Papers in Evolutionary Economic Geography

21.18

Technological breakthroughs in European regions: the role of related and unrelated combinations

Ron Boschma, Ernest Miguélez, Rosina Moreno & Diego B. Ocampo-Corrales

Utrecht University Human Geography and Planning

Technological breakthroughs in European regions: the role of related and unrelated combinations

Ron Boschma
Department of Human Geography and Spatial Planning, Utrecht University
UiS Business School, University of Stavanger
r.a.boschma@uu.nl (corresponding author)

Ernest Miguélez
GREThA, UMR CNRS 5113 – Université de Bordeaux

AQR-IREA, University of Barcelona
ernest.miguelez@u-bordeaux.fr

Rosina Moreno
AQR-IREA, University of Barcelona
rmoreno@ub.edu

Diego B. Ocampo-Corrales
AQR-IREA, University of Barcelona
diegob_ocampo@ub.edu

18 May 2021

Abstract

This paper analyzes if the emergence and occurrence of breakthrough technologies in 277 European regions in the period 1981 to 2010 is related to the existing technological portfolio of regions. The study shows that, by far, most combinations breakthrough inventions make are between related technologies: almost no breakthrough patent makes combinations between unrelated combinations only. We also find that breakthrough inventions primarily combine and cite technological classes that are present in the region. Regressions show that the occurrence of breakthrough patents in a technology in a region is positively affected by the local stock of technologies that is related to such technology, but we do not find such an effect for the local stock of unrelated technologies, in contrast to studies that suggest otherwise. However, the region's ability to enter new breakthrough inventions in a technology relies on the combination of knowledge that is both related and unrelated to such technology.

Keywords: relatedness, unrelatedness, technological breakthroughs, regional diversification, European regions

JEL codes: O18, O31, O33, R11

1. Introduction

Research in Economics of Innovation has sought to identify the sources of breakthrough technologies, as such breakthroughs are considered to have a large impact on subsequent technological change, and to achieve high economic value (e.g. Carnabuci and Operti 2013; Arts and Veugelers 2015; WIPO 2015; Verhoeven et al. 2016). There is a growing body of literature that has looked into the geography of such breakthrough technologies (e.g. O'hUallichain 1999; Varga 2000; Carlino et al. 2007; Kerr 2010; Balland and Rigby 2017; Castaldi and Los 2017; Berkes and Gaetani 2020; Li 2020; De Noni and Belussi 2021). This literature has been heavily influenced by the work of Jacobs (1969), claiming that regions with a diverse pool of knowledge trigger new ideas that result in breakthrough inventions that make atypical combinations (Bettencourt et al. 2007; Desrochers and Leppälä 2011; Mewes 2019; Abbasiharofteh et al. 2020).

Scholars have argued that it is not necessarily variety in regions *per se*, but unrelated variety that enhances the occurrence of technological breakthroughs (Frenken et al. 2007; Castaldi et al. 2015; Martynovich and Taalbi 2020). Unrelated variety in regions would enhance the occurrence of radical inventions because they make combinations across unrelated knowledge fields (Saviotti and Frenken 2008). As unrelated combinations involve low cognitive proximity between the technologies combined, geographical proximity might be a factor contributing to the likelihood of such uncommon combinations (Li et al. 2021). In other words, these combinations are more likely to occur and become more successful when available in the same region (Phene et al. 2006; Kelley et al. 2013).

However, there are still a few issues that require further attention. First, empirical studies indicate it is yet unclear whether related variety has an effect on the occurrence of breakthrough technologies in regions, besides unrelated variety. While Martynovich and Taalbi (2020) found a negative effect and Castaldi et al. (2015) found no effect of related variety, Miguelez and Moreno (2018) and Hesse and Fornahl (2020) found a positive effect. Second, these findings tend to challenge the implicit assumption that breakthroughs primarily make unrelated combinations. They suggest that new activities in regions, even the more radical ones, build on and combine both related and unrelated capabilities (Boschma 2017). Systematic evidence is lacking. Third, these studies did not examine whether the unrelated technologies in the region were actually combined in technological breakthroughs. So, what has hardly been researched to our knowledge is the extent to which breakthrough inventions also build on related combinations, besides unrelated combinations. Because if they would, it would mean local search processes still play a key role in the emergence of breakthrough technologies (Kaplan and Vakili 2015), and the regional presence of related technologies might still be a factor that affects their emergence and occurrence (Boschma 2017).

This paper aims to address these gaps, based on an empirical study on breakthrough patents in 277 European regions in the period from 1981 to 2010. The first objective is to analyze the extent to which breakthrough patents exploit related combinations versus unrelated combinations. Our study shows that, by far, most combinations breakthrough patents make is between related technologies: almost no breakthrough patent makes combinations between unrelated combinations only. The second objective is to show whether the production of breakthrough patents is shaped by the existing knowledge base of regions. Doing so, we test the implicit assumption that breakthrough patents in a region primarily combine and cite technologies that are present in the region. We find strong support for that. The third objective is to test econometrically whether the ability of regions to produce breakthrough patents in a technology is conditioned by the stock of local technologies related and unrelated to that technology. Here, our study deviates from previous studies (Castaldi et al. 2015; Miguélez and Moreno 2018) that measured these effects in terms of related and unrelated variety, because these studies do not provide insights on whether the local technologies in the variety measures are actually related or unrelated to the technology of the breakthrough inventions. Inspired by the regional diversification literature (Boschma 2017), we develop instead measures of relatedness and unrelatedness density that capture how close and how distant the technology of a breakthrough invention actually is to the existing portfolio of technologies in a region. We find evidence for the effect of relatedness density on the occurrence of breakthrough inventions in regions but no effect of unrelatedness density, no matter whether breakthrough patents are defined as highly cited or as combinations of two technologies for the first time. The fourth objective is to examine the effects of relatedness and unrelatedness density on the ability of regions to enter and develop breakthrough patents in technologies not yet present in the region. We find that the regional entry of breakthroughs in a technology for the first time is enhanced by a local stock of both related and unrelated technologies.

The paper is structured as follows. Section 2 presents a literature review. Section 3 introduces the data and explains how we define and measure breakthrough patents, and relatedness and unrelatedness density. Section 4 shows descriptive statistics concerning the type of combinations breakthrough patents make, and the extent to which they rely on local knowledge. Section 5 presents estimations explaining the share of breakthrough patents in regions in Europe as well as the entry of breakthroughs in technologies in a region. Section 6 concludes.

2. Technological breakthroughs, unrelated variety, and new combinations

A large body of literature shows that breakthrough inventions tend to concentrate in large cities, due to the presence of talent and high-skilled people, an advanced knowledge infrastructure, and high-complex knowledge (O'hUallichain 1999; Varga 2000; Carlino et al. 2007; Balland and Rigby 2017; Castaldi and Los 2017). This tendency of concentration of technological breakthroughs in the largest cities has accelerated since the early twentieth century (Mewes 2019), although breakthroughs occur also outside large cities (Fritsch and Wyrwich 2021). Jacobs (1969) argued that a diverse pool of knowledge in large cities would trigger new ideas and innovations because it provides opportunities to explore new knowledge combinations (Bettencourt et al. 2007; Desrochers and Leppälä 2011).

