



The evolution of networks of innovators within and across borders: Evidence from patent data



Andrea Morescalchi^{a,b,*}, Fabio Pammolli^a, Orion Penner^a, Alexander M. Petersen^a, Massimo Riccaboni^{a,c}

^a IMT Institute for Advanced Studies Lucca, Piazza San Ponziano 6, 55100 Lucca, Italy

^b Econometrics and Applied Statistics Unit, Joint Research Centre, European Commission, Via Enrico Fermi 2749, 21027 Ispra (Varese), Italy

^c Department of MSI, KU Leuven, Leuven, Belgium

ARTICLE INFO

Article history:

Received 24 June 2013

Received in revised form

29 September 2014

Accepted 20 October 2014

Available online 22 November 2014

Keywords:

Geography of knowledge

Networks of innovators

European integration

Spatial proximity

Cross-border collaboration

Gravity model

ABSTRACT

Recent studies on the geography of knowledge networks have documented a negative impact of physical distance and institutional borders upon research and development (R&D) collaborations. Though it is widely recognized that geographic constraints and national borders impede the diffusion of knowledge, less attention has been devoted to the temporal evolution of these constraints. In this study we use data on patents filed with the European Patent Office (EPO) for OECD countries to analyze the impact of physical distance and country borders on inter-regional links in four different networks over the period 1988–2009: (1) co-inventorship, (2) patent citations, (3) inventor mobility and (4) the location of R&D laboratories. We find the constraint imposed by country borders and distance decreased until mid-1990s then started to grow, particularly for distance. The intensity of European cross-country inventor collaborations increased at a higher pace than their non-European counterparts until 2004, with no significant relative progress thereafter. For geographical networks of mobility, R&D activities and patent citations we cannot detect any substantial progress in European research integration above and beyond the common global trend.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Rapid progress in information, communication, and transportation technologies and the overall trend of globalization have led to the assertion “distance is dead” (Castells, 1996; Cairncross, 1997). A natural tension exists, however, between this view and knowledge “stickiness”: human activities and social interactions are known to geographically cluster to take advantage of knowledge spillovers, social capital and other agglomeration economies (Feldman, 1994). While the literature on innovation systems has focused on the interplay between clusters and networks of innovators (Breschi and Malerba, 2005), the “death of distance” conjecture has been thoroughly investigated in the literature on international trade and globalization studies. The most significant recent advances in that vein have been made by means of panel gravity regressions and

indicate distance, borders and free trade areas still play a key role in trade networks (the so-called “tyranny of distance”).

Since the seminal contribution of Freeman (1991), networks of innovators have attracted a great deal of interest as a tool for representing and analyzing division of innovative labor. Many types of network have been investigated, ranging from the informal scientific connections in invisible colleges and communities of practice, to the formal collaborative agreements between firms and other research organizations. With increasing frequency, growing data on scientific collaborations, collaborative R&D projects, and patents have been widely exploited to gain insight into the structure and evolution of networks in different industries, countries and time frames (see Powell and Grodal, 2005 and Ozman, 2009 for reviews).

Despite significant efforts in a growing body of literature analyzing networks of innovators, there is still a lack of large-scale quantitative understanding of the evolution of networks of innovators in space and time. The complexity of this problem arises from the variety of competing forces that underlie the economics and sociology of R&D collaboration. Prevailing wisdom states the spread of tacit knowledge and the formation of informal ties are uninhibited over short distances, but barriers increase with distance. However, for the transmission of codified knowledge and formal

* Corresponding author at: Econometrics and Applied Statistics Unit, Joint Research Centre, European Commission, Via Enrico Fermi 2749 T.P. 361, 21027 Ispra (Varese), Italy.

