to be used continuously in the future.

In an alternative specification, we combine the two models (emergence and intensification of relatedness) into a Zero-Inflated Negative Binomial (ZINB) regressions model, which addresses the issue of the access of zeros in our dependent variable (Figure 1C). The result of the ZINB regression are shown in Table 7 in the Appendix. As expected, they confirm the earlier findings.

Even though our main variables and estimations are based on different (temporal) subsets of the patent database, endogeneity is a crucial concern. Frequently, this is addressed by means of an instrumental variable regression. However, our empirical observations are based on technology pairs, which complicates the identification of proper instrumental variables. Nevertheless, to gain a better understanding of the potential causal directions of the observed relations, we follow the approach of Broekel (2015) who applies an empirical strategy in a similar setting that is popular in the firm growth literature (see, e.g., Coad, 2009). In Table 8 in the Appendix, we transform our central variables (*RELATEDNESS*, *COAGGLOM*, *COMPLEX_SUM* and *COMPLEX_ABSDIFF*) into annual growth rates and use these as explanatory and dependent variables in different OLS regression settings. More precisely, we regress onto a focal variable's growth in t, its level in t, and the growth rates and the respective levels of all the other variables in t. We find that while *RELATEDNESS* explains *COAGGLOM* and both complexity variables in all models, co-location and the complexity of technology pairs seems to be not linked in most models. Notably, none of the models contradicts our previous findings, which highligh their robustness. However, they contain some additional insights, which we will address in more detail in the concluding section.

6. Discussion

This paper seeks to explore the evolution of technological relatedness over time and to identify what factors contribute to this. Thereby, we shed light on an issue that received little attention so far in studies on relatedness in economic geography and innovation studies: relatedness is usually treated as independent, almost exogenous factor shaping the co-location of technologies (e.g., Kogler, Rigby, & Tucker, 2013; Rigby 2015), rather than a endogenous structure that requires explanation in itself.

We argue that geographical co-concentration of technologies provides mechanisms that impact the evolution of their relatedness. We provide empirical support for this and show that the more the spatial distributions of two technologies overlap, the more likely they become related, and the more this relatedness tends to grow in the future. This suggests that mechanisms such as local knowledge spillovers enhance the emergence and intensification of technological relatedness. This finding is important for the interpretation of outcomes in studies on related diversification, which suggest an impact of relatedness on the location of industrial, technological, and economic activities (Neffke, Henning, & Boschma, 2011; Kogler, Rigby, & Tucker, 2013). The present study emphasizes that this process is most likely bi-directional: co-location drives the development of relatedness, and vice versa. Knowledge spillovers stimulated by co-location contribute to (re-)combinatorial invention processes, which facilitate technologies converging in terms of relatedness.

A somewhat different role seems to be played by technological complexity in this context. In contrast to geographic co-location, it we do not find indications of its influence on the emergence of relatedness. However, once complementarity between technologies has been achieved (i.e. a minimum level of relatedness has been reached), we show that complex technologies are more likely to be combined subsequently. We interpret this is to be driven by the higher economic incentives associated to the combination of complex technologies.

Our paper has several implications for future research. Firstly, while our study makes a first step to increase our understanding of how technological relatedness emerges and evolves over time, we need more research to fully comprehend the evolution of technologies and their relations over time (Desrochers & Leppälä 2011; Castaldi, Frenken, & Los, 2015). This includes gaining more knowledge about where and when re-combinatorial processes create new patterns of relatedness and what type of benefits are related to this (Pinheiro et al. 2018).

Secondly, our study has shown that differences in technologies' levels of complexity can act as barriers to the development of technological relatedness, while two highly complex technologies are more likely to intensify their level of relatedness over time as compared to two less complex ones. Still, these are early insights requiring further validation in future research. For instance, it might be the case that simple and complex technologies represent different underlying knowledge infrastructures and human capital, which restrict their complementarity and consequently their integration. In this sense, complexity might reflect fundamental differences between technologies. Put differently, in the extremes, it might be another way of identifying unrelatedness. This certainly calls for additional research.