Building on Jacobs' idea, scholars have claimed that different types of regional variety favor different types of inventions (Saviotti and Frenken 2008; Castaldi et al. 2015; Miguélez and Moreno 2018). Castaldi et al. (2015) argued that related variety in a region would favor inventions in general, because related technologies can be more easily and effectively combined (Tavassoli and Carbonara 2014; Aarstad et al. 2016; Miguélez and Moreno 2018; Martynovich and Taalbi 2020). When local agents search for new combinations, they focus on pieces of knowledge in their immediate surroundings they have prior experience in, and look into combinations that have been combined before, in order to reduce uncertainty and lower adjustment costs (Nelson and Winter 1982; Nooteboom 2000). Therefore, unrelated variety would slow down inventions in a region because the cognitive distance between technologies would be too large, and therefore, it would be too risky and too costly to combine those. However, unrelated variety would enhance the occurrence of radical inventions as they make combinations across unrelated knowledge fields (Saviotti and Frenken 2008). Castaldi et al. (2015) found indeed a positive correlation between unrelated variety and technological breakthroughs at the level of US states, but no correlation with related variety. Martynovich and Taalbi (2020) found for Swedish regions a positive effect of unrelated variety and negative effect of related variety on radical innovations.

However, Miguelez and Moreno (2018) found that also variety in related knowledge domains in a region, and not only variety in unrelated knowledge domains, enhance breakthrough inventions. A similar observation was made by Hesse and Fornahl (2020). They found that the local presence of related and unrelated knowledge domains supported radical inventions in German regions, but especially related variety. De Noni and Belussi (2021) found that the likelihood of a breakthrough in an industry in a region increased when related industries are present in the region. These findings tend to challenge the implicit assumption that breakthroughs primarily build on unrelated combinations. They suggest that new activities in regions, even the more radical ones, may combine both related and unrelated capabilities (Boschma 2017). Moreover, studies did not examine whether the unrelated technologies in the region were actually combined in technological breakthroughs. So, what has not yet been researched is the extent to which breakthroughs rely on unrelated combinations, and to what extent they also rely on related combinations. This requires a thorough investigation of what types of combinations technological breakthroughs make.

In recombinant search theory (Weitzmann 1998), radical inventions are considered to make combinations across existing knowledge domains not combined before (Ahuja and Lampert 2001; Schoenmakers and Duysters 2010). Recombining knowledge from distant technology fields are perceived to result in novel and more valuable inventions (Trajtenberg et al. 1997; Fleming 2001; Dahlin and Behrens 2005). Arts and Veugelers (2015) argued that reusing more familiar, well-understood components foster breakthrough inventions but only when the familiar components are recombined in novel ways for the first time. Distant search is needed to avoid the familiarity trap (March 1991), and new combinations of familiar components make this happen (Arts and Veugelers 2015). However, this recombinant view has also been challenged. Kaplan and Vakili (2015) found evidence that breakthrough patents that represent novel topics are more likely to be associated with local, not distant search. Local search is needed to identify anomalies that requires a deep understanding of a particular knowledge domain. According to this view, recombinations of distant (diverse) knowledge are regarded as detrimental to breakthrough inventions (Weisberg 1999; Taylor and Greve 2006; Kaplan and Vakili 2015). This debate shows there is a strong need to identify whether breakthroughs rely on unrelated or related combinations, or a combination of both.

This also makes relevant the question whether breakthrough patents make combinations that are local or not from a geographical perspective. Studies show that new technologies tend to draw on local knowledge sources (Jaffe et al. 1993; O'hUallichain, 1999; Acs et al. 2002; Audretsch and Feldman 2004; Sonn and Storper 2008; Breschi and Lissoni 2009; Hervás-Oliver et al. 2018; Arant et al. 2019; Grashof et al. 2019). Scholars have argued that this applies especially to breakthrough patents, as these are assumed to stem from combinations of unrelated technologies. As unrelated combinations represent high cognitive distance between the combined technologies, these are facilitated by geographical proximity (Phene et al. 2006; Kelley et al. 2013). Li et al. (2021) investigated breakthrough inventions in solar photovoltaics and found evidence that unrelated technologies are more likely to be combined when present in the same region. However, unrelated knowledge might also be accessed through non-local sources (Boschma 2005; Miguélez and Moreno 2018). Hesse and Fornahl (2020) found indeed that breakthrough inventions are enhanced through inter-regional linkages. This begs the question whether technological breakthroughs depend on local combinations, and to what extent the related and unrelated combinations breakthrough patents make concern combinations of technologies inside the region.

3. Definitions, Method and Data

3.1 Breakthrough inventions

Most inventions are incremental in the sense that they improve and refine existing inventions whereas a few are breakthroughs. Despite being a minority, breakthroughs are considered the foundational inventions that serve extensively as the basis for many subsequent technological inventions and that introduce new solutions (Ahuja and Lampert 2001; Fleming 2001). Although the definition and measurement of breakthrough inventions can take into account different dimensions, the most common ones are novelty (break from the past) and large impact with respect to some technological or economic dimension.

Verhoeven et al (2016) operationalized the different meanings given to breakthroughs in terms of the ex-post technological impact of inventions and their ex-ante characteristics. As for the former, breakthrough inventions are viewed as introducing new paradigms on which many future inventions build (Dosi 1982; Fleming 2001). As for the latter, breakthroughs are defined in terms of their underlying technology in the sense that they incorporate technologies that move away from existing practices (Carlo et al. 2012).

Kaplan and Vakili (2015) took into account the cognitive dimension (novelty of idea) and the economic dimension (realization of value) of breakthroughs. A breakthrough introduces a potential new technological path that may, or may not, turn out to have an impact in terms of generating economic returns to the owners of the invention. Kaplan and Vakili (2015) showed that combining knowledge in an exploratory search (Gavetti and Levinthal 2000) tends to produce inventions that break from pre-existing technological modes and eventually become highly cited (Phene et al 2006). This way, the impact of an exploratory search on value occurs through the mechanism of novelty (Kaplan and Vakili 2015). The correlation between novelty of ideas and impact has been analysed in terms of scientific papers (Wang et al 2017; Uzzi et al 2013) and in terms of patents (Schoenmakers and Duysters 2010; Kim et al 2016; Verhoeven et al 2016). The latter concluded that the combination of previously disconnected components leads to novel ideas and high impact results. Similarly, patents recombining two technology subfields for the first time in history are more likely to become breakthroughs, as measured by forward citations (Fleming 2001).

In line with previous studies, we rely on patent data to analyze invention activities that result from recombination of previous pieces of knowledge (Arts and Veugelers 2015; Kim et al 2016). We define breakthrough inventions in two ways. First, we define breakthroughs from the ex-post technological impact of the invention, that is, as those patents with more forward citations, under the assumption that if a patent receives many citations, it should contain influential knowledge for the creation of new ideas (Jaffe and de Rassenfosse 2016). Studies have indeed shown that the number of forward citations received conveys information about the importance and economic value of patents (Trajtenberg 1990; Harhoff et al. 1999; Lanjouw and Schankerman 2004; Hall et al. 2005). We are aware that according to the year of release of a patent, it is subject to different period lengths to be cited, so that older patents have higher chance to be cited. To smoothen it, we classified patents according to their priority year. To control for the fact that some technologies are more dynamic in patenting and citing (Schmoch 2008), we calculate the citations in each technology, considering 35 technologies. We define a patent as a breakthrough when it is in the upper 10% of the distribution of number of forward citations for the technology and year it belongs to.[1] Second, we define breakthrough patents based on the ex-ante characteristics of inventions. A patent is considered a breakthrough when it combines two technological classes for the first time, following the recombinant search approach (Fleming 2001; Fleming et al. 2007). We do not expect our findings to differ between the two definitions, because studies have shown that the more a patent combines formerly unconnected technologies, the higher its impact in terms of forward citations (Arts and Veugelers 2015).

3.2 Measuring relatedness and unrelatedness

To determine the degree of (un)relatedness between technologies, we start by computing the co-occurrence of any two IPC classification codes (for a total of 621 four-digit technology classes) in the same patent document.[2] We employ the count of times any two technologies appear together in a patent (Breschi et al. 2003). To control for the fact that this co-occurrence can be random and caused by chance, we normalize it using the association probability measure presented in Eck and Waltman (2009). All in all, we consider a probabilistic measure resulting in a co-occurrence measure between any two technologies i and j (ϕ_{ij}) as:

$$\phi_{ij} = \frac{mc_{ij}}{s_i s_j} \tag{1}$$

[1] We also make robustness analyses with different thresholds, that is, the upper 5% and 1% of the distribution of forward citations.

