Working Paper Series

#2023-036

Technologies fly on the wings of science: Exploring the role of science in technologies' spatial evolution

Michele Pezzoni, Reinhilde Veugelers and Fabiana Visentin

Published 16 October 2023

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Boschstraat 24, 6211 AX Maastricht, The Netherlands
Tel: (31) (43) 388 44 00

UNU-MERIT Working Papers
ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology
UNU-MERIT | Maastricht University

Technologies fly on the wings of science:

Exploring the role of science in technologies' spatial evolution

Michele Pezzoni
Université Côte d'Azur, CNRS, GREDEG, France
Observatoire des Sciences et Techniques, HCERES, Paris, France
ICRIOS, Bocconi University, Milan, Italy
michele.pezzoni@univ-cotedazur.fr

Reinhilde Veugelers
KULeuven
CEPR, Bruegel and PIIE
reinhilde.veugelers@kuleuven.be

Fabiana Visentin
UNU-MERIT, Maastricht University, Maastricht, Netherlands
visentin@merit.unu.edu

Abstract: This paper investigates how science affects the geographical evolution of technological trajectories. We define a technological trajectory as a series of inventions re-using the same technology, and we follow the geographical development of trajectories by geo-localizing inventors. Following 10,782 trajectories, we find that technologies with a high scientific content travel longer distances and are more likely to generate new inventors' clusters than technologies with low scientific content, especially in their growth and maturity phases.

Keywords: Patents, Technological trajectories, Geographical evolution, Scientific content

JEL code: O30, O33

1. Introduction

Science-based technologies have gained importance in recent decades (Tijssen, 2002). Several examples confirm the importance of technological applications resulting from fundamental scientific discoveries (Nelson and Winter, 1982; Rosenberg and Nelson, 1994). For instance, in the mid '80ies, the discovery of the onco-mouse, a genetically modified mouse likely to develop cancer, developed at the Harvard Medical School was the starting point of a series of patented inventions relying on the common technological idea of creating genetically modified mammals able to generate diseases such as Alzheimer, diabetes, or cystic fibrosis (Murray, 2010). This series of inventions building on the transgenic mammal technology dramatically improved the effectiveness of research aiming to find treatments for a multitude of diseases (Hanahan et al., 2007). The onco-mouse represented the origin of a technological trajectory linking multiple patented inventions relying on the same idea of combining the "genome editing" and "rearing mammals from fertilized eggs" technological components to achieve the goal of growing genetically modified mammals developing diseases (Hanahan et al., 2007; Pezzoni et al., 2022). In the next 20 years, after the transgenic mammal technology appeared in the onco-mouse invention, more than 200 patented inventions relying on the transgenic mammal technology were developed worldwide.

The onco-mouse is an example of a technology heavily reliant on science, but not all technologies rely on science to the same extent (Marx and Fuegi, 2020). Several studies have investigated the role of science in the technological development process (Nelson and Winter, 1982; Pezzoni et al., 2022; Sorenson and Fleming, 2004) and the attractiveness of science for companies (Bikard and Marx, 2020). However, extant literature has a limited understanding of the effect of science on the technologies' geographical evolution. An exemption is Sorenson and Fleming's work (2004), which looks at the geographical visibility of inventions. They find that science-based inventions attract more citations from patents listing inventors who are located farther away in space than non-science-based inventions. Their explanation for this result is that the scientific information on which inventions rely travels farther in space than non-scientific information.

Understanding the role of science in influencing the geographical evolution of the innovation process has crucial repercussions in anticipating the potential of development for a territory. For instance, for policymakers, it is important to understand if science leads technologies to stay geographically concentrated, like in the case of ICT technologies in Silicon Valley, biotechnologies in the Boston area, cybersecurity in Tel Aviv, and robotics in Tokyo, or if scientific content makes technologies travel long distances leading to a more spread model of technology development.

In this paper, we investigate how the scientific content of technologies is associated with the pattern of geographical evolution of technological trajectories. In other words, we explore if science-based technologies have the potential to fly longer distances on the wings of science.

We expect that a high level of scientific content leads technologies to travel longer distances than technologies with low scientific content due to the different socio-economic rules governing the work of scientists and technologists (Dasgupta and David, 1994). Once scientific knowledge is produced, scientists aim to disseminate it as much as possible. This behavior results from the norm of "communism" governing science, according to which scientists, contrary to technologists, are incentivized to spread the information concerning their discoveries as much as possible to establish priority on their scientific achievements (Merton, 1942; Stephan, 2012). Once the priority is established, typically through publication, scientists have the further incentive to engage in outreach activity to gain peer recognition and, consequently, reward in terms of prestige and career from the scientific community (Dasgupta and David, 1994; Merton, 1968; Stephan, 2012). Moreover, professional collaborations among scientists have become a prevalent way of organizing research work (Wuchty et al., 2007), making universities and research centers highly connected worldwide and favoring knowledge diffusion. Therefore, scientists' incentives to share scientific knowledge at the base of technologies make this knowledge available at long distances. The long distance traveled by scientific knowledge favors its reuse by other inventors far from where that knowledge originated. Unlike scientists, technologists working in the private sector tend to keep the results of their research secret to preserve their advantage over competitors. Alternatively, they disclose their research results using intellectual property rights to guarantee legal protection to assure returns on companies' R&D investments (Dasgupta and David, 1994). Although disclosing knowledge, using intellectual property rights introduces constraints in knowledge reuse, for instance, due to the risk of patent infringement. Therefore, knowledge about non-science-based technologies circulates less in space than the one embedded in science-based technologies. The main mechanisms through which the knowledge about non-science-based technologies circulates are localized knowledge spillovers that keep the reuse of the technology close to its origin (Feldman and Kogler, 2010; Sorenson and Fleming, 2004). The above-stated differences between science-based and non-science-based technologies lead us to hypothesize that technologies with high scientific content travel longer distances than those with low scientific content.

Nonetheless, the effect of having high scientific content might vary according to the development phase of the technological trajectory. Specifically, we expect the scientific content to have a negligible effect on the distance traveled by the technology at the beginning of the technological trajectory and a larger effect during the trajectory's maturity phase. The following considerations drive our expectations. Even though scientific achievements are codified and made publicly available in publications immediately after their production, their reuse is not rapid since the understanding and inclusion of the knowledge in textbooks and engineers' training programs require time (Cowan and Foray, 1997; Howells, 2002; Klevorick et al., 1995; Kuhn, 1962; Von Hippel, 1998). For this reason, in the early phases, only a few places might pioneer and be able to embed in the technology the scientific knowledge newly available. Therefore, we do not expect any boost in the geographical spread of science-based technologies compared to non-science-based technologies. Only in the maturity phase of the technological trajectory, the geographical spread of scientific principles at the base of the technology is expected to benefit its worldwide diffusion (Merton, 1942; Sorenson and Fleming, 2004). In fact, over time, scientists travel, attend conferences, and train other scientists, spreading worldwide information about their scientific achievements at the base of technology. On the contrary, technologies with a low scientific content do not experience a similar boost in the diffusion of the knowledge on which they are based, traveling shorter distances.

The paper most similar to ours is Sorenson and Fleming (2004). The authors investigate if science benefits technological innovation by stimulating knowledge flow in space. They conduct an analysis at the patent level and find differences in the citation patterns of patents referring to scientific publications and patents without those references. They find that the patents with scientific content are cited by patents farther in space than patents with no scientific content. Differently from them, in our empirical analysis, we do not consider citations between patents, but we define technological trajectories. Each patent in a trajectory is an additional invention reusing the underlying common technological idea of the previous patents. By linking together patents sharing the same technological idea, we overcome the drawbacks of using citations. Indeed, citations are legal concepts connecting patents also for strategic reasons, not necessarily linking technological content (Jaffe et al., 2000; Strumsky and Lobo, 2015). Moreover, defining a technological trajectory allows us to identify different phases in developing technological ideas, i.e., take off, growth, and maturity. Our underlying assumption is that the effect of the science content of a technology on the distance the technology travels, might vary according to its development phase.

According to our definition of technological trajectory, we measure its geographical evolution in two ways. First, we look at the ‘jump’ in the geographical space made when a new invention is added to the trajectory; second, we consider if inventors reusing the technological idea are located in places so far away from the ones already reached by the technology to form a new cluster of inventors. To measure the scientific content of a technology at a given point of its development, we calculate the share of science-based inventions introduced until that moment over the total number of inventions related to the technology. Finally, we estimate the impact of the scientific content on geographical diffusion during three trajectory phases: take off, growth, and maturity phase.

We find that a higher scientific content negatively affects the spatial jump length in the early phase of the trajectory. On the contrary, scientific content increases the spatial jump length in the trajectory growth and maturity phases, when the trajectory has already generated a considerable number of inventions. Similarly, we observe that in the early phase of a trajectory, the probability of generating a new cluster of inventors is negatively affected by the scientific content, while when the trajectory is in its maturity phase, the probability of generating a new cluster increases with the scientific content.