Thirdly, our study points towards endogenous relations shaping the interplay of relatedness, co-location, and complexity of technologies. For instance, our results indicate that the relation between co-location and relatedness is bi-directional: co-location is linked to the emergence and intensification of relatedness, and vice versa. Our study thereby complements the rich literature on relatedness that has identified relatedness as being a crucial driver behind the spatial distribution of industries and technologies (e.g. Hidalgo, Klinger, Barabási & Hausmann, 2007; Neffke, Henning, & Boschma, 2011; Rigby, 2015). Our results suggest that this relationship is more complex and characterized by the existence of co-evolutionary dynamics between the spatial concentration, complexity, and relatedness of different technologies. This has hardly been addressed in the literature. However, the empirical approaches employed do not provide causal evidence on which dimension is driving which. To establish this, future research needs a more elaborate approach utilizing causal modelling strategies. While we have made an attempt to get first insights into the magnitude and direction of these relations, we are far from disentangling the interplay of complexity, co-location, and relatedness. Undoubtedly, such endeavors are crucial, as all these dimensions are applied and used to an

increasing extent in current regional innovation policies (Balland, Boschma, Crespo, & Rigby, 2019; Fitjar, Benneworth, & Asheim, 2019).

Finally, while the paper provides insights on the evolution of relatedness and explored the role of several factors therein, it leaves a universe of other dimensions unexplored. The consideration of other factors such as the role of institutions or policy interventions will surely add to our understanding of the underlying mechanisms. Moreover, our empirical study relies exclusively on patent data. Despite the many advantages of this approach, it is undoubtedly advisable to repeat and extend the analysis using other types of data in the future (Boschma, 2017).

References

Audretsch, D., & Feldman, M. P. (1996). R&D Spillovers and the Geography of Innovation and Production. *American Economic Review*, 18 (1), 1-28.

Balland, P. A., Boschma, R. & Frenken, K. (2015). Proximity and Innovation: From Statics to Dynamics. *Regional Studies*, 49 (6), 907-920.

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

Balland, P. A., Boschma, R., Crespo, J. & Rigby, D. (2019). Smart Specialization Policy in the European Union: Relatedness, Knowledge Complexity and Regional Diversification. *Regional Studies*, 53 (9), 1252-1268.

Boschma, R. (2017). Relatedness as driver behind 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.

Breschi, S., Lissoni, F. & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy*, 32 (1), 69-87.

Broekel, T. (2015). The Co-evolution of Proximities – A Network Level Study. *Regional Studies*, 49 (6), 921-935.

Broekel, T. (2019). Using structural diversity to measure the complexity of technologies. *PLOS One*, 14 (5), ee0216856.

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

Broekel, T, & Brachert, M. (2015). The structure and evolution of inter-sectoral technological complementarity in R&D in Germany from 1990 to 2011. Journal of Evolutionary Economics, 25(4), 755–785.

Buerger, M., Broekel, T. & Coad, A. (2012). Regional dynamics of innovation – Investigating the co-evolution of patents, R&D, and employment. *Regional Studies*, 46(5), 565–582.

Castaldi, C., Frenken, K. & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies*, 49 (5), 767-781.

Coad, A. (2009). *The growth of firms: A survey of theories and empirical evidence*. Edward Elgar, Cheltenham, UK.

Cowan, R., Jonard, N. & Zimmermann, J.B. (2007). Bilateral collaboration and the emergence of innovation networks, *Management Science*, 53, 1051–1067.

Dalmazzo, A. (2002). Technological complexity, wage differentials and unemployment. *Scandinavian Journal of Economics*, 104(4), 515-530.

Desrochers, P. (1998). On the Abuse of Patents as Economic Indicators. *The Quarterly Journal of Austrian Economics*, 1(4), 51–74.

Desrochers P. & Leppälä S. (2011). Opening up the 'Jacobs spillovers' black box: local diversity, creativity and the processes underlying new combinations. *Journal of Economic Geography* 11, 843–863.

Diodato, D., Neffke, F. & O'Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics*, 1-26. doi: 10.1016/j.jue.2018.05.022

Dosi, G. (1982). Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change. *Research Policy,* 11, 147–162.