[2] Taking the four-digit disaggregation of IPC classification codes, our dataset contains 640 technologies, but the 621 included in the analyses are those that are present for the whole period under consideration.

where c_{ij} denotes the number of times technologies i and j occur together in the same patent, s_i and s_j are the total number of times technologies i and j appear, and m is the total number of patents.[3] We compute a co-occurrence matrix for each of the time periods under consideration – see below.

For the measure of unrelatedness, we consider values of $\phi_{ij} \leq 1$ to indicate that technologies i and j are observed together as often as if they would co-occur by chance: they are considered to be unrelated. We obtain a matrix of unrelatedness for each time period of dimension 621x621, where each element ψ_{ij}=1 if $\phi_{ij} \leq 1$, and zero otherwise. As per relatedness, since those values of $\phi_{ij} > 1$ indicate that technologies i and j are observed together more frequently than would be expected by chance, we consider them to be related. We construct a matrix for each time period of dimension 621x621, where each element $\eta_{ij} = 1$ if $\phi_{ij} > 1$, and zero otherwise.

Based on the matrices above, we construct regional measures that will allow us to assess whether the development of a breakthrough invention in a region in a technology i is affected by the local presence of technologies related to that technology i (relatedness density) and the local presence of technologies unrelated to technology i (unrelatedness density). Recall that these two density measures are different from the entropy measures capturing related and unrelated variety that are commonly used in this literature (e.g. Castaldi et al. 2015; Miguélez and Moreno, 2018; Martynovich and Taalbi 2020). The latter measures are not fully satisfactory in our framework, as they do not make specific whether the local technologies in the variety measures are actually related or unrelated to the technology of the breakthrough inventions. Inspired by the regional diversification literature (Boschma 2017;

[3] Following van Eck and Waltman (2009), the use of a probabilistic similarity measure at the patent level is superior to other types of similarity measures based on co-occurrence. The co-occurrence of two objects can be driven by two independent effects: the similarity effect and the size effect. The similarity effect is the one in which two objects co-occur because they are related to each other. The size effect is the one in which a high frequency of co-occurrence of two objects can be due to the fact that one of them occurs a lot. van Eck and Waltman (2009) offer a detailed discussion of why our similarity measure remains unchanged when the occurrence of an object doubles, as well as the co-occurrences. This is not the case in other measures that fail to capture the size effect in co-occurrence. This indicator has been used before by Boschma et al. (2014) and Balland et al. (2019), among others.

Balland et al. 2019), we develop instead new measures of relatedness and unrelatedness density that capture how close and how distant the technology of a breakthrough invention actually is to the existing portfolio of technologies in a region. These density measures will be used to analyze whether the ability of regions to produce breakthrough patents in a technology is conditioned by the local stock of technologies related or unrelated to that technology, or a combination of both.

To develop relatedness and unrelatedness density measures, we need to construct variables for each set region-technology that indicate how close a technology is to the existing technological base of a region, with the purpose of studying to what extent this closeness favors the emergence and the occurrence of breakthroughs in a region. The technological base of a region is considered to consist of those technologies in which a region has developed a specialization, measured as the Revealed Comparative Advantage (RCA) (Hidalgo et al. 2007, Boschma et al. 2013). Specifically, a region r has RCA in technology i if the share of patents in technology i in its technological portfolio is higher than the share of technology i in the portfolio of all European regions:

$$RCA_{ir} = 1\ if\ \frac{patents_{ir} / \sum_i patents_{ir}}{\sum_r patents_{ir} / \sum_r \sum_i patents_{ir}} > 1\ and\ 0\ otherwise. \tag{2}$$

where $patents_{ir}$ represents the total number of patents in technology i in region r. Having RCA in technology i would imply that the region is more specialized in technology i than the EU average.

To assess the degree of relatedness and unrelatedness that each technology has with the technologies in which a region has RCA, we combine the RCA measure with the co-occurrence matrix to derive a density indicator, as done in Boschma et al. (2014):

$$UnrelD_{ir} = \frac{\sum_{i\epsilon r} \psi_{ij} RCA_{ir}}{\sum_i \psi_{ij}} \tag{3}$$

$$RelD_{ir} = \frac{\sum_{i\epsilon r} \eta_{ij} RCA_{ir}}{\sum_i \eta_{ij}} \tag{4}$$

The unrelated ($UnrelD_{ir}$) density indicator determines how close the set of technologies unrelated to technology *i* is to the technologies in which region *r* has a RCA. This is computed as the sum of all technologies *j* that are unrelated to technology *i* (ψ_{ij}=1) in which region *r* has a RCA, divided by the sum of unrelatedness of technology *i* to all other technologies *j* in all regions. A similar reasoning applies for the relatedness density indicator ($RelD_{ir}$). The values of both lie between 0% and 100%. A relatedness density of 60% would imply that region *r* has RCA in 60% of the technologies that are related to technology *i*.[4] An unrelatedness density of 20% would imply that region *r* has RCA in 20% of the technologies that are unrelated to technology *i*. These two density measures are our main variables of interest to explain the emergence and occurrence of truly radical and novel inventions.

3.3. Data and Methods

We use EPO patents unit-record data from OECD REGPAT database (September 2015 edition; Maraut et al. 2008), as well as forward citations (EPO-to-EPO, including indirect links through patent families) from ICRIOS database (Tarasconi and Coffano 2014). We consider patent applications in 277 NUTS2 European regions in 30 countries – EU-27, plus UK, Norway and Switzerland – in the period 1981 to 2010. When patents are produced by several inventors resident in different NUTS2 regions, they have been fully assigned to different regions (full counting). Patent statistics in the REGPAT database incorporate fractional counting when there are multiple inventors residing in different regions. However, knowledge is arguably a non-divisible asset, and since we are interested in knowledge

[4] While at the level of the co-occurrence matrices, being unrelated and related are opposite (either any two technologies are related or otherwise they are unrelated), once we transfer this to the regional level and consider only technologies in which the region has RCA, they are no longer opposite. Indeed, the related and unrelated density measures are positively correlated.

production at the regional level, non-fractional counts are preferred. In order to smooth the yearly lumpiness of patent data, we create time windows of five years starting from 1981 and lasting until 2010, combining the data over non-overlapping five-year periods (1981-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, and 2006-2010).[5]

The first set of empirical results presents some descriptive analyses at the patent level, concerning the extent to which breakthrough inventions (proxied by the two definitions discussed above) rely on unrelated and related combinations, and whether breakthroughs in a region combine and cite technological classes that are unrelated and available in the region. The second set of empirical results is based on regression analyses. In our estimations, the dependent variable is a measure of the occurrence of breakthrough patents in technology *i* in region *r*. This is regressed on measures of the degree of relatedness and unrelatedness that such technology *i* maintains with the technological portfolio of the region, while controlling for regional and technological characteristics. All specifications are estimated at the region-technology level for the different time windows, using the following specification:

$$BP_{ir,t} = \beta_0 + \beta_1 RelD_{ir,t-1} + \beta_2 UnrelD_{ir,t-1} + \beta X_{ir,\,t-1} + \omega_r + \varphi_i + \alpha_t + \varepsilon_{irt} \qquad (5)$$

where *r* refers to region, *i* to technology, and *t* to time period. *BP* is a measure of breakthrough patents, *RelD* and *UnrelD* stand for related and unrelated density, respectively, and *X* is a set of controls. All the estimations include region, technology and time fixed effects (ω_r, φ_i and α_t, respectively), to control for unobserved heterogeneity at these three dimensions. To consider deviations from the theory, a well-behaved error term is introduced, ε_{irt}.