Our paper brings four main contributions to the innovation literature. First, unique to our study, we consider a large set of trajectories when investigating the distance traveled by technologies. So far, the main contributions in studying the geographical evolution of technologies have relied on case studies focusing on individual technologies or countries. For example, Feldman et al. (2015) studied the case of the evolution of rDNA methods in the United States. Graziano and Gillingham (2015) considered the evolution of solar photovoltaic technology in Connecticut, Fontana et al. (2009) studied the LAN network technology, and Verspagen (2007) studied fuel cell technology. Second, while previous studies have mapped the geographical evolution of innovative activities using administrative borders, e.g., regions and countries (Feldman et al., 2015), as the most fine-grained level of analysis, we use geocoded inventors’ locations to compute the exact distance between inventions. Geocoding allows us to conduct a worldwide analysis going beyond extant studies that focus on specific places, countries, or sets of countries. Third, we go beyond the stylized mapping of technological trajectories over space (Nomaler and Verspagen, 2016), investigating the factors leading to a certain spatial evolution. In particular, we consider the technology’s science-based content as a driver. Finally, different from the extant literature that focuses on static analyses, we conduct a dynamic analysis assessing the determinants of the spreading of innovative activities over time along the different development phases of the technological trajectories (Balland et al., 2015).

2. Data, variables, and methodology

2.1 Data

Our study investigates the spatial evolution of a large pool of 10,782 technological trajectories that originated between 1985 and 1996, following each technology over a period of 20 years[1]. To identify the trajectories of each technology, we rely on the patent data provided by Patstat, the statistical database of the European Patent Office, from 1985 to 2015. To geo-localize the inventors, we use the information on the latitude and longitude of the inventors' addresses reported in patent documents. We retrieved the latitude and longitude of the addresses from de Rassenfosse et al. (2019) and Morrison et al. (2017)'s databases. In our analysis, we consider patent applications at the European Patent Office.

2.2 Technological trajectories

We aim to analyze the relationship between science and the geographical evolution of technological trajectories. As a first step, we need to define technological trajectories. To do so, we use the methodology proposed by Pezzoni et al. (2022). We define the origin of a trajectory, i.e., the appearance of a "technological idea," as an unprecedented combination of existing technological components (Arthur, 2009, 2007). Using patent data, we proxy technological components with IPC classes[2] and the origin of a trajectory as the patent embedding an unprecedented combination of IPC classes (Verhoeven et al., 2016). Then, we trace the technological trajectory by following the reuse of the same IPC class combination in follow-on patents[3]. We consider follow-on patents over a fixed window of 20 years. Following the

---

[1] In order to ensure the same observation window for each technological trajectory originating in different years, we set an observation window of 20 years for each trajectory. Indeed, for our last cohort of trajectories originating in 1996 we consider 20 years of observation until 2015, that is our last year of observation with reliable patent data.

[2] IPC (International Patent Classification) classes are hierarchical codes used to classify patent documents at the European Patent Office. Each code is made of six parts: Section, Class, Subclass, Main group, Subgroup (WIPO, 2019). Following Verhoeven et al. (2016), we proxy technological components with IPC codes at Main group level. For instance, the IPC code at Main group level "A01K67" corresponds to "Rearing or breeding animals, not otherwise provided for; New breeds of animals".

[3] Alternatively, technological trajectories have been reconstructed by other studies linking patents belonging to the same trajectory through the patent citation network (Fontana et al., 2009; Nomaler and Verspagen, 2016). However, we refrain from using a citation-based approach because we need to map a large set of technological trajectories, while studies using citations focus on a single technology. Moreover, there are some drawbacks in

methodology described above, we reconstruct 10,782 technological trajectories started between 1985 and 1996, ordering chronologically the N_i patents attributed to each trajectory i, from the oldest to the most recent one, according to the patent application date[4] (Equation 1).

$$Trajectory_i = \{p_{i,1}, p_{i,2}, \dots, p_{i,s}, \dots, p_{i,N_i}\}$$

Equation 1

In Equation 1, $p_{i,1}$ is the patent in trajectory i with the oldest application date, while p_{i,N_i} is the patent with the most recent application date. We observe trajectories along their development process. Each trajectory counts, on average, 54.92 patented inventions. Our sample includes a total of 592,173 trajectory-patent pairs. Depending on the location of their inventors, each patented invention in the development process brings the technology to a new place, incrementally or drastically expanding the trajectory's geographical space coverage. We measure the spatial evolution of the trajectory over geographical space by calculating two variables: *Jump length* and *New cluster*.

2.3 Jump length

The technological trajectory i's geographical coverage expands each time an inventor of a patent s brings the technology to a new place. To map the spread of the trajectory i over the geographical space, we calculate the variable $Jump\ length_{i,s}$ in two steps. First, we calculate the distance between the location of each inventor of patent s and the closest location reached by all the inventors in the stock of patents preceding patent s in the trajectory i, i.e., patents from 1 to $s-1$. Second, among the distances calculated for the inventors of patent s in the first step, we consider the geographical distance covered by the inventor who has brought the technology i the farthest. For example, consider the third patent in trajectory i, i.e., patent 3. The preceding patents, i.e., patents 1 and 2, represent the stock of patents before patent 3. Patent 1 lists two inventors, A and B, one located in Chestnut Hill (MA) and the other in San Francisco (CA). Patent 2 lists one inventor C, located in Huston (TX). Patent 3 lists three inventors, D and E, located in New York City and F in Strasbourg (France). Calculating $Jump\ length_{i,s}$ for $s=3$ implies calculating the distances D-A (288.1 km), E-A (289.4 km), F-A (5,932.1 km), D-B (4,143.7 km), E-B (4,143.3 km), F-B (9,253.2 km), D-C (2,300.6 km), E-C (2,299.5 km), and F-C (8,478.2 km). Once we calculated those distances, we selected the ones corresponding to

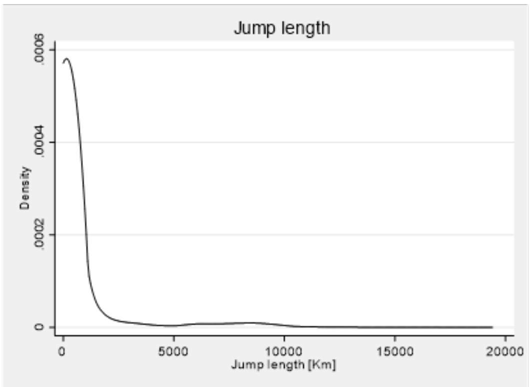
using citation networks to identify technological trajectories. For instance, citations do not fully capture the technological content of an invention (Strumsky and Lobo, 2015) being about 44% of citations weakly associated with the technological content of inventions (Jaffe et al., 2000).

[4] We replicated our econometric analyses using the priority date obtaining similar results.

the closest inventor to D, E, and F, namely the distance D-A (288.1 km), E-A (289.4 km), F-A (5,932.1 km). In other words, the closest inventor to D in the stock of previous patents is 288.1 km from D, the closest inventor to E is 289.4 km from E, and the closest to F is 5,932.1 km from F. These distances represent the expansion of the geographical coverage of the trajectory *i* due to each inventor of patent 3. Among these three distances, we select the maximum distance and assign its value to the variable *Jump length*$_{i,s}$, when $s=3$. In this example, technology *i* is brought farther by inventor F, and the variable *Jump length*$_{i,3}$ equals 5,932.1 km.

Figure 1 shows the kernel distribution of the variable *Jump length* for the patents forming our technological trajectories. The average jump length of a patent in a trajectory is 570 km (Table 1). We observe that short jumps, i.e., less than 1,000 km, are highly frequent (1,000 km corresponds roughly to the 90th percentile of the distribution). On the contrary, long-length jumps, i.e., more than 1,000 km, are rare. These figures align with the extant literature showing that the knowledge diffusion process is geographically localized (Breschi and Lissoni, 2001; Feldman and Kogler, 2010).

Figure 1. Jump length distribution for the patents forming technological trajectories



2.4 New cluster

Inventors of patent *s* can locate close to the inventors of the preceding patents in trajectory *i,* becoming part of an existing cluster, or locate far away from the inventors of the preceding patents, generating a new cluster. We define *New cluster*$_{i,s}$ equal to one if *Jump length*$_{i,s} > 1{,}000$ km, meaning that at least one inventor of the patent *s* is located at more than 1000 km from all the inventors of the preceding patents in the trajectory *i*. On the contrary, *New cluster*$_{i,s}$ equals zero when all the inventors of patent *s* join an existing cluster, locating at less than 1,000 km from the inventors listed in the patents preceding patent *s*. The distance threshold of 1,000 km corresponds roughly to the 90th percentile of the distribution of the variable *Jump length*$_{i,s}$ (see Figure 1).

2.5 Trajectory scientific content

We define a proxy for the scientific content of the trajectory i at patent s-1, looking at the scientific content of the patents from 1 to s-1. Specifically, we identify all the patents from 1 to s-1, citing at least one scientific document (Tijssen, 2002). As a scientific document, we consider the non-patent literature cited that includes the name of a scientific journal or a conference proceeding[5] (Sorenson and Fleming, 2004). Then, we calculate at patent s-1 the share of patents citing at least one scientific document for the trajectory i ($\textit{Share of scientific content}_{i,s\text{-}1}$).