E-mail addresses: andrea.morescalchi@jrc.ec.europa.eu, andrea.morescalchi@mail.com (A. Morescalchi).

contractual collaborations, distance plays less of a role at large scales, even if national borders between different institutional settings can still reduce the effectiveness of contractual solutions. On the one hand, because technological advancements have increased the capacity to codify and share knowledge across large distances, it follows that the barriers induced by distance should be decreasing, and possibly vanishing, in R&D networks. On the other hand, innovators are increasingly attracted by large innovation “hubs” which combine local agglomeration economies with centrality advantages in knowledge and social networks also known as preferential attachment or “rich get richer” effect. The effect of globalization forces, which reduce the cost to collaborate at distance, compounds with the attractiveness of central regions in innovation networks, which tend to limit the geographical span of collaboration of smaller regions in their gravitation fields. Innovation networks can be more or less constrained by geographical distance depending on the interplay of these two forces. Also, the dynamic role of physical distance and institutional borders may differ significantly across different R&D networks depending upon the type of knowledge that is exchanged (tacit vs. codified) and the nature of the links: arm’s length market transactions, hierarchical relations or network forms of coordination (Whittington et al., 2009).

Cross-network interdependencies should also be taken into account. For example, international mobility should have a positive impact on regional citation flows as inventor movement is thought to be an important driver of knowledge spillovers (Agrawal et al., 2006). International mobility may, in turn, have a positive impact on large distance collaborations as mobile inventors act as bridges across teams of inventors working for different organizations (Breschi and Lissoni, 2009). Conversely, one could argue that, the more individual inventors and research teams can freely move, the less R&D organizations will feel the need to locate R&D labs abroad or to sign collaborative agreements with foreign partners. In this sense, understanding the extent to which globalization reduces constraints on geographical mobility is important for assessing side-effects in other dimensions of R&D networks.

Here we employ a gravity approach to quantify simultaneously the strength of borders and distance on multiple innovation networks constructed from about 2.4 million patents recorded by the European Patent Office (EPO). We analyze a large sample of developed nations over many years to investigate the dichotomy arising from localizing constraints of R&D spillovers and agglomeration economies in R&D clusters vis-à-vis the tendency to expand R&D networks via long-range collaborations.

Our study moves beyond previous efforts to understand the geography of research collaboration in many respects. First, we study a large set of developed countries at a low level of spatial aggregation. Second, we analyze a set of interrelated patent networks using the same analytic approach. Third, in our analysis we investigate, jointly, the physical distance effect and the country-border effect on a global scale. Few previous studies have investigated the dynamics of the geographical distance and border-effects simultaneously, focusing either on Europe (Hoekman et al., 2010) or the United States (Singh and Marx, 2013). Fourth, we focus on the evolution of barriers to the internationalization of knowledge networks over two decades, whereas most of previous studies in this field have had to take a static viewpoint. Our more comprehensive analysis allows us to examine and ultimately quantify the effect of European integration efforts in the context of global network evolution trends.¹

¹ Overall, our approach poses some numerical challenges since we analyze a network of around 2.7 billion links: 5552^2 (potential links between 5552 regional nodes) \times 4 (networks) \times 22 (years) = 2,712,573,952 possible connections.

Beyond scientific relevance of developing methods to quantify R&D network evolution, a better understanding of how distance and borders influence the structure and evolution of R&D networks is important to orient the policy debate. In particular, the European Research Area (ERA) vision of an “open space for knowledge and growth” stands as the most recent in a long line of integration efforts within the European Union (EU). The establishment of the ERA has been highlighted as key component of the competitiveness of the EU’s Europe 2020 growth strategy. This is an attempt to reduce, perhaps even eliminate, the effect of national borders on scientific and R&D networks to create an area in which ideas and high skill human capital are free to flow and capitalize on transnational synergies and complementarities. Furthermore, a better understanding of the role distance and borders play in the structure and evolution of networks of innovators is key, not only for crafting effective policy, but even more simply, for assessing the true effectiveness of past, present, and future policy measures.

This paper proceeds in the following manner. Section 2 presents a review of the relevant literature. In Section 3 we describe the data and methodology used. Section 4 presents the results of our analysis. Finally, in Section 5 we discuss our results and natural extensions of this research direction deriving some policy implications for the European Research Area.