Our first dependent variable measures breakthrough patents in a region and technology. As explained above, we use two definitions for a patent being considered as a breakthrough. The first one concerns a patent that is in the upper 10% of the distribution of number of forward citations for the technology and year it belongs to. We also check the robustness of our results to the utilization of the upper 5% and 1% of the distribution of number of forward citations

[5] The consideration of data in five-year periods is common practice in empirical analyses using patent data (among many, see Montresor & Quatraro (2019) and Santoalha & Boschma (2021)).

for the technology and year it belongs to – see the Appendix. As per the second definition, we label a patent as a breakthrough when it combines two technological classes for the first time within such technology in the region and year under consideration.

For each of the two definitions of breakthroughs, we look at both the occurrence of breakthrough inventions as well as their emergence in the region. In the first case, we consider two different dependent variables proxying for the occurrence of breakthrough inventions in a technology and region: the total number of breakthrough patents (*BP_Tot*) as well the share of breakthroughs out of the total number of patents in such technology (*BP_Share*). In a second step, we measure the emergence (entry) of breakthroughs in a NUTS2 region that did not have any breakthrough in a specific technology in the previous period with two different proxies. The first entry variable is a binary indicator which switches to 1 if the region in a given period has at least one breakthrough patent in a specific technology while it was not the case in the previous period (*BP_Entry*).[6] The second entry variable is also a binary variable which switches to 1 if the region in a given period acquires a relative advantage in producing breakthrough patents in a given technology compared to the EU average (*BP_Spec*), as follows:[7]

$$BP\_Spec_{irt} = 1 \quad \text{if} \quad \frac{BP_{irt}/P_{irt}}{\sum_r BP_{irt}/\sum_r P_{irt}} > 1 \text{ and } \frac{BP_{irt-1}/P_{irt-1}}{\sum_r BP_{irt-1}/\sum_r P_{irt-1}} < 1 \qquad (6)$$

and 0 otherwise

where BP_{irt} represents the total number of breakthrough patents and P_{irt} refers to the total number of patents in technology i, region r and time period t.

[6] We restrict to only one jump in a region for a technology, and forcing missing values in case the region already had breakthroughs in the first period of the analysis.

[7] If the region continues having a share which is larger than for the rest of regions in the following periods, it becomes missing value since a region can only enter if it was outside in the previous period.

As for the control variables, we merge the REGPAT database with other data primarily derived from EUROSTAT. At the regional level, we include the GDP per million inhabitants ($GDPpc$) to control for the economic wealth in the region, the population density as a proxy for urbanization economies (*Pop Dens*), and its square term to pick up non-linearities (*Pop Dens Sq*). We also include the technological stock measured as the total number of technological claims in a region to proxy for the number of ideas that could potentially be combined in a given region (*Tech Stock*). Additionally, we include technological size (*Tech Size*), measured as the number of technological claims of a technology, with the same reasoning as for the previous variable but at the technology level. Finally, when the dependent variable refers to the total number of breakthrough patents as well as the entry measure, (*BP_Tot* and *BP_Entry*, respectively), we include a measure of the overall patents produced in a region and a technology in a period of time (*TotPat*), to proxy for the overall innovativeness capacity of a region.

All variables are z-standardized, so that the coefficients can be compared within the estimation. To dampen potential endogeneity issues in the regressions, all explanatory variables are lagged by one period, denoted by t-1. Tables A1 and A2 in the Appendix provide summary statistics of the explanatory variables and the correlation matrix, respectively.

4. Findings from the descriptive analysis

We first present some descriptive analyses on the extent to which breakthrough inventions rely on unrelated and related combinations, and whether breakthrough inventions in a region combine technological classes that are unrelated and available in the region.[8,9]

[8] In this descriptive analysis, when a patent is breakthrough in one technology, it is considered as breakthrough in all the technologies in which it appears (even if in these other technologies it is not strictly classified as breakthrough).

[9] For each of the six time periods, we compute the shares of pairs of technologies that are related (observed together in the same patent document more frequently than expected by chance) and unrelated (observed together as rarely, as if they would co-occur by chance). It is much more common that two technologies are unrelated: around 90% of all pairs of technologies are unrelated, and around 10% related (Table A3 in Appendix).

Table 1 shows the share of patents with respect to the number of technological classes combined for breakthrough patents (BP defined as *Highlycited*) and all patents (P). Around 30% of all breakthrough patents belong to one technological class only (except for the last period 2006-2010), compared to around 50% for all patents. Interestingly, slightly more than 50% of all breakthroughs combine only related technologies, whereas only around 16% of all breakthroughs make unrelated combinations, and only 2% concerns combinations across unrelated technologies only. So, it is very rare that breakthrough patents (defined as highly cited) make unrelated combinations, and when they do, they combine both related and unrelated combinations.

Table 1. Share of patents according to the number of technological classes combined at the patent level. BP as *BP-Highlycited*

	Monotechnological Patents		Combining only related tech.		Combining Unrelated techs.		Combining only UR techs.	
	P	BP	P	BP	P	BP	P	BP
1981-1985	49.7%	31.8%	40.5%	51.8%	9.8%	16.4%	2.3%	1.9%
1986-1990	47.7%	31.9%	42.2%	52.5%	10.1%	15.6%	2.1%	1.7%
1991-1995	46.1%	28.9%	43.4%	53.6%	10.6%	17.5%	2.0%	1.6%
1996-2000	46.5%	28.9%	43.5%	54.7%	9.9%	16.4%	1.9%	1.5%
2001-2005	50.4%	38.4%	40.3%	48.5%	9.3%	13.1%	2.1%	1.8%
2006-2010	58.6%	52.0%	35.9%	41.5%	5.5%	6.6%	1.7%	1.6%

Note: BP is defined as those patents that are in the top 10% of forward citations. P refers to all patents.

Table 2 shows the average ratio of the number of pairwise combination of unrelated technologies within a patent document, with respect to its total pairwise technological combinations. The share of unrelated combinations used by breakthrough patents (defined as *BP-Highlycited*) is low, with an average of less than 10% in all periods, not being statistically different from patents in general. However, this share is higher when breakthroughs are defined as combining technologies for the first time, between 29% and 47%, being statistically different from general patents in this case.

Table 2. Number of pairwise combinations of unrelated technologies over total number of pairwise technologies within a patent. Average ratio

	P	BP-Highlycited	Diff	BP-Newcomb	Diff

1981-1985	0.098091	0.099215		0.294846	***
1986-1990	0.093728	0.092952		0.348666	***
1991-1995	0.090859	0.094369		0.357887	***
1996-2000	0.086258	0.08756		0.402626	***
2001-2005	0.095395	0.090876	**	0.469076	***
2006-2010	0.074894	0.069653	***	0.282352	***

Note: P refers to all patents. *Diff* refers to the t-test of equality of means between P and BP. * denotes significant at 10%, ** 5% and ***1%

Next, we investigate the extent to which patents in a region combine technological classes that are available in the region. To do this, for each patent p in a region r, we created the following index I_{pr}, as follows:

$$I_{pr} = \frac{S_{pr}}{T_{pr}} \tag{7}$$

where S_{pr} is the set of all pairwise technology combinations of such a patent that are also present in the set of all possible pairwise combinations of technologies of region r (after removing patent p), and T_{pr} is the set of all pairwise technology combinations of a patent p assigned to region r. This index is equal to one when the same elements are present in both sets. This means that the technological combinations set of patent's p is fully included in the region's r technological combinations set. This index is equal to zero when the intersection set is empty, meaning that none of the technological combinations of patent p is present in region r. This index is not defined for patents that have one single technology, nor for patents that are the only one in a region.