2.6 Trajectory development phase

To take into account possible heterogeneity in geographical evolution, we split the development of each technology into three phases, i.e., take off, growth, and maturity. To define the development phase reached by the trajectory, we look at the position of patent s-1 over the entire trajectory i and define three dummy variables accordingly. We define the dummy variable $\textit{Take off}_{i,s\text{-}1}$ equals one if the patent s-1 belongs to the first 25% of the total number of patents in the trajectory ($[(s\text{-}1)/N_i] \leq 0.25$). The dummy variable $\textit{Growth}_{i,s\text{-}1}$ equals one if the patent s-1 is between 25% and 75% of the total number of patents in the trajectory ($0.25 < [(s\text{-}1)/N_i] \leq 0.75$). Finally, the dummy variable $\textit{Maturity}_{i,s\text{-}1}$ equals one if the patent s-1 is over 75% of the total number of patents in the trajectory ($[(s\text{-}1)/N_i] > 0.75$).

2.7 Descriptive statistics

Table 1 reports the descriptive statistics calculated over the 592,173 trajectory-patent observations included in our study sample. Table 1 shows that the *Take off* phase is characterized by longer jumps and a higher probability of generating a new cluster than the *Growth* and *Maturity* phases. Indeed, the average jump length in the take off phase equals 1,375.43 km, while during the growth and maturity phases, the average jump lengths are only 351.89 and 246.42 km. Similarly, we observe a lower probability of generating a new cluster after the *Take off* phase. These latter two figures are expected because the trajectory has already spread in the geographical space during the *Take off* phase, reducing the probability of long jumps and generating new clusters during *Growth* and *Maturity*. Interestingly, the average

---

[5] Patstat, in its 2019 version, provides a classification of the scientific documents that is reliable staring from the early 1990s. To ensure the coverage of this information before the 1990, we matched the text of the non-patent literature citations with a list of scientific journals and conferences covered by the Web of Science (Clarivate) bibliometric dataset.

Share of scientific content does not vary substantially over the trajectory's three phases, maintaining a value of around 37 percent.

Table 1. Descriptive statistics

	(1)	(2)	(3)	(4)	(5)
	Mean (Trajectory)	Mean (Take off)	Mean (Growth)	Mean (Maturity)	ANOVA P-value*
Jump length	579.96	1,375.43	351.89	246.42	0.00
New cluster	0.11	0.24	0.07	0.05	0.00
Share of scientific content	37.34	37.75	36.98	37.68	0.00
Trajectory-patent observations	592,173	146,785	301,409	143,979	

Note: * The null hypothesis of the one-way ANOVA is that the mean of the variable considered is equal for the three phases of the trajectory, H0: $\mu_{take\ off} = \mu_{growth} = \mu_{maturity}$. Rejecting the null hypothesis tells us that at least two means are statistically different.

2.8 *An illustrative example of trajectory geographical evolution: The transgenic mammal technology*

As an illustrative example of the empirical strategy followed, we calculate the geographical evolution of the *transgenic mammal* technology. The *transgenic mammal* technology originated in 1985 when it appeared as the result of combining ‘genome editing’ (IPC C07H21) with ‘rearing mammals from fertilized mouse eggs’ (IPC A01K67). At the Harvard Laboratories, Philip Leder and Timothy Steward joined their competencies and patented the so-called ‘onco-mouse’, a mouse genetically modified to study cancer that embedded for the first time the transgenic mammal technology. Subsequent patented inventions re-used the same technology, and after 20 years from its origin, 218 inventions reusing the same combination of IPC codes were patented[6].

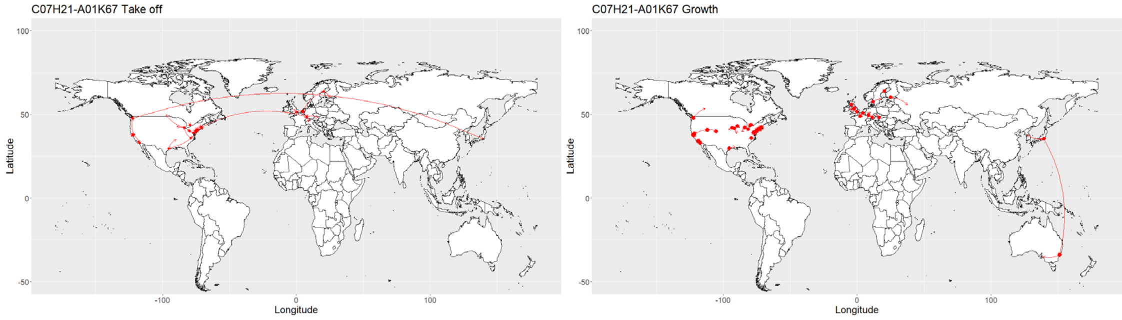
To assess the geographical evolution of the *transgenic mammal* technology along its development trajectory, we identify three phases: take off (0%-25% of the total number of patents included in the trajectory), growth (25%-75%), and maturity (75%-100%). Figure 2 shows a graphical representation of the geographical jumps of the trajectory. Following our definition of *Jump length*, lines in the figures connect inventors of patent *s* with the closest inventor in preceding patents from *1* to *s-1*. We report the inventors’ connections for each trajectory phase. Panel A in Figure 2 shows the jumps of the *transgenic mammal* technology during the take off phase of their trajectory. Panels B and C show the jumps for the trajectory during the growth and maturity phases, respectively.

Figure 2. Jumps of the transgenic mammal technology

Panel A: Take off
Transgenic mammal (54 patents)

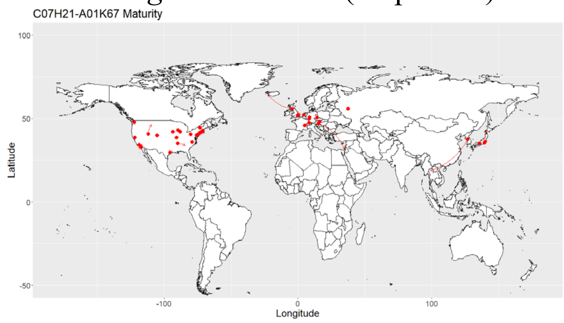
Panel B: Growth
Transgenic mammal (109 patents)

[6] The number of inventions patented were 221, however, we were able to geolocalize inventors for 218 patents. The 218 patents are the ones included in our analysis.



Panel C: Maturity

Transgenic mammal (54 patents)



3. Econometric methodology

Our econometric analysis aims to associate the technology's scientific content with the trajectory spatial evolution. Equation 2 reports the algebraic representation of the model that we estimate in our econometric exercise.

$$\begin{aligned} Spatial\ evolution_{i,s} &= \beta_0 + \beta_1 Share\ of\ scientific\ content_{i,s-1} * Trajectory\ phase_{i,s-1} \\ &+ \beta_3 Trajectory\ phase_{i,s-1} + \beta_4 Other\ controls_{i,s-1} + \gamma_i + \varepsilon_{i,s} \end{aligned}$$

Equation 2

The dependent variable *Spatial evolution*$_{i,s}$ takes, in turn, the values of *Jump length*$_{i,s}$ and *New Clusters*$_{i,s}$. The variable *Share of scientific content*$_{i,s-1}$ is our main explanatory variable. As argued in our hypotheses, *Share of scientific content* might show different effects according to the phase of the technological trajectory in which the trajectory is, i.e., take off, growth, or maturity phase. To assess the different effects of *Share of scientific content*$_{i,s-1}$ along the trajectory, we interact the variable with the vector *Trajectory phase*$_{i,s-1}$ that includes three dummy variables[7]: *Take off*$_{i,s-1}$, *Growth*$_{i,s-1}$, and *Maturity*$_{i,s-1}$. Finally, Equation 2 includes time-variant controls (*Other controls*$_{i,s-1}$) and technology fixed effects (γ_i) to control for all the

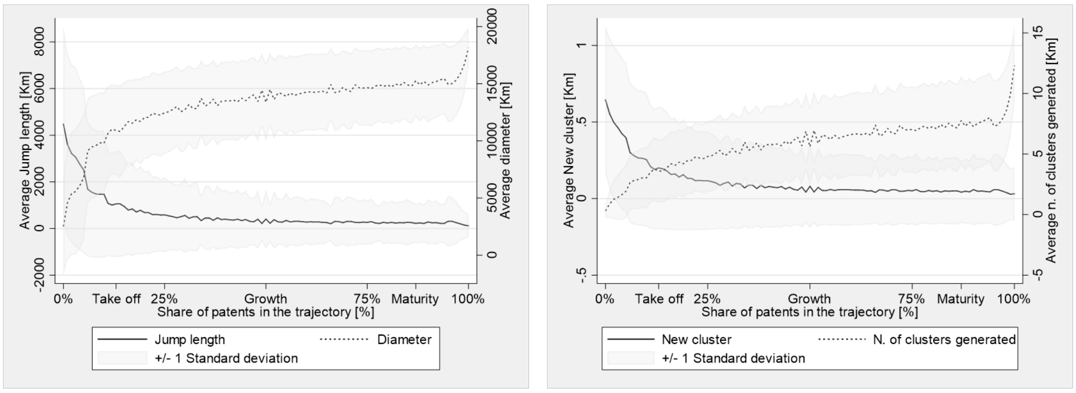
[7] The model reported in Equation 2 does not include the non-interacted variable *Share of scientific content*$_{i,s-1}$. The variable *Share of scientific content* is collinear with the full set of interactions between *Scientific content*$_{i,s-1}$, *Take off*$_{i,s-1}$, *Growth*$_{i,s-1}$, and *Maturity*$_{i,s-1}$.

unobservable time-invariant characteristics of the technology and an idiosyncratic error term ($\varepsilon_{i,s}$).