2. The role of geography in networks of innovators

Regional networks of knowledge portray the knowledge exchange between entities located in different regions. A knowledge exchange is put into effect whenever the benefits that individuals or organizations receive from accessing new pieces of knowledge are greater than the associated costs. In regional networks, the costs of accessing remote knowledge are related to different forms of proximity that can be summarized in five dimensions: physical, institutional, cognitive, social and organizational (Boschma, 2005). The likelihood of a knowledge exchange is positively affected by these different forms of proximity. Most previous analyses of the globalization of knowledge production have focused on two specific spatial biases. First, the degree to which travel and communication costs result in physical distance being an impediment to collaboration. Second, the extent to which institutional friction arising from country-to-country differences create challenges for collaboration across national systems of innovation (Freeman, 1995; Lundvall, 1992; Nelson, 1993; Gertler, 1995). This second aspect has been investigated with particular reference to the European Union, where specific actions have been taken to favor scientific and technological collaboration between Member States. Table 1 summarizes previous findings on spatial biases in innovation networks.

Geographical networks of knowledge can be modeled according to a number of empirical strategies (Broekel et al., 2013). Application of gravity models to scientific and technological collaboration have provided strong evidence for a negative effect of physical distance and country borders on the likelihood of collaboration (Ponds et al., 2007; Maggioni and Uberti, 2007; Scherngell and Barber, 2009, 2011; Frenken et al., 2009; Hoekman et al., 2009, 2010, 2013; Scherngell and Hu, 2011; Scherngell and Lata, 2012; Pan et al., 2012). This body of evidence is robust over various kinds of data (scientific publications, patents), the type of network (collaborations between individuals/institutions, citations, labor mobility) and the geographic unit of analysis (country, regional, sub-regional).

When focusing on the evolution of spatial biases, the prevailing wisdom is that globalization and advances in information, transportation, and communication technologies should reduce the role of distance in socio-economic interactions (Castells, 1996;

Table 1
Literature survey, selected contributions.

Author (year)	Data source	Time frame	Network	Method	Distance	Border	EU
Aldieri (2011)	USPTO	1979–2002	C	Poisson, ll, cs	-, ↗
Belderbos et al. (2013)	EPO	1995–2002	F	OLS, NUTS0, s
Cassi and Plunket (2012)	EPO	1990–2006	A	r.e. Logit, NUTS3, s
Chessa et al. (2013)	EPO, WoS	1985–2010	A, C, M	ZINB(g), NUTS3, d	=, =
Fischer et al. (2006)	EPO	1985–2002	C	Poisson(g), NUTS2, s
Frenken et al. (2009)	WoS	1988–2004	A	ZINB(g), NUTS3, s	n.s., ..
Hoekman et al. (2009)	EPO, WoS	1988–2004	A	ZINB(g), NUTS3, s
Hoekman et al. (2010)	WoS	2000–2007	A	NB(g), NUTS2, cs	-, ↗
Hoekman et al. (2013)	WoS, EUPRO	2000–2007	A	NB(g), NUTS2, d	+, ..
LeSage et al. (2007)	EPO	1985–2002	C	Poisson s.i.(g), NUTS2, s
Maggioni and Uberti (2007)	EPO	1998–2002	A	OLS(g), NUTS2, s
Maurseth and Verspagen (2002)	EPO	1979–1996	C	Tobit/NB, NUTS1/2, s
Miguélez and Moreno (2013a)	EPO	1996–2001	A, M	OLS-ZINB, NUTS2, cs	-, ↗	-, ↘	..
Montobbio and Sterzi (2013)	USPTO	1990–2004	A, F	Poisson(g), NUTS0, d	-, =
Paci and Usai (2009)	EPO	1978–1998	C	OLS-Poisson, NUTS2, cs	-, ↗	-, ↘	..
Peri (2005)	USPTO	1975–1996	C	NB, NUTS1, s	-, ↗
Picci (2010)	EPO	1990–2005	A, F	OLS/Poisson(g), NUTS0, cs	-, ↗	..	+, ↗
Ponds et al. (2007)	WoS	1988–2004	A	NB(g), NUTS3, s
Ponds (2009)	WoS	1988–2004	A	Probit, .., d, =	..
Schergell and Hu (2011)	CNKI	2007	A	Poisson, NB(g), .., s
Singh and Marx (2013)	USPTO	1975–2004	C	WESML, CBSA, d	-, ↗	-, ↗	..