Table 3 presents the average of this index for every period for breakthrough patents and all patents. What can be observed first is that the index is close to 1. Breakthrough patents use mostly regional knowledge. This applies to all combinations and unrelated combinations. When we look at all combinations of technologies, between 86% and 97% of them in a breakthrough patent (Highly Cited) are also present in the region (column 3). These figures are between 70% and 86% of all combinations in a breakthrough patent defined as NewComb (column 5). Second, the index for breakthrough patents defined as those Highly Cited is significantly higher than for general patents: on average, breakthrough patents present a higher rate of technology combinations of the region than the average patent (P). The opposite is true for breakthrough patents defined as NewComb. When we look at unrelated combinations, between 83% and 96% (column 8) and between 70% and 87% (column 10) of unrelated combinations in a breakthrough patent are present in the region. This is similar though slightly less compared to all combinations. All in all, no matter the definition used, breakthrough inventions tend to combine primarily technological classes that are already present in the region.

Table 3. Average share of technological combinations of a patent that are present in the region

	All combinations					Unrelated combinations				
	P	**BP-Highlycited**	**Diff**	**BP-Newcomb**	**Diff**	**P**	**BP-Highlycited**	**Diff**	**BP-Newcomb**	**Diff**
1981-1985	0.83	0.86	***	0.7	***	0.78	0.83	***	0.7	***
1986-1990	0.87	0.89	***	0.74	***	0.83	0.86	***	0.73	***
1991-1995	0.91	0.94	***	0.8	***	0.89	0.92	***	0.79	***
1996-2000	0.95	0.97	***	0.86	***	0.93	0.95	***	0.85	***
2001-2005	0.95	0.97	***	0.84	***	0.93	0.95	***	0.82	***
2006-2010	0.95	0.97	***	0.81	***	0.94	0.96	***	0.87	***

Note: P refers to all patents. *Diff* refers to the t-test of equality of means between P and BP. * denotes significant at 10%, ** 5% and ***1%

Another way of looking at how breakthrough patents rely on local knowledge is to connect their technology classes through the patents they cite. We use backward citations EPO-to-EPO, including indirect links through patent families from the ICRIOS database (Tarasconi and Coffano 2014). We define 'inherited technologies' of a focal patent *p* to all technological classes that belong to patents cited by such patent. We identify all possible pairwise technology combinations coming from all the cited patents by the focal patent ('inherited combinations'). First, we construct the set of all possible pairwise technology combinations of the region where the focal patent belongs to, after removing the focal patent from the count. Then, for every patent, we define an index where the denominator is the total number of inherited combinations of patent *p*, and the numerator is the number of inherited combinations that are also present in the region's technology combinations portfolio. This is similar to the index in equation (7), but now based on combinations of technologies that breakthroughs actually cite.

Table 4 shows that the inherited combinations of technologies in breakthroughs are also present in the region. This applies to all combinations and unrelated combinations. When we look at all combinations of technologies, the index is higher than 0.8 for all the periods for any breakthrough patent (defined as highly cited) (column 3): the share of inherited combinations that are present also in the region is very high relative to its total portfolio of technological combinations. The index is slightly lower for patents in general (column 2), although the difference is marginal and not significant in many periods. Columns 5-6 present the same analyses for breakthrough patents, defined as those that combine for the first time two technologies. We observe similar findings in the sense that breakthrough patents share to lesser extent inherited combinations with the region than the average patent, but their reliance on the region is still considerable: between 76% and 87% of the combinations of technologies in a breakthrough patent are also present in the region. When we look at unrelated combinations, between 79% and 90% (column 8) and between 75% and 86% (column 10) of unrelated combinations in a breakthrough patent are present in the region. This is slightly less compared to all combinations.

Table 4. Average share of technological combinations in the cited patents that are present in the region

	All combinations					Unrelated combinations				
	P	**BP-Highlycited**	**Diff**	**BP-Newcomb**	**Diff**	**P**	**BP-Highlycited**	**Diff**	**BP-Newcomb**	**Diff**
1981-1985	0.833	0.834		0.758	***	0.777	0.791		0.749	**
1986-1990	0.853	0.859	*	0.778	***	0.803	0.81		0.745	***
1991-1995	0.885	0.892	***	0.828	***	0.848	0.851		0.807	***

1996-2000	0.922	0.928	***	0.871	***	0.893	0.899	***	0.857	***
2001-2005	0.923	0.923		0.864	***	0.895	0.894		0.844	***
2006-2010	0.907	0.911	***	0.835	***	0.876	0.879	***	0.82	***

Note: P refers to all patents. *Diff* refers to the t-test of equality of means between P and BP. * denotes significant at 10%, ** 5% and ***1%

5. Findings from regression analysis

Table 5 presents the FE estimations for the occurrence of breakthrough patents, defined as highly cited (columns 1 and 3) and combining two technological classes for the first time in the region (columns 2 and 4). Results indicate that it is mainly Relatedness Density (RelD) that is positively and significantly related to both the absolute number (columns 1-2) as well as the share of breakthrough patents in a region (columns 3-4). In contrast, Unrelatedness Density (UnrelD) is not significant (BP as *Highlycited*) or negatively and significantly related (BP as *Newcomb*). This indicates that a region has a higher number and a higher share of breakthrough inventions in a given technology if the overall technological portfolio of the region is related to such given technology. In sum, the occurrence of breakthrough patents in a region is enhanced by the local presence of related technologies, but not by the local presence of unrelated technologies. This latter variable has even a negative and significant coefficient when BP's are defined as making new combinations for the first time.

Table 5. Occurrence of Breakthroughs

	Total number (BP_Tot)		**Share (BP_Share)**	
	BP_Highlycited	BP_Newcomb	BP_Highlycited	BP_Newcomb
TotPat (st)	0.729***	0.107***		
	(0.059)	(0.018)		
RelD (st)	0.037***	0.038***	0.011***	0.042***
	(0.007)	(0.007)	(0.001)	(0.011)
UnrelD (st)	0.003	-0.038***	0.001	-0.059**
	(0.010)	(0.010)	(0.001)	(0.027)
Tech Stock (st)	-0.060***	0.232***	0.004***	0.344***
	(0.011)	(0.012)	(0.001)	(0.035)

Tech Size (st)	-0.089***	0.004	0.001	-0.103***
	(0.012)	(0.006)	(0.000)	(0.011)
GDPpc (st)	0.020	-0.039**	0.002	0.025
	(0.013)	(0.018)	(0.002)	(0.066)
Pop Dens (st)	0.377**	-0.601*	-0.004	0.007
	(0.152)	(0.316)	(0.022)	(1.197)
Pop Dens Sq (st)	-0.156***	0.168	0.005	-0.352
	(0.058)	(0.120)	(0.008)	(0.454)
Constant	-0.001***	-0.000	0.034***	0.309***
	(0.000)	(0.000)	(0.000)	(0.000)
Adjusted R2	0.59	0.08	0.08	0.02
N	666,333	666,333	666,333	666,333
Field FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Region-clustered standard errors. Explanatory variables are standardized.

We turn now to the analysis of the emergence of breakthrough inventions in a region. First, we consider in Table 6 (columns 1-2) the entry of a breakthrough patent in a region and technology for the first time in one of the periods under consideration (*BP_Entry*). Second, we contemplate in Table 6 (columns 3-4) when the region in a given period acquires a relative advantage in producing breakthrough patents in a given technology compared to the EU average which was not the case in the previous period (*BP_Spec*). What we observe again is that Relatedness Density (RelD) always shows a positive and significant coefficient: the local presence of related technologies enhances the emergence of breakthrough inventions in European regions. However, Tables 6 also shows that Unrelatedness Density (UnrelD) has now a significant positive impact on the emergence of breakthroughs inventions in regions. The magnitude of the effect of the local presence of related technologies is higher than the one of unrelated technologies when BP's are defined as highly-cited (columns 1-2), but this is the opposite case when BP's incorporate a new combination for the first time (columns 3-4).