Controls

The geographical space the technology covers from patent *s* to patent *s-1* mechanically affects the jump length and probability of generating a new cluster at patent *s*. If the technology has already spread significantly in space before patent *s*, the likelihood of observing a long jump and creating a new cluster is expected to be short due to the lack of unexploded geographical space. On the contrary, if the technology did not spread significantly before patent *s*, the availability of a large unexplored space increases the likelihood of long jumps. Similarly, the probability of generating a new cluster of inventors developing the technology is high when few clusters of inventors exist worldwide, while it is lower when several clusters have already been generated. To control for the geographical space coverage of the trajectory and presence of previously generated clusters of inventors, we calculate the variables $Diameter_{i,s-1}$, and $Clusters\ generated_{i,s-1}$, which we include as control variables in all our econometric exercises. The variable *Diameter* corresponds to the distance between the two most distant inventors listed in the stock of patents from *1* to *s-1* while $Clusters\ generated_{i,s-1}$ corresponds to the cumulated number of clusters appearing in the trajectory, from patent *1* to patent *s-1*. Figure 3, panel A, illustrates the mechanical relationship between the variable *Jump length* and *Diameter,* while panel B illustrates the relationship between *New cluster* and *Clusters generated*. As expected, Figure 3, panel A, shows that in the early phases of the trajectory, the average *Jump length* is characterized by high values because most of the geographical space is not yet explored by the technology, as measured by low values of the average *Diameter*. On the contrary, in latter phases of the trajectory, most of the geographical space has already been explored, i.e., high values of the variable *Diameter*, and making long jumps becomes less likely, i.e., low values of average *Jump length*. Similarly, panel B shows that in the early phases of the trajectory, the lack of existing clusters of inventors developing the technology, i.e., low values of *Generated clusters*, makes it more likely to generate a *New cluster*, while in the latter phases of the trajectory, the existence of several clusters makes less likely to generate a new one.

Figure 3. *Jump length* versus *Diameter* (panel A) and *New cluster* versus *Generated clusters* (panel B) over the trajectory development phases



We also include as control the number of distinct inventors in the stock of patents in the technological trajectory, from patent *1* to patent *s-1* (*Stock of inventors*$_{i,s-1}$). A larger number of inventors in the trajectory is expected to decrease the chance of observing long jumps because inventors are likely to have already occupied a consistent part of the geographical space. For the same reason, the probability of observing a new cluster is expected to decrease with an increase in the number of inventors.

As time passes, the inventors' community becomes familiar with the technology. A longer time is expected to allow for reusing the technology at longer distances and generating new clusters where the technology is developed. To control for the number of years elapsed since the appearance of the technology *i*, we calculate the variable *Years passed since the entry year* as the years passed from the application year of the first patent in the technological trajectory to the application year of patent *s*.

The countries where the technology develops might also affect its spatial evolution. We created three variables *Share of US patents*$_{i,s-1}$, *Share of DE patents*$_{i,s-1}$, and *Share of JP patents*$_{i,s-1}$, calculating the share of patents from patent *1* to patent *s-1* with at least one inventor in the US, Germany, or Japan, respectively. We selected these countries because, historically, they produce the largest number of patents.

Technologies might have heterogeneous intrinsic tendencies to spread geographically. The econometric specification of our model reported in Equation 2 accounts for the unobservable time-invariant propensity of the technology to spread with technology fixed effects (γ_i). In an alternative model specification without fixed effects, we add a set of dummy variables to control for the technology specificities. We construct 118 dummy variables, one for each IPC class's 3-digits that appear in the IPC combinations that generate novel technologies (*Dummy technology classes*$_i$).

The historical period might affect the geographical evolution of the trajectories. For instance, in recent years, fast and low-cost transportation connections can favor geographical spread. In each model specification, we included a set of dummy variables for the application year of the patent *s*. The application years range from 1985 to 2015.

Table 4 reports the descriptive statistics of the control variables. The average number of distinct inventors in the technological trajectory equals 224.57. Looking at the country of origin of the patented inventions in our trajectories, we observe that the US is the most inventive country, with 43.87% of patents in our trajectories listing at least one inventor based in the US. Japan and Germany follow with a share of 19.68% and 12.68%, respectively. On average, the trajectories in our sample generate 5.78 clusters at patent *s* and cover a geographical space with a diameter of 13,184.54 km. At patent *s*, the average time passed since entry equals 11.22 years.

Table 4. Descriptive statistics for the control variables

592,173 trajectory-patent observations	Mean	SD	Median	Min	Max
Stock of inventors	224.57	353.97	86	1	3283
Share of DE patents [%]	12.68	14.73	8.24	0	100
Share of US patents [%]	43.87	24.46	41.67	0	100
Share of JP patents [%]	19.68	20.34	14.29	0	100
Years passed since the entry year	11.22	4.62	12	0	19
Diameter [km]	13,184.54	4,236.27	12,228.39	0	19,979.54
Clusters generated	5.78	3.62	5	0	21

4. Results

Table 5 reports the estimates of the coefficients of six regression models, having *Jump length* (Columns 1 to 3) and *New cluster* (Columns 4 to 6) as dependent variables, respectively. Columns 1 and 4 include *Share scientific content* and the control variables as explanatory variables. Columns 2 and 5 report a model where we add the full interaction terms between *Share of Scientific content* and the trajectory phase dummies, *Take off*, *Growth*, and *Maturity*. Columns 3 and 6 estimate the fully specified model described in Equation 2, including trajectory fixed effects.

Table 5: Regression results

	(1) OLS Jump length	(2) OLS Jump length	(3) OLS Jump length	(4) LPM New cluster	(5) LPM New cluster	(6) LPM New cluster
Share of scientific content	2.47***			0.00041***		
	(0.12)			(0.000021)		
Share of scientific content * Take off		-0.77***	-0.88***		-0.000045*	-0.00022***
		(0.15)	(0.22)		(0.000027)	(0.000041)
Share of scientific content * Growth		4.81***	6.27***		0.00071***	0.0013***
		(0.15)	(0.27)		(0.000028)	(0.000050)
Share of scientific content * Maturity		5.60***	8.86***		0.00093***	0.0021***
		(0.19)	(0.30)		(0.000036)	(0.000055)
Growth	-160***	-375***	-81.4***	-0.034***	-0.063***	-0.032***
	(6.63)	(9.40)	(10.6)	(0.0012)	(0.0017)	(0.0020)
Maturity	40.0***	-223***	2.13	-0.0035**	-0.044***	-0.029***
	(9.53)	(12.9)	(15.0)	(0.0018)	(0.0024)	(0.0028)
Log(Stock of inventors)	-323***	-322***	-281***	-0.062***	-0.061***	-0.032***
	(3.26)	(3.26)	(6.89)	(0.00060)	(0.00060)	(0.0013)
Share of DE patents	-3.75***	-3.70***	-3.13***	-0.00073***	-0.00073***	-0.00050***
	(0.19)	(0.19)	(0.29)	(0.000035)	(0.000035)	(0.000054)
Share of US patents	-0.96***	-0.94***	-2.79***	0.000100***	0.00010***	-0.00022***
	(0.14)	(0.14)	(0.22)	(0.000025)	(0.000025)	(0.000040)
Share of JP patents	-1.87***	-1.80***	0.70***	-0.00043***	-0.00042***	-0.00014***
	(0.16)	(0.16)	(0.27)	(0.000029)	(0.000029)	(0.000050)
Years passed since the entry year	-28.8***	-25.1***	-196***	-0.0042***	-0.0036***	-0.030***
	(0.86)	(0.86)	(3.11)	(0.00016)	(0.00016)	(0.00058)
Diameter	-0.097***	-0.095***	-0.20***	-8.4e-06***	-8.2e-06***	-0.000015***
	(0.00081)	(0.00081)	(0.0011)	(1.5e-07)	(1.5e-07)	(2.1e-07)
Clusters generated	62.6***	54.7***	-22.1***	0.0032***	0.0020***	-0.037***
	(1.25)	(1.27)	(2.22)	(0.00023)	(0.00024)	(0.00041)
Constant	4,690***	4,791***	4,286***	0.66***	0.67***	0.58***
	(71.2)	(71.2)	(72.7)	(0.013)	(0.013)	(0.014)
Observations	592,173	592,173	592,173	592,173	592,173	592,173
N. of technological trajectories	10,782	10,782	10,782	10,782	10,782	10,782
R-squared / R-squared within	0.190	0.192	0.215	0.150	0.151	0.149
Dummy application year	Yes	Yes	Yes	Yes	Yes	Yes
Dummy technology classes	Yes	Yes	No	Yes	Yes	No
Technological trajectory fixed effects	No	No	Yes	No	No	Yes

NOTE: When adding fixed effects, we drop the variables *Dummy technology classes* that are time-invariant variables for each trajectory. In the model using fixed effects, we report the R-squared for the within model.