Note: A: co-authorship; C: citations; CBSA: US core based statistical area; CNKI: China National Knowledge Infrastructure; cs: comparative statics; d: dynamics; EPO: European Patent Office; EUPRO: EUPRO database; F: location of firm R&D labs; (g): gravity model; ll: longitude-latitude; M: mobility; NB: Negative Binomial; n.s.: not significant; NUTS: Nomenclature of Territorial Units for Statistics; OLS: Ordinary Least Square; Poisson: Poisson model; ZINB: Zero Inflated Negative Binomial; ZIP: Zero Inflated Poisson; r.e.: rare event; s: statics; s.i.: spatial interaction; USPTO: US Patent and Trademark Office; WESML: weighted exogenous sampling maximum likelihood; WoS: Web of Science; ↗: increasing in absolute terms; ↘: decreasing in absolute terms; =: approximately constant.

Cairncross, 1997). This issue has been thoroughly explored in the literature on trade through the lens of gravity models (Coe, 2002; Brun et al., 2005).

Perhaps counter intuitively, results obtained from a wide variety of approaches and data have led to an emerging consensus that distance still plays an important, if not increasing, role in constraining trade flows. In the context of the geography of R&D networks, given the rapid decrease in the costs of coordinating R&D activities over relatively long distances, possibly remotely, coupled with overall globalization, it is widely assumed that the bias to collaborate domestically and at short distance has been decreasing. This may be especially true for European countries, where specific steps have been taken at the political level to stimulate integration in R&D. Several studies have found evidence for a decreasing burden of distance and borders interpreted as the result of directed policies. However, other scholars have presented conflicting results (see Frenken et al., 2009a for a survey).

The conclusion that spatial biases are attenuating is the general interpretation invoked when observing an increase in the cross-border shares of collaborations and an increase in the average distance of collaborations. However, using different methodologies, some studies provide contrary evidence that the constraint of distance is becoming *more* binding over time (Hoekman et al., 2010; Singh and Marx, 2013; Sonn and Storper, 2008; Boerner et al., 2006) and that the country-border effect is *not* decreasing (Singh and Marx, 2013; Ponds, 2009; Frenken, 2002).

Among all studies examining the dynamics of spatial biases, two in particular are worth mentioning in the context of our analysis. Hoekman et al. (2010) estimate gravity models using data on co-publications between NUTS2 regions in 33 European countries for the period 2000–2007. They find that the negative effect of distance on inter-regional collaborations increases over the focus period and that the country-border effect decrease, though not statistically significantly.² Singh and Marx (2013) analyze citations to US patents applied for over the period 1975–2004, with the United States Patent and Trademark Office (USPTO). Using an approach

in which the unit of analysis are pairs of patents representing actual and potential citations, and the probability of observing an actual citation is modeled with a weighted logistic regression, they observe an increase over time in the citations received from US patents relative to citations received from non-US patents.³ That study also finds that the rate of decay in the probability of citation as a function of distance has slightly increased over time. That is to say, the effect of distance is increasing.

Here we aim to bring coherence to the issue of the dynamics of distance and borders by considering a broad range of countries (EU and non-EU) and by applying our methodology to four different R&D networks, each with their own dependence upon tacit vs. codified knowledge, thus providing an overall consistency check for our results. More specifically, we study 50 OECD and OECD-partner countries at the NUTS3 level of spatial aggregation.⁴ Most previous studies used NUTS2 and the few that used NUTS3 focused on a single country (Ponds et al., 2007; Frenken et al., 2009) or a few countries (an exception being Hoekman et al., 2009 who analyze EU27 countries plus Norway and Switzerland). We apply the same analytic approach to four patent networks: (I) the network of patent co-inventorship, (II) the location of R&D labs (the applicant-inventor network), (III) patent citations and (IV) inventor mobility. This is in contrast to previous studies that generally focused only on one network at a time (see Table 1).