Table 6. Emergence of Breakthroughs

	Entry (*BP_Entry*)		Entry (*BP_Spec*)	

	BP_Highlycited	*BP_Newcomb*	*BP_Highlycited*	*BP_Newcomb*
TotPat (st)	0.255***	0.000		
	(0.015)	(0.000)		
RelD (st)	0.050***	0.008***	0.062***	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)
UnrelD (st)	0.036***	0.014***	0.040***	0.014***
	(0.007)	(0.004)	(0.007)	(0.004)
Tech Stock (st)	-0.022***	-0.001	-0.008*	-0.001
	(0.005)	(0.001)	(0.005)	(0.001)
Tech Size (st)	-0.005	-0.016***	0.003	-0.016***
	(0.003)	(0.002)	(0.003)	(0.002)
GDPpc (st)	0.029**	0.001	0.027*	0.001
	(0.014)	(0.009)	(0.015)	(0.009)
Pop Dens (st)	-0.033	-0.227	-0.048	-0.228
	(0.241)	(0.150)	(0.248)	(0.150)
Pop Dens Sq (st)	-0.011	0.079	-0.005	0.079
	(0.099)	(0.051)	(0.102)	(0.051)
Constant	0.350***	0.221***	0.331***	0.221***
	(0.018)	(0.010)	(0.018)	(0.010)
Adjusted R2	0.23	0.90	0.23	0.90
N	88,505	32,151	88,505	32,151
Field FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

* p<0.1; ** p<0.05; *** p<0.01. Region-clustered standard errors. Explanatory variables are standardized.

All in all, we can conclude that for regions in Europe, breakthrough inventions rely especially on a stock of related technologies that are present in the region. For their emergence, technological breakthroughs rely not only on a local stock of related technologies, but also the local presence of unrelated technologies seems to matter. To recall, Tables 5 and 6 are reproduced in the online appendix with different breakthrough thresholds (5% and 1%), and the main conclusions of the paper remain (Tables A4 to A7).

6. Concluding remarks

This paper addressed critically the question whether technological breakthroughs build and rely on unrelated combinations in regions, as suggested by some research. Our study on the emergence and occurrence of breakthrough patents in European regions produced a number of interesting findings. First, we found that, by far, most combinations breakthrough patents make is between related technologies: almost no breakthrough patent makes combinations between unrelated combinations only. Even when breakthrough patents are defined as patents that make a combination not made previously, they tend to rely primarily on related combinations. Second, we found that breakthrough patents primarily combine and cite technological classes already present in the region. So, the knowledge base of a region, and more in particular the availability of local technologies seems to matter for the combinatory process of breakthrough patents. Third, we found that the relatedness density between a given technology and the overall technological portfolio of a region enhances both the occurrence and emergence of breakthrough inventions in a region, whereas unrelatedness density (that is, a local stock of technologies unrelated to the technology of a breakthrough patent) has a positive impact only on the emergence of breakthroughs in a region.

These results seem to challenge the radical nature of technological breakthroughs and the overall importance of unrelatedness for breakthrough patents. This seems to be in line with other papers in innovation studies (e.g. Kaplan and Vakili 2015) and economic geography (e.g. Miguélez and Moreno 2018; Hesse and Fornahl 2020; De Noni and Belussi 2021). It also contributes to the regional diversification literature, in which there is a tendency to make an artificial distinction between related and unrelated diversification. As Boschma (2017) suggested, it might be misleading to put it in terms of either or, as in practice it will be a matter of degree, or a mixture of related and unrelated combinations when regions diversify. Our findings on breakthrough inventions point in that direction, as local stocks of related and unrelated technologies seem to matter both for the emergence of technological breakthroughs in regions. This suggests that breakthroughs (even the more radical ones) rely to a considerable degree on pre-existing and well-known local knowledge sources that have been used and combined before. This might be even considered essential for breakthroughs to survive, because relying on unrelated combinations only would be too risky.

This paper has also raised issues that call for further research. First, the study has revealed the importance of a local stock of related technologies for the occurrence and the emergence of breakthrough inventions in regions. What still needs to be determined in future research is which technological breakthroughs are the most successful ones. The study suggests that making related combinations might be a necessary condition for breakthroughs to survive (the more so when they make unrelated combinations), but this needs to be investigated more closely. Second, the study found there are a few breakthrough inventions that make only unrelated combinations. It would be very interesting to explore in detail how these emerge in regions, seemingly against all odds. Third, the study has focused on cognitive distance (degree of relatedness) and geographical proximity (local or not) but other dimensions like social proximity might favor the development of breakthrough inventions in regions as well (Fleming et al. 2007; Crespo et al. 2014). It would be interesting to investigate whether breakthrough inventions are more likely to make combinations that involve technologies with low or high social proximity. Fourth, we did not explore the role of inter-regional linkages for breakthrough patents (Balland and Boschma 2021). Miguelez and Moreno (2018) showed that extra-regional knowledge linkages promote radical breakthroughs when the external knowledge is related (but not similar) to the knowledge base of the region. Hesse and Fornahl (2020) found that inter-regional linkages mattered at the dyadic level of unrelated combinations. A next step is to research whether a technological breakthrough relies on inter-regional linkages with respect to the related combinations it makes, or to its unrelated combinations (Balland and Boschma 2021). Fifth, we did not investigate the role of regional institutions (and policy) that might be crucial for the development of technological breakthroughs (Boschma and Capone 2015). We envisage that bridging institutions favor the recombinant search process, as they enhance interaction and collective action (Cortinovis et al. 2017). These may be even more important for unrelated combinations that are harder and riskier to combine, and thus might require institutions (and institutional change) to bridge the cognitive distance between the combined technologies. This still needs to be investigated.

7. REFERENCES

Aarstad, J., O. A. Kvitastein and S.-E. Jakobsen (2016). Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study. *Research Policy* 45(4), 844-856.

Abbasiharofteh, M., D. F. Kogler and B. Lengyel (2020) Atypical combination of technologies in regional co-inventor networks, Papers in Evolutionary Economic Geography (PEEG) 20.55, Utrecht University, Utrecht.

Acs, Z. J., L. Anselin and A. Varga (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy* 31(7), 1069-1085.

Ahuja, G. and C. M. Lampert (2001) Entrepreneurship in the large corporation: a longitudinal study of how established firms create breakthrough inventions, *Strategic Management Journal* 22, 521–543.

Arant, W., D. Fornahl, N. Grashof, K. Hesse and C. Söllner (2019) University-industry collaborations—The key to radical innovations? *Review of Regional Research* 39, 119–141.

Arts, S., and Veugelers, R. (2015) Technology familiarity, recombinant novelty, and breakthrough invention. *Industrial and Corporate Change* 24 (6), 1215–1246.

Audretsch, D. B., and Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. *Handbook of Regional and Urban Economics* 4, 2713–2739.

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

Balland, P. and R. Boschma (2021) Complementary inter-regional linkages and Smart Specialisation. An empirical study on European regions, *Regional Studies*, doi: 10.1080/00343404.2020.1861240

Balland, P.A. and Rigby, D. (2017) The geography of complex knowledge. *Economic Geography* 93 (1), 1–23.

Berkes, E. and Gaetani, R. (2020) The Geography of Inconventional Innovation. *The Economic Journal*. https://doi.org/10.1093/ej/ueaa111.

Bettencourt, L. M. A., Lobo, J., and Strumsky, D. (2007) Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research Policy* 36 (1), 107–120.

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

Boschma, RA (2017). Relatedness as driver behind regional diversification: A research agenda, *Regional Studies*, 51(3), 351–364.