In Column 1, we find that the higher the share of scientific content of a trajectory, the longer the average jump length. We find that a 10 percentage points higher *Share of scientific content* is associated with a 24.7 km longer jump[8]. The value 24.7 km corresponds to 4.25% of the average jump length observed in our sample (579.96 km). Although the coefficient is small when considering the whole trajectory, it is tightly linked to the phase reached by the technology, as shown in Columns 2 and 3. In the *Take off* phase, trajectories with a higher *Share of scientific content* have lower jump lengths (Column 3). However, a trajectory with a greater *Share of scientific content* in the growth and maturity phases spreads with longer jumps: a 10 percentage points higher *Share of scientific content* is associated with a jump of 62.7 km longer if the technology is in its *Growth* phase (17.69% of the average jump length during the growth phase), and of 88.6 km longer if the technology is in its *Maturity* phase (35.95% of the average jump length during the maturity phase), respectively[9].

When looking at the probability of observing a *New cluster* when the technology evolves, we find that a higher *Share of scientific content* is positively associated with the probability that the technology travels farther away from the existing areas, generating a new cluster (Column

[8] The value of 24.7 Km is calculated as the estimated coefficient for the variable *Share of scientific content* (2.47) times the increment of the variable *Share of scientific content* (10).

[9] We tested for the null hypothesis that $\beta_{\text{Share of scientific content * Maturity}} = \beta_{\text{Share of scientific content * Take off}}$ and for the null hypothesis that $\beta_{\text{Share of scientific content * Maturity}} = \beta_{\text{Share of scientific content * Growth}}$, in both cases we rejected the null hypothesis at 1% level.

4). In Column 6, where we control for *Technological trajectory fixed effects*, and we consider the interactions between *Share of scientific content* and the phase reached by the technology, we find that the *Share of scientific content* in the trajectory is negatively associated with the probability of generating a new cluster in the *Take off* phase. A ten percentage points higher *Share of scientific content* is associated with a 0.13 percentage points (1.86% of the unconditional probability of generating a new cluster during the growth phase) and 0.21 percentage points (4.2% of the unconditional probability of generating a new cluster during the maturity phase) higher probability of generating a new cluster in the *Growth* or *Maturity* phase, respectively[10].

Looking at the control variables, we find that *Diameter* shows the expected mechanical effect: the larger the *Diameter*, the lower the *Jump length*. The number of *Clusters generated* shows a similar mechanical effect: the larger the number of *Clusters generated*, the lower the probability of generating a *New cluster*. As explained in Section 3, including this control is crucial to avoid an omitted variable bias in estimating the coefficient of *Share of scientific content*[11]. Concerning the other control variables, we find that trajectories characterized by a larger number of inventors (*Stock of inventors*) have shorter jumps and a lower probability of generating a new cluster. This result is expected since having a high number of inventors in the trajectory makes it more likely to observe an inventor at a close distance to the inventors of patent *s*, leading to shorter jump lengths. Similarly, a larger number of inventors is associated with a lower probability of generating new clusters. Along the same line of reasoning, the variables *Year passed since the entry year*, *Growth*, and *Maturity*, show negative associations with jump length and the probability of generating a new cluster. Indeed, the longer the time elapsed from the entry year or the latter is the trajectory phase, the shorter the jumps and the lower the probability of generating new clusters. Interestingly, a higher share of patents in the trajectory with at least one inventor based in US, Germany, or Japan, is associated with shorter jumps and a lower probability of generating a new cluster, suggesting that trajectories concentrated in large highly innovative countries tend to stay in those countries, decreasing the length of observed jumps.

---

[10] We tested for the null hypothesis that $\beta_{\text{Share of scientific content * Maturity}} = \beta_{\text{Share of scientific content * Take off}}$ and for the null hypothesis that $\beta_{\text{Share of scientific content * Maturity}} = \beta_{\text{Share of scientific content * Growth}}$, in both cases we rejected the null hypothesis at 1% level.

[11] Omitting the variable *Diameter* would introduce a negative bias in the estimates of the coefficient of *Share of scientific content*. Indeed, *Diameter* is positively correlated (0.31) with *Share of scientific* content and, as shown in Table 5, it is negatively associated with both the dependent variables in all regression models.

4.1 Additional results

This section provides six alternative model specifications when estimating the relationship between *Share of scientific content* and the distance traveled by technologies. To detail the dynamics of the trajectory evolution, (1) we replace the dummy variables identifying the trajectory phases, i.e., take off, growth, and maturity, with a continuous variable, and (2) we split our study sample into three sub-samples according to the trajectory phase. To further explore the role of scientific knowledge in the trajectory evolution, (3) we investigate how the age of the scientific documents cited by patents relates to the distance traveled by the technology. To test the robustness of our findings, we calculate two alternative proxies for the science-based nature of the trajectory, such as (4) the citations to scientific articles by the first patent in the trajectory and (5) the share of university-owned patents. We calculate also an alternative proxy for (6) the distance traveled by the technology based on national borders.

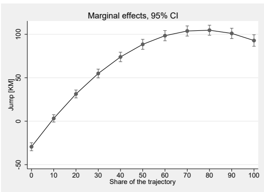
In the first exercise (1), we interact the variable *Share of scientific content* with the share of completion of the trajectory at patent *s-1*. Specifically, we define *Share of the trajectory* as a variable calculated as 100*($s\text{-}1/N_i$), where N_i is the total number of patents in the trajectory *i*. *Share of the trajectory* ranges from 0 to 100 and represents the point where the patent *s-1* is positioned in the development trajectory. Table 6 shows the results of this regression exercise, explaining the *Jump length* (Column 1) and the probability of generating a *New cluster* (Column 2). Based on the regression exercises in Table 6, we report in Figure 4 the estimated marginal effect of *Share of scientific content* for different values of the *Share of the trajectory* variable. The figure illustrates that for a 10 percentage points increase in the *Share of scientific content* in the trajectory, the jump length is associated with a reduction of 26.42 kilometers at the beginning of the trajectory (i.e., *Share of the trajectory* = 0%). For a 10 percentage points increase in the *Share of scientific content* in the trajectory, the jump length is associated with an increment of 88.42 and 103.79 kilometers at the 50% and 70% of the trajectory, respectively. For the same increase in the *Share of scientific content,* the probability of generating a new cluster equals -0.59, +1.71, and +2.17 percentage points at 0%, 50%, and 70% of the trajectory, respectively. The results of this regression exercise are aligned with those discussed in Table 5.

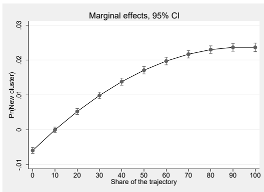
Table 6: Additional regression results substituting the three dummy variables indicating the trajectory phases with the continuous variable *Share of the trajectory*

	(1) OLS Jump length	(2) LPM New Cluster
Share of scientific content	-2.94***	-0.00059***
	(0.23)	(0.000044)
Share of scientific content * Share of the trajectory	0.35***	0.000062***
	(0.011)	(2.0e-06)
Share of scientific content * Share of the trajectory²	-0.0023***	-3.3e-07***
	(0.00010)	(1.9e-08)
Share of the trajectory	4.78***	0.00029**
	(0.71)	(0.00013)
Share of the trajectory²	0.045***	8.6e-06***
	(0.0058)	(1.1e-06)
Log(Stock of inventors)	-2.96***	-0.00048***
	(0.29)	(0.000054)
Share of DE patents	-2.83***	-0.00023***
	(0.22)	(0.000040)
Share of US patents	0.78***	-0.00013**
	(0.27)	(0.000050)
Share of JP patents	-197***	-0.030***
	(3.11)	(0.00058)
Year passed since the entry year	-0.20***	-0.000013***
	(0.0011)	(2.1e-07)
Diameter	-40.5***	-0.040***
	(2.26)	(0.00042)
Clusters generated	4,646***	0.64***
	(72.9)	(0.014)
Constant	-2.94***	-0.00059***
	(0.23)	(0.000044)
Observations	592,173	592,173
Number of trajectories	10,782	10,782
R-squared within	0.219	0.153
Dummy application year	Yes	Yes
Dummy technological classes	No	No
Technological trajectory fixed effects	Yes	Yes

NOTE: We report the R-squared for the within model.

Figure 3: Marginal effects of increasing 10 percentage points the values of the variable *Share of scientific content* on *Jump length* and Probability of generating a *New cluster*





In the second exercise (2), we explore an alternative strategy to estimate the effect of the variable *Share of scientific content* in the three phases of the trajectory. Specifically, we split our study sample into three sub-samples according to the trajectory phase. Table 7 shows the results of the regressions. In each column of Table 7, we report the estimates considering the fully specified model described in Equation 2[12]. Columns 1 and 4 show the results on the sub-sample of patents belonging to the take-off phase, columns 2 and 5 consider the growth phase.

[12]Differently from Equation 2, we exclude the interaction terms between the phases of the trajectory and the variable scientific content that do not make sense when the sample is split according to the trajectory phase.

Finally, columns 3 and 6 report the patents belonging to the maturity phase. Results reported in Table 7 align with those discussed in Table 5.