Ejermo, O. (2003). Patent diversity as a predictor of regional innovativeness in Sweden. CESPRI Working Paper, 140.

Ellison, G., Glaeser, E. L. & Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100 (3), 1193-1213.

Emmert-Streib, F. & Dehmer, M. (2012). Exploring statistical and population aspects of network complexity. *PLoS ONE*, 7(5).

Fitjar, R. D., Benneworth, P., & Asheim, B. T. (2019). Towards regional responsible research and innovation? Integrating RRI and RIS3 in European innovation policy. *Science and Public Policy*, 1–12.

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

Fornahl, D., Broekel, T. & Boschma, T. (2011). What drives patent performance of German biotech firms? The impact of R&D subsidies, knowledge networks and their location. *Papers in Regional Science*, 90 (2), 395-418.

Frenken, K., Van Oort, F. & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41 (5), 685-697.

Gómez-Herrera, E. (2012). Comparing alternative methods to estimate gravity models of bilateral trade. *Empirical Economics*, 44(3), 1087–1111.

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

Gross, A. E., & McMullen, P. A. (1982). The Help-Seeking Process. In V. J. Derlega & J. Grezlak (Eds.), *Cooperation and Helping Behavior*. Academic Press, New York.

Hidalgo, C. A. (2009). The dynamics of economic complexity and the product space over a 42 year period. *CID Working Papers*, 189.

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

Hidalgo, C. A. & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106 (26), 10570-10575.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A, Neffke, F., Rigby, D., Stern, S., Zheng, S. & Zhu, S. (2018).

The Principle of Relatedness. *International Conference on Complex Systems*, Springer, Cham, 451-457.

Jacobs, J. (1969). *The Economy of Cities*. Vintage, New York, NY.

Jaffe, A.B., Trajtenberg, M. & Henderson, R. (1993). Geographic localization and knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* 108, 577–598.

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

Krafft, J., Quatraro, F. & Saviotti, P.P. (2011). The knowledge-base evolution in biotechnology: a social network analysis. *Economics of Innovation and New Technology* 20(5), 445-475.

Krafft, J., Quatraro, F. & Saviotti, P.P. (2014). The dynamics of knowledge-intensive sectors' knowledge base. Evidence from biotechnology and telecommunications. *Industry and Innovation* 21(3), 215-242.

Lundvall, B. A. (1993). Explaining Interfirm Cooperation and Innovation. Limits of the Transaction-cost Approach. In G. Grabher (Ed.), *The embedded firm*. Routledge, London.

Maskell, P., Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics* 23 (2), 167-186.

Menzel, M.P. (2015). Interrelating Dynamic Proximities by Bridging, Reducing and Producing Distances. *Regional Studies* 49(11), 1892-1907.

Mewes, L. (2019). Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836-2010. *Economic Geography*, doi: 10.1080/00130095.2019.1567261

Mewes, L., & Broekel, T. (2020). Subsidized to change ? The impact of R & D policy on regional technological diversification. *The Annals of Regional Science*, https://doi.org/10.1007/s00168-020-00981-9.

Neffke, F. (2009). *Productive Places: The influence of technological change and relatedness on agglomeration externalities*. Doctoral Thesis.

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.

Neffke, F., Henning, M. & Boschma, R. (2012). The impact of aging and technological relatedness on agglomeration externalities: a survival analysis. *Journal of Economic Geography*, 12 (2), 485-517.

Nooteboom, B. (2000). *Learning and Innovation in Organizations and Economies*. Oxford University Press, Oxford.

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R. & Hidalgo, C. (2018). Shooting Low or High: Do Countries Benefit from Entering Unrelated Activities? *Papers in Evolutionary Economic Geography*, 18 (7).

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. & Barber, M. J. (2009). Spatial interaction modeling of cross-region R&D collaboration. Empirical Evidence from the 5th EU Framework Programme. *Papers in Regional Science*, 88(3), 531–546.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106 (6), 467-482.

Sorenson, O. & Fleming, L. (2004). Science and the diffusion of knowledge. *Research Policy*, 33 (10), 1615-1634.