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

Boschma, R., Heimeriks, G. and Balland, P. A. (2014) Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy* 43 (1), 107–114

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

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

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

Carlino, G. A., Chatterjee, S. and Hunt, R. M. (2007) Urban density and the rate of invention. *Journal of Urban Economics* 61(3), 389–419.

Carlo, J., Lyytinen, K. and Rose, G. (2012) A Knowledge-Based Model of Radical Innovation in Small Software Firms. *MIS Quarterly, 36*(3), 865-895.

Carnabuci, G. and E. Operti (2013) Where Do Firms' Recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination, *Strategic Management Journal* 34 (13), 1591–1613.

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

Castaldi, C., and Los, B. (2017) Geographical patterns in US inventive activity 1977– 1998: The 'regional inversion' was underestimated. *Research Policy* 46 (7), 1187–97.

Coffano, M., and G. Tarasconi (2014) CRIOS-Patstat database: sources, contents and access rules. *Center for Research on Innovation, Organization and Strategy, CRIOS working paper*, (1).

Cortinovis, N., J. Xiao, R. Boschma and F. van Oort (2017) Quality of government and social capital as drivers of regional diversification in Europe, *Journal of Economic Geography* 17 (6), 1179–1208.

Crespo, J., Suire, R., & Vicente, J. (2014) Lock-in or lock-out? How structural properties of knowledge networks affect regional resilience. *Journal of Economic Geography*, 14(1), 199–219.

Dahlin, K. B., and Behrens, D. M. (2005). When is an invention really radical? Defining and measuring technological radicalness. *Research Policy*, 34(5), 717-737.

De Noni, I. and F. Belussi (2021). Breakthrough invention performance of multispecialized clustered regions in Europe*, Economic Geography*, DOI: 10.1080/00130095.2021.1894924

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

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy* 11 (3), 147-162.

Eck, N. J. V., and Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the Association for Information Science and Technology*, 60(8), 1635-1651.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1), 117-132.

Fleming, L., S. Mingo and D. Chen (2007), Collaborative brokerage, generative creativity and creative success, *Administrative Science Quarterly* 52(3), 443–475.

Frenken, K, van Oort, FG and Verburg, T (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685–697.

Fritsch, M. and M. Wyrwich (2021) Is innovation (increasingly) concentrated in large cities? An international comparison, *Research Policy* 50, https://doi.org/10.1016/j.respol.2021.104237

Gavetti G. and Levinthal D. (2000) Looking forward and looking backward: cognitive and experiential search, *Administrative Science Quarterly* 45 (1), 113–137.

Grashof, N., K. Hesse and D. Fornahl (2019) Radical or not? The role of clusters in the emergence of radical innovations, *European Planning Studies* 27 (10), 1904-1923.

Hall, B. H., Jaffe, A., and Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics* 36(1), 16–38.

Harhoff, D., F. Narin, F. Scherer and K. Vopel (1999) Citation frequency and the value of patented inventions, *The Review of Economics and Statistics* 81(3), 511–515.

Hervás-Oliver J.L., J. Albors-Garrigos,S. Estelles-Miguel and C. Boronat-Moll (2018) Radical innovation in Marshallian industrial districts, *Regional Studies* 52 (10), 1388-1397.

Hesse, K. and D. Fornahl (2020), Essential ingredients for radical innovations? The role of (un-)related variety and external linkages in Germany, *Papers in Regional Science* 99 (5), 1165-1183.

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

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

Jaffe, A. B. and de Rassenfosse, G. (2016) Patent Citation Data in Social Science Research: Overview and Best Practices (January 2016). NBER Working Paper No. w21868, Available at SSRN: https://ssrn.com/abstract=2713593

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

Kaplan, S. and K. Vakili (2015) The double-edge sword of recombination in breakthrough innovation, *Strategic Management Journal* 36, 1435-1457.

Kelley, D.J., A. Ali and S.A. Zahra (2013) Where Do Breakthroughs Come From? Characteristics of High-Potential Inventions, *Journal of Product Innovation Management*, 30 (6), 1212-1226.

Kerr, W. R. (2010) Breakthrough Inventions and Migrating Clusters of Innovation, *Journal of Urban Economics* 67, 1, 46–60.

Kim, D., Cerigo, D.B., Jeong, H. and Youn, H. (2016) Technological novelty profile and invention's future impact. *EPJ Data Science* 5 (1), 721.

Lanjouw J.O. and Schankerman M. (2004) Patent quality and research productivity: measuring innovation with multiple indicators. *Economic Journal* 114 (495), 441–465.

Li, D. (2020), Place dependence of renewable energy technologies. Connecting the local and global scale, PhD thesis, Utrecht University, Utrecht.

Li, D., G. Heimeriks and F. Alkemade (2021) Recombinant invention in solar photovoltaic technology: can geographical proximity bridge technological distance?, *Regional Studies*, 55 (4), 605-616.

Maraut, S., Dernis, H., Webb, C., Spiezia, V. and Guellec, D. (2008). The OECD REGPAT Database (OECD Science, Technology and Industry Working Papers). Paris: Organisation for Economic Co-operation and Development.

March J.G. (1991) Exploration and exploitation in organizational learning. *Organization Science* 2 (1), 71–87.

Martynovich, M. and J. Taalbi (2020) Related variety, recombinant knowledge and regional innovation. Evidence for Sweden, 1991-2010, Papers in Evolutionary Economic Geography, no. 20.15, Utrecht University, Utrecht.

Mewes, L. (2019). Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836 to 2010. *Economic Geography*, 95 (4), 341-361

Miguélez, E. and R. Moreno (2018) Relatedness, external linkages and regional innovation in Europe, *Regional Studies* 52, 688–701.

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

Nelson, R.R. and Winter, S.G. (1982) *An Evolutionary Theory of Economic Change*. Cambridge: The Belknap Press of Harvard University Press

Nooteboom, B. (2000). *Learning and innovation in organizations and economies*. Oxford, Oxford University Press

O'hUallichain, B. (1999). Patent places: Size matters. *Journal of Regional Science* 39 (4): 613–36.

Phene, A., K. Fladmoe-Lindquist and L. Marsh (2006), Breakthrough innovations in the U.S. biotechnology industry: the effects of technological space and geographic origin, *Strategic Management Journal* 27 (4), 369-388.

Santoalha, A. and R. Boschma (2021) Diversifying in green technologies in European regions: does political support matter?, *Regional Studies* 55 (2), 182-195.

Saviotti, P. P., and Frenken, K. (2008). Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, 18(2), 201-218.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Organization, WIPO. Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany.

Schoenmakers, W., and Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, 39(8), 1051-1059.

Sonn, J. W., and Storper, M. (2008). The increasing importance of geographical proximity in knowledge production: An analysis of US patent citations, 1975–1997. *Environment and Planning A* 40 (5), 1020–1039.

Strumsky, D., and Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. Research Policy, 44(8), 1445-1461.

Tavassoli, S. and N. Carbonara (2014). The role of knowledge variety and intensity for regional innovation. *Small Business Economics* 43(2): 493-509.

Taylor A. and Greve H.R. (2006) Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Academy of Management Journal* 49 (4), 723–740.

Trajtenberg, M. (1990) A penny for your quotes: patent citations and the value of innovations, *RAND Journal of Economics* 21(1), 172–187.

Trajtenberg, M., R. Henderson and A. Jaffe (1997) University versus corporate patents: a window on the basicness of invention, *Economics of Innovation and New Technology* 5 (1), 19–50.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B. (2013) Atypical Combinations and Scientific Impact. *Science* 342, 468–472.

Varga, A. (2000) Local academic knowledge transfers and the concentration of economic activity. *Journal of Regional Science* 40 (2), 289–309.