(I) *The patent citation network.* Since the pioneering work of Jaffe et al. (1993), patent citations have been utilized extensively to measure the diffusion of knowledge across a variety of dimensions: geographic space, time, technological fields, organizational boundaries, alliance partnerships, and social networks (Almeida and Kogut, 1999; Jaffe and Trajtenberg, 2002; Peri, 2005;

³ They also find that the state-border effect decreases over time as consistent with Hoekman et al. (2010).

⁴ The Nomenclature of Units for Territorial Statistics (NUTS) is a geo-code standard for referencing the subdivisions of countries for statistical purposes. The nomenclature has been introduced by the EU for its member states. The OECD provides an extended version of NUTS, called Territorial Levels (TLs), for its non-EU member and partner states. For European countries, TL2 and TL3 are largely consistent with the Eurostat classifications NUTS2 and NUTS3 (Maraut et al., 2008).

² However they find a decreasing border effect with respect to regional borders.

Gomes-Casseres et al., 2006). A principal assumption underlying this approach is that citations trace out knowledge flows and technological learning as knowledge embedded in the cited patent is transmitted to inventors of the citing patent. Given that access to codified knowledge typically does not require interaction between individuals, it is recognized that distance and institutional borders should be relatively less important in this network. Such studies focus on citations as means to transfer codified knowledge but acknowledge that citations are less effective means of spreading tacit knowledge than personal, face-to-face contacts.

(II) *The co-inventor collaboration network.* Though many empirical studies have analyzed the role of patent citations as measures of knowledge flows, it has also been stressed that economic agents can access knowledge from many sources other than just codified knowledge. In particular, a distinction between two means of spreading tacit knowledge has been made in the literature, operating either through informal social interactions, arm-length market-based relationships, inter-organizational alliances, or hierarchical solutions within R&D organizations. Examples of the first case are social ties with current and former colleagues and those developed in social events (conferences, membership in professional associations, etc.). Geography is relevant here as proximity facilitates the development of social relationships and raises incentives to invest in social capital (Agrawal et al., 2006). In the second case, the transmission of knowledge is regulated by a contract, such as a labor contract, licensing or formal collaborations, which explicitly set a compensation for the exchange of knowledge (Breschi and Lissoni, 2009). Geography matters either because labor mobility among different institutions or laboratories can be constrained in space, or because formal agreements require frequent interactions and monitoring that are more easily conducted locally. The network of co-inventions stands somehow in between these two categories as either the collaboration can be ruled by a formal agreement or inventors can decide to collaborate informally with colleagues located in different areas. The co-inventor network is affected by geography as spatial proximity and co-location may facilitate the transfer of complex knowledge and frequent face-to-face interactions maybe required. Though easing of communication and travel constraints is expected to reduce the importance of spatial proximity in this network (Giroud, 2013), the result can depend on the degree of complementarity between remote and face-to-face interactions.

The popularity of patent citations and collaborations as representations of knowledge flows is likely due to the pursuit in economics of pure externalities (spillovers), *i.e.* a transfer of knowledge which is not mediated by the market (Breschi and Lissoni, 2009). On the other hand, the next two networks we present operate through market-based channels. Specifically, these networks capture the relationships between organizations and affiliated inventors, and the mobility of inventors moving across organizations or across regional laboratories within the same institution.⁵