Table 7: Regression results splitting the sample by trajectory phase

	(1) OLS Jump length	(2) OLS Jump length	(3) OLS Jump length	(4) LPM New cluster	(5) LPM New cluster	(6) LPM New cluster
	Take off	Growth	Maturity	Take off	Growth	Maturity
Share scientific content	-0.26	0.87	3.77**	-0.00016**	0.00041***	0.00086**
	(0.46)	(0.56)	(1.74)	(0.000076)	(0.00012)	(0.00042)
Log(Stock of inventors)	-4.83***	1.42**	9.78***	-0.00071***	0.00060***	0.0028***
	(0.64)	(0.70)	(2.20)	(0.00011)	(0.00015)	(0.00053)
Share of DE patents	-4.12***	-2.35***	-0.82	-0.00037***	-0.0013***	-0.0023***
	(0.48)	(0.53)	(1.68)	(0.000079)	(0.00011)	(0.00041)
Share of US patents	1.78***	-16.0***	-14.3***	-0.00021**	-0.0017***	-0.0026***
	(0.62)	(0.64)	(1.82)	(0.00010)	(0.00014)	(0.00044)
Share of JP patents	-124***	-62.9***	-9.34**	-0.023***	-0.012***	0.0019*
	(6.74)	(4.92)	(4.43)	(0.0011)	(0.0011)	(0.0011)
Years passed since the entry year	-0.40***	-0.17***	-0.27***	-0.000026***	-8.6e-06***	-0.000018***
	(0.0028)	(0.0020)	(0.0041)	(4.7e-07)	(4.4e-07)	(9.8e-07)
Diameter	-220***	-175***	-259***	-0.10***	-0.078***	-0.13***
	(7.87)	(3.66)	(6.50)	(0.0013)	(0.00079)	(0.0016)
Clusters generated	4,119***	4,789***	4,457***	0.55***	0.67***	0.54***
	(121)	(488)	(306)	(0.020)	(0.11)	(0.074)
Constant	-0.26	0.87	3.77**	-0.00016**	0.00041***	0.00086**
	(0.46)	(0.56)	(1.74)	(0.000076)	(0.00012)	(0.00042)
Observations	146,785	301,409	143,979	146,785	301,409	143,979
N. of technological trajectories	10,782	10,782	10,782	10,782	10,782	10,782
R-squared within	0.285	0.056	0.071	0.200	0.052	0.073
Dummy application year	Yes	Yes	Yes	Yes	Yes	Yes
Dummy technology classes	No	No	No	No	No	No
Technological trajectory fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns 1 and 4 report the results on the sub-sample of patents in the take-off phase. Columns 2 and 5 show the results on the sub-sample of patents in the growth phase. Columns 3 and 6 show the results on the sub-sample of patents in the maturity phase. We report the R-squared for the within model.

In the third exercise (3), we dig into the characteristics of the scientific documents cited by patents. Scientists share their achievements and codify their discoveries in scientific publications. We expect that the longer these publications are available to the public, the higher the probability that inventors worldwide become aware of them and incorporate the codified knowledge in their inventions. In other words, we expect the age of the scientific publications cited by patent documents to correlate positively with the distance traveled by technologies. To test our hypothesis, we calculate in three steps the variable *Average citation age*. First, we retrieve the publication year of each scientific article cited in patent documents. Second, we calculate the age of each article cited as the difference between the patent application and the article publication year[13]. Finally, we calculate the *Average citation age* as the average age of the publications cited by all the patents in the trajectory preceding the focal patent *s*, i.e., from patent *1* to patent *s-1*. In our study sample, 525,373 observations (88.72% of the full sample) have in the patent stock, from *1* to *s-1,* at least one patent citing a scientific article and, for those observations, the variable *Average citation age* can be calculated. For the remaining 66,800 observations (11.28% of the full sample) that do not have citations to any scientific article in the patent stock, we assign a value of zero to the variable *Average citation age*, and we calculate the dummy variable *No citations to articles*. The variable *No citations to articles* equals one for

[13] In the case of multiple articles cited by the same patent, we calculate the average citation age.

patent *s* if there are no patent citations to any scientific article in the patent stock from *1* to *s-1*. In the restricted sample of 525,373 observations showing a positive value of the *Average citation age*, the *Average citation age* equals 5.65 years with a standard deviation of 4.00 years. Table 8 reports the results of including the *Average citation age* and *No citations to articles* in our regression exercise. As expected, we observe a positive and significant correlation between the *Average citation age* and *Jump length*. For each additional year in the age of the cited scientific articles in the trajectory, we observe an increase of 4.55 kilometers in the distance traveled by the trajectory (Column 1). Similarly, for each additional year in the age of the cited scientific articles, we observe an increase of 0.073 percentage points in the probability of generating a new cluster (0.66% of the unconditional probability of generating a new cluster).

Table 8: Regression results, including the age of the cited scientific articles

	(1) OLS Jump length	(2) LPM New cluster
Average citation age	4.55***	0.00073***
	(0.99)	(0.00019)
No citations to articles	671***	0.12***
	(14.2)	(0.0026)
Share of scientific content	6.05***	0.0011***
	(0.24)	(0.000045)
Growth	181***	0.022***
	(8.02)	(0.0015)
Maturity	376***	0.060***
	(11.8)	(0.0022)
Log(Stock of inventors)	-219***	-0.021***
	(6.96)	(0.0013)
Share of DE patents	-3.17***	-0.00051***
	(0.29)	(0.000054)
Share of US patents	-2.94***	-0.00025***
	(0.22)	(0.000040)
Share of JP patents	0.046	-0.00029***
	(0.27)	(0.000050)
Year passed since the entry year	-196***	-0.031***
	(3.11)	(0.00058)
Diameter	-0.20***	-0.000014***
	(0.0011)	(2.1e-07)
Clusters generated	-13.2***	-0.033***
	(2.15)	(0.00040)
Constant	3,584***	0.46***
	(73.8)	(0.014)
Observations	592,173	592,173
N. of technological trajectories	10,782	10,782
R-squared within	0.215	0.148
Dummy application year	Yes	Yes
Dummy technology classes	No	No
Technological trajectory fixed effects	Yes	Yes

NOTE: We report the R-squared for the within model.

In the fourth exercise (4), we define an alternative proxy for the science-based nature of the technological trajectory by calculating a dummy variable indicating if the patent at the origin of the trajectory has scientific content or not. Specifically, we define the variable *First patent with scientific content* as a variable equal to one if the first patent in the trajectory has scientific content, zero otherwise. In 38% of the cases, our trajectories originate from a patent with scientific content, and the variable *First patent with scientific content* correlates positively with the explanatory variable *Share of scientific content* (0.54). Table 9 shows the regression results using *First patent with scientific content* as a proxy for the trajectory science-based nature. Results are consistent with the ones reported in Table 5.

Table 9: Regression results including the variable *First patent with scientific content* as proxy for the science-based nature of the trajectory

	(1) OLS Jump length	(2) LPM New cluster
First patent with scientific content * Take off	-91.7***	-0.012***
	(8.58)	(0.0016)
First patent with scientific content * Growth	81.1***	0.012***
	(6.20)	(0.0011)
First patent with scientific content * Maturity	107***	0.019***
	(8.71)	(0.0016)
Growth	-238***	-0.045***
	(7.73)	(0.0014)
Maturity	-51.9***	-0.018***
	(10.8)	(0.0020)
Log(Stock of inventors)	-312***	-0.060***
	(3.23)	(0.00060)
Share of DE patents	-3.88***	-0.00076***
	(0.19)	(0.000035)
Share of US patents	-0.70***	0.00014***
	(0.13)	(0.000025)
Share of JP patents	-1.96***	-0.00045***
	(0.16)	(0.000029)
Year passed since the entry year	-30.6***	-0.0045***
	(0.85)	(0.00016)
Diameter	-0.096***	-8.2e-06***
	(0.00081)	(1.5e-07)
Clusters generated	61.5***	0.0030***
	(1.26)	(0.00023)
Constant	4,793***	0.68***
	(71.2)	(0.013)
Observations	592,173	592,173
N. of technological trajectories	10,782	10,782
R-squared	0.190	0.150
Dummy application year	Yes	Yes
Dummy technology classes	Yes	Yes
Technological trajectory fixed effects	No	No

NOTE: The econometric model does not include *Technological trajectory fixed effects* to avoid collinearity between them and one of the three interaction terms of the variable *First patent with scientific content*.

In the fifth exercise (5), we use the information about the ownership of the patents in a trajectory to construct an alternative proxy for the trajectory's scientific content. We define the variable *Share university*$_{i,s-1}$ as the share of patents in the trajectory that list at least one university applicant. We expect longer jumps and a higher probability of generating new clusters when universities own a high share of inventions in the trajectory. Our technologies sample shows a 7.76% share of university patents in the trajectories. Table 10 shows the results of using *Share university*$_{i,s-1}$ as a proxy for the scientific content of the trajectory. In the growth and maturity phases, we find that the share of university-owned patents in the trajectory is positively associated with the length of the jumps and the probability of generating a new cluster. These results are coherent with the positive impact observed for the trajectory's *Share of scientific content* in Table 5. Indeed, trajectories with a high share of university-owned patents are likely to have high scientific content (the correlation between *Share university* and *Share of scientific content* equals 0.545).