Verhoeven, D., Bakker, J., and Veugelers, R. (2016) Measuring technological novelty with patent-based indicators. *Research Policy* 45 (3), 707–723.

Wang, J., Veugelers, R., Stephan, P. (2017) Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* 46 (8), 1416‑1436.

Weisberg RW. (1999) Creativity and knowledge: a challenge to theories. In *Handbook of Creativity*, Sternberg RJ (ed). Cambridge University Press: Cambridge, UK, pp. 226–250.

Weitzman, M. L. (1998) Recombinant growth. *The Quarterly Journal of Economics* 113(2), 331-360.

WIPO (2015) World Intellectual Property Report 2015 - Breakthrough Innovation and Economic Growth (WIPO Economics & Statistics Series). World Intellectual Property Organization - Economics and Statistics Division, Geneva, Switzerland.

Appendix

Table A1. Descriptive of variables

Variable	Mean	Std. Dev.	Min	Max
Number of patents	3.67	22.07	0	3118
Number of BP as in top 1%	0.04	0.45	0	42
Number of BP as in top 5%	0.22	1.65	0	172
Number of BP as in top 10%	0.43	3.07	0	318
Number of BP as new combination	0.01	0.17	0	14
Share of BP as in top 1%	0.00	0.04	0	1
Share of BP as in top 5%	0.02	0.09	0	1
Share of BP as in top 10%	0.03	0.13	0	1
Share of BP as new combination	0.00	0.04	0	1
Entry of BP as in top 1%	0.32	0.47	0	1
Entry of BP as in top 5%	0.34	0.47	0	1
Entry of BP as in top 10%	0.35	0.48	0	1
Entry of BP as new combination	0.21	0.41	0	1
Entry of specialization as in top 1%	0.32	0.47	0	1
Entry of specialization as in top 5%	0.34	0.47	0	1
Entry of specialization as in top 10%	0.35	0.48	0	1
Entry of specializacion as new combination	0.21	0.41	0	1
Rel Dens	24.35	14.80	0	100
Unrel Dens	19.87	9.86	0	47.81
Tech Stock	3052.09	5493.93	2	55845.2
Tech Size	4101.39	11742.31	1.5	255406
GDPpc	0.02	0.01	0.0014151	0.1628184
Pop Dens	0.36	0.87	0.0030805	9.613589

Table A2. Correlation matrix

	1	2	3	4	5	6	7	8
1. Number of patents	1.00							
2. Rel Dens	0.19	1.00						
3. Unrel Dens	0.19	0.54	1.00					
4. Tech Stock	0.22	0.20	0.45	1.00				
5. Tech Size	0.20	-0.17	-0.10	-0.08	1.00			
6. GDPpc	0.06	0.11	0.21	0.26	-0.08	1.00		
7. Pop Dens	0.02	-0.02	0.04	0.09	-0.05	0.60	1.00	
8. Pop Dens Sq	0.00	-0.03	0.01	0.03	-0.03	0.68	0.92	1.00

Table A3. Share of related/unrelated technologies within a patent as in co-occurrence matrices

	Related technologies (%)	**Unrelated technologies (%)**
1981-1985	8.7%	91.3%
1986-1990	10.0%	90.0%
1991-1995	11.8%	88.2%
1996-2000	11.6%	88.4%
2001-2005	12.0%	88.0%
2006-2010	10.3%	89.7%

Table A4. Occurrence of Breakthroughs: Total number (*BP_Tot*). Definition of highly cited patents (5% and 1% top)

	# BP5%	# BP1%
TotPat (st)	0.663***	0.497***
	(0.070)	(0.075)
Rel Dens (st)	0.041***	0.038***
	(0.009)	(0.011)
Unrel Dens (st)	-0.003	-0.009
	(0.010)	(0.009)
Tech Stock (st)	-0.055***	-0.033*
	(0.013)	(0.017)
Tech Size (st)	-0.081***	-0.061***
	(0.014)	(0.018)
GDPpc (st)	0.022	0.015
	(0.014)	(0.012)
Pop Dens (st)	0.301*	0.087
	(0.158)	(0.166)
Sq Pop Dens (st)	-0.131**	-0.043
	(0.060)	(0.063)
Constant	-0.002***	-0.002***
	(0.000)	(0.000)
Adjusted R2	0.49	0.29
N	666,333	666,333
Field FE	Yes	Yes
Region FE	Yes	Yes
Time FE	Yes	Yes

* $p<0.1$; ** $p<0.05$; *** $p<0.01$. Region-clustered standard errors. Explanatory variables are standardized.

Table A5. Occurrence of Breakthroughs: Share (*BP_Share*). Definition of highly cited patents (5% and 1% top)

	% BP5%	% BP1%
Rel Dens (st)	0.005***	0.001***
	(0.000)	(0.000)
Unrel Dens (st)	0.000	0.000**
	(0.001)	(0.000)
Tech Stock (st)	0.001***	0.000**
	(0.000)	(0.000)
Tech Size (st)	-0.000	-0.000**
	(0.000)	(0.000)
GDPpc (st)	0.002**	0.000
	(0.001)	(0.000)
Pop Dens (st)	-0.003	-0.006
	(0.014)	(0.008)
Sq Pop Dens (st)	0.002	0.002
	(0.006)	(0.003)
Constant	0.017***	0.003***
	(0.000)	(0.000)
Adjusted R2	0.04	0.01
N	666,333	666,333
Field FE	Yes	Yes
Region FE	Yes	Yes
Time FE	Yes	Yes

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Region-clustered standard errors. Explanatory variables are standardized.

.

Table A6. Emergence of Breakthroughs: Entry (*BP_Entry*). Definition of highly cited patents (5% and 1% top)

	BP5%	BP1%
TotPat (st)	0.118***	0.029***
	(0.011)	(0.005)
Rel Dens (st)	0.049***	0.035***
	(0.004)	(0.005)
Unrel Dens (st)	0.027***	0.024***
	(0.007)	(0.008)
Tech Stock (st)	-0.015***	-0.006
	(0.005)	(0.005)
Tech Size (st)	-0.008**	-0.012***
	(0.003)	(0.004)
GDPpc (st)	0.033*	0.011
	(0.018)	(0.020)
Pop Dens (st)	0.104	0.236
	(0.242)	(0.386)
Sq Pop Dens (st)	-0.079	-0.106
	(0.100)	(0.153)
Constant	0.310***	0.265***
	(0.023)	(0.053)
Adjusted R2	0.24	0.27
N	66,885	28,836
Field FE	Yes	Yes
Region FE	Yes	Yes
Time FE	Yes	Yes

* *p*<0.1; ** *p*<0.05; *** *p*<0.01. Region-clustered standard errors. Explanatory variables are standardized.

Table A7: Emergence of Breakthroughs: Entry (*BP_Spec*). Definition of highly cited patents (5% and 1% top)

	BP5%	BP1%
Rel Dens (st)	0.060***	0.043***
	(0.004)	(0.005)
Unrel Dens (st)	0.029***	0.021**
	(0.007)	(0.008)
Tech Stock (st)	-0.002	0.004
	(0.005)	(0.005)
Tech Size (st)	0.001	-0.004
	(0.003)	(0.003)
GDPpc (st)	0.032*	0.009
	(0.018)	(0.021)
Pop Dens (st)	0.082	0.193
	(0.251)	(0.398)
Sq Pop Dens (st)	-0.071	-0.087
	(0.104)	(0.158)
Constant	0.308***	0.273***
	(0.023)	(0.055)
Adjusted R2	0.24	0.27
N	66,885	28,836
Field FE	Yes	Yes
Region FE	Yes	Yes
Time FE	Yes	Yes

* p<0.1; ** p<0.05; *** p<0.01. Region-clustered standard errors. Explanatory variables are standardized