(III) *The location of R&D labs.* The geographical links between applicants and affiliated inventors is relevant to the analysis of the geographic distribution and globalization of the innovative activities of firms (see Keller, 2004 and Narula and Zanfei, 2005 for surveys). Multinational firms are well known to be drivers of the internationalization of innovation activities (see Wolfmayr et al., 2013) as international location of a firm's subsidiaries facilitates knowledge transfer across borders. The literature on the

internationalization of business suggests a number of different reasons for undertaking technological activities outside the home country (Dunning and Lundan, 2009; Florida, 1997; von Zedtwitz and Gassmann, 2002). Among these, knowledge-seeking motives such as proximity to university and innovative firms as a means to benefit from spillovers and agglomeration advantages, and access to high quality scientific and technical talent, have become considered extremely relevant since the late 1990s (Florida, 1997; von Zedtwitz and Gassmann, 2002; Patel and Vega, 1999; Granstrand, 1999). For example, von Zedtwitz and Gassmann (2002) show that these knowledge related factors are by far the most important motives for performing "research" (rather than "development") activities at foreign locations. Indeed, localized foreign knowledge that is tacit can be accessed or imported for firms by moving closer to the source. This goal can be achieved by setting up subsidiaries abroad (Phene and Almeida, 2003) and by hiring scientists (learning-by-hiring), or by sending a firm's scientist abroad to the subsidiaries (Kim et al., 2009). Evidence from the international business literature suggests that knowledge outflows from the multinational corporation's home base are outweighed by inflows from its foreign-based subsidiaries (Singh, 2007; Kogut and Zander, 1993; Dunning, 1992), and that both knowledge flows appear to follow from personnel flows (Singh, 2007). Focusing on the location of patent inventors is not a novel way to map the geographical distribution of a firm's innovation activities (Cantwell, 1989) but has attracted less attention than it deserves due to data limitations (Harhoff and Thoma, 2010). Patents are extensively used as indicator of the location of firms' R&D activities since systematic firm-level data on R&D expenditures by location are either not collected or not available for analysis. Prior studies have shown the existence of a home country bias in R&D thus proving the existence of strong institutional barriers to the internationalization of R&D activities (Belderbos et al., 2013).

(IV) *The inventor mobility network.* Inventor mobility data can be used to measure the geographical distribution of knowledge spillovers (Breschi and Lissoni, 2009; Almeida and Kogut, 1999; Kim et al., 2009; Agrawal et al., 2006; Miguèlez and Moreno, 2013a). Mobile individuals are endowed carriers of knowledge stock and play a key role in the diffusion of knowledge by acting as vehicles for knowledge spillovers across organizations and locations through person-to-person interaction. The role of individuals as active agents in the creation and spatial diffusion of knowledge is often emphasized in the literature (Almeida and Kogut, 1999; Howells, 2012), particularly because person-to-person contact involving a transfer or exchange of personnel is gathered as an efficient means of transmission across organizational boundaries for tacit knowledge (Kim et al., 2009). For example, Breschi and Lissoni (2009) argue that the most fundamental reason why geography matters in constraining the diffusion of knowledge is that mobile researchers are not likely to relocate in space, since this is responsible to a large extent for localization of co-inventions or citations (Miguèlez and Moreno, 2013a).

Only a handful of very recent contributions adopted a long-term dynamic perspective to investigate the evolution of knowledge networks (see Table 1). Our analysis spans two decades from 1998 to 2009. Moreover, our methodology provides the opportunity to evaluate the overall effectiveness of EU integration policies toward cross-border cooperation. By using a regression approach that is capable of determining evolution of the country-border effect for European vs. non European countries, our analysis provides an important and timely insight into the rate of European R&D integration relative to the rest of the world. Specifically we seek to test three basic research questions:

- (i) Is the effect of physical distance decreasing in magnitude?

⁵ Beyond mobile workers in the strict sense, *i.e.* workers switching employer or the establishment they work in, mobile inventors can be also consultants or academic scientists that offer their services to different companies (Breschi and Lissoni, 2009).

- (ii) Is the effect of country borders decreasing in magnitude?
- (iii) Is the country border effect decreasing in magnitude faster within Europe than among the rest of the developed world?