Table 10: Regression results using *Share university* as alternative proxy for scientific content

	(1) OLS Jump length	(2) OLS Jump length	(3) LPM New cluster	(4) LPM New cluster
Share university	3.14***		0.00050***	
	(0.20)		(0.000038)	
Share university * Take off		-0.081		-0.00039***
		(0.36)		(0.000066)
Share university * Growth		7.27***		0.0010***
		(0.52)		(0.000096)
Share university * Maturity		9.69***		0.0019***
		(0.64)		(0.00012)
Growth	-168***	124***	-0.035***	0.011***
	(6.61)	(8.57)	(0.0012)	(0.0016)
Maturity	37.1***	313***	-0.0041**	0.044***
	(9.53)	(12.5)	(0.0018)	(0.0023)
Log(Stock of inventors)	-310***	-265***	-0.059***	-0.029***
	(3.22)	(6.93)	(0.00059)	(0.0013)
Share of DE patents	-3.84***	-3.27***	-0.00075***	-0.00054***
	(0.19)	(0.29)	(0.000035)	(0.000054)
Share of US patents	-0.95***	-2.89***	0.00010***	-0.00024***
	(0.14)	(0.22)	(0.000025)	(0.000040)
Share of JP patents	-1.89***	0.30	-0.00044***	-0.00025***
	(0.16)	(0.27)	(0.000029)	(0.000051)
Year passed since the entry year	-31.1***	-208***	-0.0046***	-0.032***
	(0.85)	(3.10)	(0.00016)	(0.00058)
Diameter	-0.097***	-0.21***	-8.4e-06***	-0.000016***
	(0.00081)	(0.0011)	(1.5e-07)	(2.1e-07)
Clusters generated	62.7***	0.55	0.0032***	-0.031***
	(1.25)	(2.15)	(0.00023)	(0.00040)
Constant	4,735***	4,253***	0.67***	0.58***
	(71.2)	(72.4)	(0.013)	(0.014)
Observations	592,173	592,173	592,173	592,173
N. of technological trajectories	10,782	10,782	10,782	10,782
R-squared / R-squared within	0.190	0.213	0.150	0.146
Dummy application year	Yes	Yes	Yes	Yes
Dummy technology classes	Yes	No	Yes	No
Technological trajectory fixed effects	No	Yes	No	Yes

NOTE: For the regressions in Columns 2 and 4, including technological trajectory fixed effects, we report the R-squared for the within model.

Finally, in the sixth exercise (6), we introduce an alternative dependent variable to *Jump length* and *New cluster*. Instead of using the geographical distance as in the definition of *Jump length* and *New cluster*, we look at the national borders to define the variable *Continent jump*. Specifically, *Continent jump* takes the value 1 if the patent *s* appears in a continent previously unexplored by one technology *i*'s patent from *1* to *s-1*, 0 if *s* appears in a continent already explored by *i*. In our sample of technologies, we observe a 4.21% probability of a continent jump at patent *s*. Table 11 shows the regression results where we use as the dependent variable *Continent jump*, and we maintain the same econometric specification as in Table 5 for the rest. *Share of scientific content* is positively associated with a higher probability of observing a continent jump by the technology, especially in the late technology development phases. Results reported in Table 11 using *Continent jump* as dependent variable align with those obtained in Table 5 using *Jump length* and *New cluster* as dependent variables.

Table 11: Regression results using continent jump as dependent variable.

	(1) OLS Continent jump	(2) OLS Continent jump
Share of scientific content	0.000062**	
	(0.000026)	
Share of scientific content * Take off		-0.000089***
		(0.000027)
Share of scientific content * Growth		0.00071***
		(0.000033)
Share of scientific content * Maturity		0.00089***
		(0.000036)
Growth	0.019***	-0.0098***
	(0.00098)	(0.0013)
Maturity	0.043***	0.0030*
	(0.0014)	(0.0018)
Log(Stock of inventors)	-0.025***	-0.026***
	(0.00084)	(0.00084)
Share of DE patents	-0.00044***	-0.00042***
	(0.000036)	(0.000036)
Share of US patents	-0.00031***	-0.00030***
	(0.000026)	(0.000026)
Share of JP patents	-0.000084**	-0.000022
	(0.000033)	(0.000033)
Year passed since the entry year	-0.026***	-0.025***
	(0.00037)	(0.00037)
Diameter	-0.024***	-0.022***
	(0.00038)	(0.00038)
Clusters generated	0.0063***	0.0031***
	(0.00026)	(0.00027)
Constant	0.45***	0.46***
	(0.0089)	(0.0089)
Observations	592,173	592,173
N. of technological trajectories	10,782	10,782
R-squared / R-squared within	0.171	0.173
Dummy application year	Yes	Yes
Dummy technology classes	No	No
Technological trajectory fixed effects	Yes	Yes

NOTE: We report the R-squared for the within model.

5. Discussion and conclusion

Technologies evolve along different trajectories due to a cumulative reuse process in which follow-on inventions build on the same technological idea. The inventive activities related to a specific technology might aggregate in local clusters or spread around the globe. This paper reconstructs technological trajectories and relates their geographical evolution to their scientific content. We measure the distance traveled by the technology, calculating how far in space inventors currently using the technology are located from the closest inventors who previously used the technology. We also observe the probability that inventors currently utilizing the technology generate a new cluster by locating far away from the previous inventors using the technology. We find that technologies travel longer distances when characterized by a high scientific content, but only in their growth or maturity phases. Looking at cluster formation, in the early phases, scientific content negatively affects the probability of generating a new cluster, while in the growth and maturity phases, the scientific content increases the probability of generating a new cluster.

Overall, our result, showing that science-based technologies travel longer distances than non-science-based technologies, aligns with the ones of Sorenson and Fleming (2004). While Sorenson and Fleming (2004) consider patented inventions as unrelated entities, we link patents embedding the same technological idea. By doing so, we follow each technology along its development process and disentangle the effect of science in the different evolution phases. Noteworthy, Sorenson and Fleming (2004)'s findings hold only when the technological trajectory is in its maturity and growth, while technologies behave differently in their early stages. We confirm that scientific publications stimulate the spread of technologies. However, there are important nuances in the average effect related to the peculiarities of the scientific knowledge diffusion processes. The knowledge publications' content needs time to be absorbed, and delays in scientific knowledge absorption translate into the tendency of technologies with scientific content to stay aggregated in a geographically restricted area in their early phases. We can claim that technologies fly on the wings of science only in their growth and maturity phases when the community of inventors has assimilated the scientific content. Our intuition that scientific knowledge absorption impacts the spread of technologies is confirmed when examining the age of the publications on which patents are built. Indeed, we find that the age of the scientific publications cited by patent documents correlates positively with the distance traveled by technologies.

As a general conclusion, we argue that technologies with science-based and non-science-based content develop according to two different models. Science-based technologies show a diffused development model with clusters of inventors emerging around the world. On the contrary, non-science-based technologies tend to follow a more agglomerated development model, traveling shorter distances and remaining concentrated in a few clusters.

From a policy perspective, our results are crucial in guiding policymakers' decisions when supporting the development of technologies. For example, in regional planning, policymakers who plan to support the development of a technology with high science content, like the transgenic mammal technology, know upfront that such technologies will develop worldwide. They can plan the infrastructure accordingly and, for instance, support agreements with other countries to develop large consortia to exploit the synergies derived from cross-countries collaborations. On the contrary, policymakers interested in developing in their region a non-science-based technology should tailor their investments considering that the technology development will stay geographically localized, and their territories can internalize the technology development. Similar reasoning should apply to private companies and organizations when promoting the development of their technologies' portfolios.

References

Arthur, W.B., 2009. The nature of technology: What it is and how it evolves. Simon and Schuster.

Arthur, W.B., 2007. The structure of invention. Research Policy 36, 274–287.

Balland, P.-A., Boschma, R., Frenken, K., 2015. Proximity and Innovation: From Statics to Dynamics. Regional Studies 49, 907–920. https://doi.org/10.1080/00343404.2014.883598

Bikard, M., Marx, M., 2020. Bridging Academia and Industry: How Geographic Hubs Connect University Science and Corporate Technology. Management Science 66, 3425–3443. https://doi.org/10.1287/mnsc.2019.3385

Breschi, S., Lissoni, F., 2001. Knowledge spillovers and local innovation systems: a critical survey. Industrial and corporate change 10, 975–1005.

Cowan, R., Foray, D., 1997. The Economics of Codification and the Diffusion of Knowledge. Industrial and Corporate Change 6, 595–622. https://doi.org/10.1093/icc/6.3.595

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research policy 23, 487–521.

de Rassenfosse, G., Kozak, J., Seliger, F., 2019. Geocoding of worldwide patent data. Sci Data 6, 260. https://doi.org/10.1038/s41597-019-0264-6

Feldman, M.P., Kogler, D.F., 2010. Chapter 8 - Stylized Facts in the Geography of Innovation, in: Rosenberg, B.H.H. and N. (Ed.), Handbook of the Economics of Innovation, Handbook of The Economics of Innovation, Vol. 1. North-Holland, pp. 381–410.