Teece, D. J., Rumelt, R., Dosi, G. & Winter, S. G. (1994). Understanding corporate coherence: theory and evidence. *Journal of Economic Behaviour Organisation*, 23, 1-30.

van Eck, N. J. & Waltman, L. (2009). How to normalize co-occurence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635–1651.

von Hippel, E. (1987). Cooperation between Rivals: Informal Know-How Trading. *Research Policy*, 16(6), 291–302.

Appendix

Table 4. Correlation matrix of the most important variables

	RELATEDNESS	COAGGLOM	COMPLEX_SUM	COMPLEX_ABSDIFF
RELATEDNESS	1			
COAGGLOM	0.019	1		
COMPLEX_SUM	0.065	0.004	1	
COMPLEX_ABSDIFF	.0.033	-0.043	-0.250	1

Source: Authors' own construction based on OECD REGPAT 2018

Note: For the construction of the correlation matrix we do not split the dataset to periods.

Table 5. Entry models in linear probability settings

	ENTRY MODELS – linear probability setting			
	(1)	(2)	(3)	(4)
COAGGLOM	0.006***			0.006***
	(0.001)			(0.001)
COMPLEXITY SUM		-0.000		-0.001*
		(0.000)		(0.000)
COMPLEXITY ABSDIFF			-0.005***	-0.005***
			(0.000)	(0.000)
No. Patents CPC1	0.023***	0.023***	0.022***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
No. Patents CPC2	0.019***	0.020***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-0.109***	-0.109***	-0.089***	-0.087***
	(0.005)	(0.005)	(0. 005)	(0. 006)
Period FE	Yes	Yes	Yes	Yes
Observations	206 976	206 976	206 976	206 976
R^2	0.172	0.172	0.172	0.172

Note: The regressions are based on all the combinations that appear at least once during the 1980-2010 period and not observed in each period. Standard errors in parentheses, *p<0.1; **p<0.05; ***p<0.01

Table 6. Exit models in logit setting

	EXIT MODELS – logit setting			
	(1)	(2)	(3)	(4)
COAGGLOM	0.019**			0.020**
	(0.008)			(0.008)
COMPLEXITY SUM		-0.028***		-0.030***
		(0.003)		(0.003)
COMPLEXITY ABSDIFF			-0.010***	-0.014***
			(0.004)	(0.004)
No. Patents CPC1	0.039***	0.068***	0.037***	0.067***
	(0.006)	(0.006)	(0.006)	(0.006)
No. Patents CPC2	0.003	0.028***	0.001	0.027***
	(0.005)	(0.006)	(0.006)	(0.006)
Constant	-19.763	-19.706	-19.717	-19.649
	(59.827)	(59.801)	(59.8295)	(59.792)
Period FE	Yes	Yes	Yes	Yes
Observations	172 313	172 313	172 313	172 313
Log Likelihood	-56 682	-56 637	-56 682	-56 628
AIC	113 385	113 294	113 384	113 280

Note: The regressions are based on all the combinations that appear at least once during the 1980-2010 period and not observed in each period. Standard errors in parentheses, *p<0.1; **p<0.05; ***p<0.01


Table 7. Results of the multivariate gravity models, zero-inflation negative binomial regressions (ZINB) on the strength of technological relatedness

	Count model	**Zero-inflation model**
	RELATEDNESS in t+1	RELATEDNESS in t+1
COAGGLOM (z-score)	0.507***	-0.017***
	(0.007)	(0.004)
COMPLEX_SUM	0.091***	0.068***
	(0.001)	(0.002)
COMPLEX_ABSDIFF	-0.196***	-0.006
	(0.002)	(0.003)
NO. PATS CPC1 log	0.424***	-0.811***
	(0.004)	(0.006)
NO. PATS CPC2 log	0.354***	-0.697***
	(0.004)	(0.005)
Constant	-4.008***	6.487***
	(0.029)	(0.040)
Period FE	Yes	Yes
Log-likelihood	-84 030	
Observations	1 092 636	