3. Data and methodology

3.1. Data

The data analyzed in this study are drawn from the OECD REGPAT database (Maraud et al., 2008; Webb et al., 2005) which compiles all patent applications filed with the European Patent Office (EPO) from the 1960s to present. In this database the geographical location of each inventor and applicant has been matched to the appropriate 5552 NUTS3 region in one of the 50 OECD or OECD-partner countries.⁶ This allows us to construct four geographical networks: (I) co-inventors, (II) applicant-inventor, (III) patent citations and (IV) inventor mobility. For each network we define $y_{m,n}$ as the number of links between NUTS3 region m and n . In (I) $y_{m,n}$ is equal to the number of patents jointly invented by the two regions. We use a full-counting approach so that a patent with $I(>1)$ inventors accounts for $\sum_{i=1}^{I-1} (I-i)$ regional links (hence, patents with only one inventor do not appear in this network by construction). Unlike (I), networks (II–IV) are directed networks in which we distinguish the pair (m, n) with respect to the pair (n, m) . In (II) the region of the applicant is linked to the regions of the affiliated inventors. The inventor's region usually indicates where the invention was made (often a laboratory or a research establishment, or the place of residence of the inventor) while the applicant's region indicates where the holder (usually a company, university or other type of entity) has its headquarters. In the database there is no direct information on affiliations, but it can be trivially retrieved for patents associated with a single applicant, the case for approximately 94% of the whole set of patents. In (III) for each pair (m, n) of NUTS3 regions we count the number of times that (a patent of an inventor in) region m cites (a patent of an inventor in) region n ($y_{m,n}$), and the number of citations that m receives from n ($y_{n,m}$). In (IV) a link indicates one inventor moving from one region to another one. Inventors regional migration can be tracked observing patent activity in at least two different years. In the case that an inventor has no patents for one or more years, we can track her region only at the beginning and at the end of the gap. In that case the flow is referred to the first year in which the inventor is observed again.⁷ Names of inventors have been cleaned and ambiguity over names have been dealt with, but have not been fully disambiguated. More precisely, we apply the same name parsing and matching procedure as in Miguèlez and Gomez-Miguèlez (2011). However, since our goal is to track inventors' mobility we cannot rely on the inventor's address field to fully disambiguate inventor's name. To circumvent this problem, we track the flow of names between regions which

is a gross proxy of pure inventor flows. Looking simply at the flows of names we may erroneously count as inventor move two authors sharing the same name and residing in two regions in two subsequent years. To minimize this source of error and get rather close to the true individual flows, we drop flows that most likely correspond to this case. Namely, we drop regional moves of names whenever the name is observed in the incoming region even in earlier years or in the source region after inventor's departure. This rule is not able to identify pure moves only in the unlikely situation in which two authors with the same name move simultaneously. Our goal is to count the number of moves between regions, not to track the careers of individual inventors over time. However, we only detect moves for inventors who have filed patents both in the source and destination regions. This results in observing a 13% of inventors who have been active on more than one NUTS3 region in the period 1981–2010.⁸

For the econometric analysis we create a balanced panel of data by networks for the period 1986–2009. In the estimation of the distance and border effects, the sample is restricted to all pairs of regions with at least twenty patents in every year.⁹ For the third exercise we retain in the sample all those pairs for which at least one link is registered in the time period.

3.2. Econometric methodology

The root of our econometric approach is the gravity model (Anderson and Van Wincoop, 2003; McCallum, 1995), a standard tool in the econometrics of trade which has been recently applied to the analysis of R&D network (see Table 1). In its most elementary form, the gravity approach models the intensity of the interaction between two nodes as a (negative) function of the geographical distance between the nodes and a (positive) function of their respective sizes.¹⁰ In applications to regional networks of knowledge it is typically assumed that knowledge transfers are hindered not only by physical distance, but also by other separation measures that can account for institutional, social, organizational and cognitive effects (Boschma, 2005).

Our econometric approach is based on three key steps that are relevant to investigate our research questions (i–iii). First, the probability density function of the dependent variable y_i , which is relevant to model the impact of each independent variable. Second, the selection of independent variables. Third, the econometric specification and measures we use to estimate the temporal evolution of the spatial effects. We describe here the main features of our approach, while we refer to the Appendix for a more detailed discussion on each of these steps.