Feldman, M.P., Kogler, D.F., Rigby, D.L., 2015. rKnowledge: The Spatial Diffusion and Adoption of rDNA Methods. Regional Studies 49, 798–817. https://doi.org/10.1080/00343404.2014.980799

Fontana, R., Nuvolari, A., Verspagen, B., 2009. Mapping technological trajectories as patent citation networks. An application to data communication standards. Economics of Innovation and New Technology 18, 311–336. https://doi.org/10.1080/10438590801969073

Graziano, M., Gillingham, K., 2015. Spatial patterns of solar photovoltaic system adoption: The influence of neighbors and the built environment. Journal of Economic Geography 15, 815–839. https://doi.org/10.1093/jeg/lbu036

Hanahan, D., Wagner, E.F., Palmiter, R.D., 2007. The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. Genes & Development 21, 2258–2270. https://doi.org/10.1101/gad.1583307

Howells, J.R., 2002. Tacit knowledge, innovation and economic geography. Urban studies 39, 871–884.

Jaffe, A.B., Trajtenberg, M., Fogarty, M.S., 2000. The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees (Working Paper No. 7631). National Bureau of Economic Research. https://doi.org/10.3386/w7631

Klevorick, A.K., Levin, R.C., Nelson, R.R., Winter, S.G., 1995. On the sources and significance of interindustry differences in technological opportunities. Research Policy 24, 185–205. https://doi.org/10.1016/0048-7333(93)00762-I

Kuhn, T.S., 1962. The structure of scientific revolutions. University of Chicago press.

Marx, M., Fuegi, A., 2020. Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations (No. w27987). National Bureau of Economic Research. https://doi.org/10.3386/w27987

Merton, R.K., 1968. The Matthew effect in science. Science 159, 56–63.

Merton, R.K., 1942. The Sociology of Science: Theoretical and Empirical Investigations. University of Chicago Press.

Morrison, G., Riccaboni, M., Pammolli, F., 2017. Disambiguation of patent inventors and assignees using high-resolution geolocation data. Scientific Data 4, 1–21. https://doi.org/10.1038/sdata.2017.64

Murray, F., 2010. The oncomouse that roared: Hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions. American Journal of sociology 116, 341–388.

Nelson, R.R., Winter, S.G., 1982. An evolutionary theory of economic change, digitally reprinted. ed. The Belknap Press of Harvard Univ. Press, Cambridge, Mass.

Nomaler, Ö., Verspagen, B., 2016. River deep, mountain high: of long run knowledge trajectories within and between innovation clusters. J Econ Geogr lbw035. https://doi.org/10.1093/jeg/lbw035

Pezzoni, M., Veugelers, R., Visentin, F., 2022. How fast is this novel technology going to be a hit? Antecedents predicting follow-on inventions. Research Policy 51, 104454. https://doi.org/10.1016/j.respol.2021.104454

Rosenberg, N., Nelson, R.R., 1994. American universities and technical advance in industry. Research policy 23, 323–348.

Sorenson, O., Fleming, L., 2004. Science and the diffusion of knowledge. Research Policy 33, 1615–1634. https://doi.org/10.1016/j.respol.2004.09.008

Stephan, P.E., 2012. How economics shapes science. Harvard University Press, Cambridge, Mass.

Strumsky, D., Lobo, J., 2015. Identifying the sources of technological novelty in the process of invention. Research Policy 44, 1445–1461. https://doi.org/10.1016/j.respol.2015.05.008

Tijssen, R.J.W., 2002. Science dependence of technologies: evidence from inventions and their inventors. Research Policy 31, 509–526. https://doi.org/10.1016/S0048-7333(01)00124-X

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

Verspagen, B., 2007. Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research. Advances in Complex Systems 10, 93–115. https://doi.org/10.1142/S0219525907000945

Von Hippel, E., 1998. Economics of product development by users: The impact of “sticky” local information. Management science 44, 629–644.

WIPO, 2019. Guide to the International Patent Classification.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. Science 316, 1036–1039. https://doi.org/10.1126/science.1136099

The UNU-MERIT WORKING Paper Series

2023-01 *Can international mobility shape students' attitudes toward inequality? The Brazilian case* by Cintia Denise Granja, Fabiana Visentin and Ana Maria Carneiro

2023-02 *Demand-led industrialisation policy in a dual-sector small open economy* by Önder Nomaler, Danilo Spinola and Bart Verspagen

2023-03 *Reshoring, nearshoring and developing countries: Readiness and implications for Latin America* by Carlo Pietrobelli and Cecilia Seri

2023-04 *The role of product digitization for productivity: Evidence from web-scraping European high-tech company websites* by Torben Schubert, Sajad Ashouri, Matthias Deschryvere, Angela Jäger, Fabiana Visentin, Scott Cunningham, Arash Hajikhani, Lukas Pukelis and Arho Suominen

2023-05 *More than a feeling: A global economic valuation of subjective wellbeing damages resulting from rising temperatures* by Stephan Dietrich and Stafford Nichols

2023-06 *Was Robert Gibrat right? A test based on the graphical model methodology* by Marco Guerzoni, Luigi Riso and Marco Vivarelli

2023-07 *Predicting social assistance beneficiaries: On the social welfare damage of data biases* by Stephan Dietrich, Daniele Malerba and Franziska Gassmann

2023-08 *Related or unrelated diversification: What is smart specialization?* By Önder Nomaler and Bart Verspagen

2023-09 *Breach of academic values and digital deviant behaviour: The case of Sci-Hub* by Giulia Rossello and Arianna Martinelli

2023-10 *The effect of lobbies' narratives on academics' perceptions of scientific publishing: An information provision experiment* by Giulia Rossello and Arianna Martinelli

2023-11 *Making impact with agricultural development projects: The use of innovative machine learning methodology to understand the development aid field* by Lindsey Moore, Mindel van de Laar, Pui Hang Wong and Cathal O'Donoghue

2023-12 *Green windows of opportunity in the Global South* by Rasmus Lema and Roberta Rabellotti

2023-13 *The green and digital transition in manufacturing global value chains in latecomer countries* by Rasmus Lema and Roberta Rabellotti

2023-14 *Impact of prestigious-STEM Education of corporate board members on innovation effort: Evidence from India* by Rituparna Kaushik, Sourabh Bikas Paul and Danilo Spinola

2023-15 *Reducing environmental impact through shared ownership: A model of consumer behaviour* by Francesco Pasimeni and Tommaso Ciarli

2023-16 *Welfare losses, preferences for redistribution, and political participation: Evidence from the United Kingdom's age of austerity* by Patricia Justino, Bruno Martorano and Laura Metzger

2023-17 *Inequality, social mobility and redistributive preferences* by Isabel Günther and Bruno Martorano

2023-18 *Automation-induced reshoring and potential implications for developing economies* by Hubert Nii-Aponsah, Bart Verspagen and Pierre Mohnen

2023-19 *Ethnic spatial dispersion and immigrant identity* by Amelie F. Constant, Simone Schüller and Klaus F. Zimmermann

2023-20 *Public R&D and Growth: A dynamic Panel VECM Data Analysis for 14 OECD Countries* by Thomas H.W. Zieseme by r

2023-21 *Innovation and the labor market: Theory, evidence and challenges* by Nicoletta Corrocher, Daniele Moschella, Jacopo Staccioli and Marco Vivarelli

2023-22 *Revisiting Schumpeter in Europe: Place-based innovation and transformative industrial policy* by Luc Soete and Johan Stierna

2023-23 *Identification of Fourth Industrial Revolution technologies using PATSTAT data* by María de las Mercedes Menéndez, Önder Nomaler and Bart Verspagen

2023-24 *Expectations and the stability of stock-flow consistent models* by Huub Meijers, Joan Muysken and Giulia Piccillo

2023-25 *Critical minerals and countries' mining competitiveness: An estimate through economic complexity techniques* by Jorge Valverde Carbonell, María de las Mercedes Menéndez, and Carlo Pietrobelli

2023-26 *Internal rates of return for public R&D from VECM estimates for 17 OECD countries* by Thomas H.W. Ziesemer

2023-27 *Does moral transgression promote anti-social behavior? Evidence from lab-in-the-field experiments* by Halefom Yigzaw Nigus, Eleonora Nillesen, Pierre Mohnen and Salvatore Di Falco

2023-28 *Does exposure to markets promote investment behavior? Evidence from rural Ethiopia* by Halefom Yigzaw Nigus, Pierre Mohnen and Eleonora Nillesen

2023-29 *The virtuous loop of quality of government (QoG) and institutional trust in OECD countries, 2006-2021 and culture* by Ahmed Hussain and Jo Ritzen

2023-30 *Alternative related variety, macroeconomic factors and diversification: Extractive vs. non-extractive products* by Beatriz Calzada Olvera

2023-31 The effect of government cuts of doctoral scholarships on science by Giulia Rossello

2023-32 *Spatial inequality during the COVID-19 pandemic in Africa using night-time lights data* by Bruno Martorano, Elena Perra and Marco Tiberti

2023-33 *How humanitarian are Germans towards refugees?* By Tobias Hillenbrand, Bruno Martorano, Laura Metzger and Melissa Siegel

2023-34 *The gender inequality effects of the COVID-19 pandemic on the Nigerian labor market* by Ana Karen Diaz Mendez and Bruno Martorano

2023-35 *Lost in aggregation? On the importance of local food price data for food poverty estimates* by Stephan Dietrich, Valerio Giuffrida, Bruno Martorano, Georg Schmerzeck and Marco Tiberti

2023-36 *Technologies fly on the wings of science: Exploring the role of science in technologies' spatial evolution* by Michele Pezzoni, Reinhilde Veugelers and Fabiana Visentin