Standard errors in parentheses, *p<0.1; **p<0.05; ***p<0.01

Table 8. OLS models on the growth rate of different dependent variables

	Growth RELATEDNESS (1)	Growth COAGGLOM (2)	Growth COMPLEX_SUM (3)	Growth COMPLEX_ABS (4)	Growth RELATEDNESS (5)	Growth COAGGLOM (6)	Growth COMPLEX_SUM (7)	Growth COMPLEX_ABS (8)
Growth RELATEDNESS t-1					-0.014***	-0.013***	0.002***	0.004**
					(0.004)	(0.002)	(0.000)	(0.002)
Growth COAGGLOM t-1					0.090*	-0.020	-0.000	0.007***
					(0.054)	(0.036)	(0.000)	(0.002)
Growth COMPLEX_SUM t-1							0.121***	
							(0.030)	
Growth COMPLEX_ABSDIFF t-1					0.053***	0.020***	0.010***	
					(0.010)	(0.007)	(0.001)	
RELATEDNESS in t	0.059***	0.035***	0.005***	-0.017***	0.050***	0.033***	0.002***	-0.013***
	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
COAGGLOM in t	0.035***	-0.270***	0.002***	-0.013***	0.040***	-0.250***	0.003***	-0.013***
	(0.002)	(0.002)	(0.000)	(0.001)	(0.003)	(0.002)	(0.000)	(0.002)
COMPELX_SUM in t	0.008***	0.002***	-0.022***	0.000	0.009***	0.003***	-0.016***	0.001**
	(0.001)	(0.001)	(0.000)	(0.001)	(0.003)	(0.002)	(0.000)	(0.002)
COMPLEX_ABSDIFF in t	-0.012***	-0.006***	0.005***	-0.135***	-0.013***	-0.007***	0.003***	-0.108***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Constant	-0.405***	0.455***	1.351***	1.881***	-0.825***	0.376***	0.206***	0.436***
	(0.014)	(0.010)	(0.003)	(0.010)	(0.026)	(0.017)	(0.002)	(0.015)
No. Patents CPC1, CPC2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	143 206	138 581	143 206	143 206	126 098	127 829	126 098	126 098
R^2	0.051	0.177	0.226	0.170	0.036	0.149	0.243	0.156
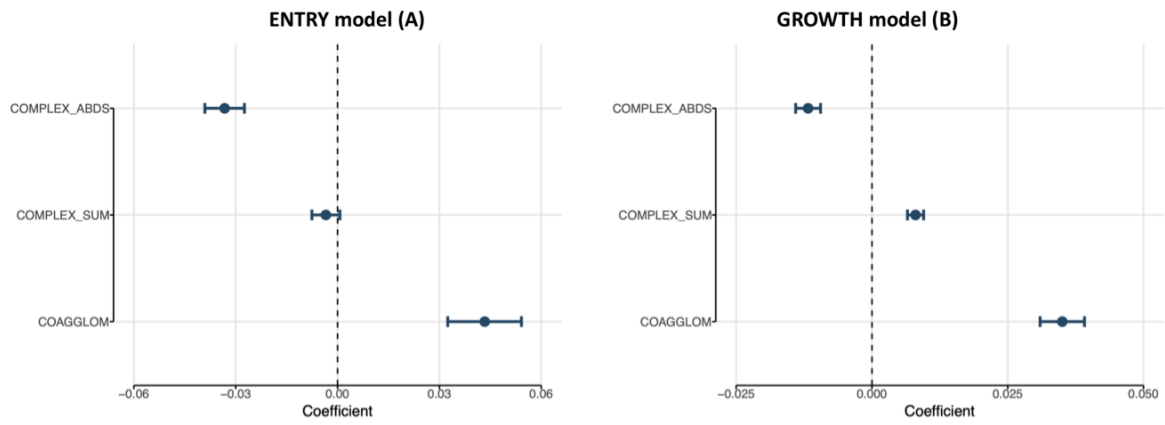
Fig3 (A) Visual representation of the influence of our main variables on the entry of relatedness from Model (4) in Table 2.

Fig3 (B) Visual representation of the influence of our main variables on the intensification of relatedness from Model (8) in Table 3.