The dependent variable is the number of links ($y_i \equiv y_{(m,n)}$) between NUTS3 regions (m and n) and we model its probability distribution with a count density. A number of models can be found

⁶ This is the full list of countries in the data set: Austria, Belgium, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Sweden, United Kingdom, Australia, Bulgaria, Brazil, Canada, Switzerland, Chile, China, Cyprus, Czech Republic, Estonia, Hong Kong, Republic of Croatia, Hungary, Israel, India, Iceland, Japan, South Korea, Liechtenstein, Lithuania, Latvia, Macedonia, Malta, Mexico, Norway, New Zealand, Poland, Romania, Russian Federation, Slovenia, Slovakia, Turkey, Taiwan, United States, and South Africa.

⁷ We stress that, while there might be some overlapping in the mobility and applicant-inventor networks, these capture very different R&D relations between regions. The applicant-inventor network captures the way in which applicant institutions organize the geographical structure of their laboratories. Any inventor move is associated to two applicant-inventor links, the one referring to the outgoing region and the other to the destination region. A data point for inventor moves can correspond to a data point in the applicant-inventor network only for the move destination region and in the particular case that the outgoing region is the same of the region of new applicant. This happens when an applicant relocate the inventor far from the applicant region.

⁸ This number is in line with the 9% that Breschi and Lissoni (2009) find analyzing moves of US inventors among Metropolitan Statistical Areas. They focus on the period 1978–2002 though and on a subset of technological fields (Organic Chemistry, Pharmaceuticals and Biotechnology). Our results are also in good agreement with recent findings by Miguèlez and Moreno (2013b). They report in Table 1 the average distance covered by inventors' movements during the 5-year time window 2001–2005 among EU27 (excluding Cyprus and Malta, but including Norway and Switzerland) NUTS2 regions. They obtain a 395.14 Km average distance, while we obtain 383.92 Km for the same sub-sample (of EU27 NUTS2 regions in 2001–2005).

⁹ We choose to apply a lower limit to the size of regions we consider in our analysis due to the large size of the networks. However, we experimented with other approaches, such as considering all regions at the NUTS2 level or selecting a lower cutoff for a limited number of countries. Results, which are robust to alternative sample selection strategies, are available upon request.

¹⁰ See Anderson and Van Wincoop (2004) and Bergstrand and Egger (2011) for two excellent surveys on gravity model applications. See Anderson (2011) for an updated review on theory.

in the literature to handle count densities, including the Poisson model, Negative Binomial model variants, and Zero-inflated models, as listed in Table 1. Since a large portion of NUTS3 region pairs have zero links, we opted for a Zero-Inflated Negative Binomial (ZINB) density, as consistent with Hoekman et al. (2009), Frenken et al. (2009) and Chessa et al. (2013). Zero-inflated models are suitable when data exhibits “excess zeros” as they take into account a large number of zero entries (Cameron and Trivedi, 1998). The expected number of links is thus a non-linear function of independent variables and is modeled as

$$E(y_i | X_i) = \left(1 - \frac{\exp(X_i \beta^0)}{1 + \exp(X_i \beta^0)}\right) * \exp(X_i \beta^1) = \frac{\exp(X_i \beta^1)}{1 + \exp(X_i \beta^0)}, \quad (1)$$

where, due to non-linearity, the impact of each regressor in X_i on the dependent variable is a function of the whole set of regressor in X_i , and due the Zero-Inflated supplement, the impact is mediated by two distinct parameters in the vectors β^0 and β^1 .

The selection of independent variables affects the specification of the two linear indices $X\beta^j$ ($j=0, 1$). First, we include the standard “gravity” variables *distance*, *size_m* and *size_n*. The continuous variable *distance* measures the distance, in kilometers, between the centroids of the NUTS3 regions. *size_m* and *size_n* denote the size of each of the two regions, which is represented by the total number of links attached to the region in a given year. Consistently with the literature, we add to the baseline gravity specification a further set of variables controlling for separation effects (see Appendix A for a more detailed discussion on the independent variables). Relevant to the research questions (ii) and (iii) is the addition of the variables *border* and *area*. The dummy variable *border* flags pairs of NUTS3 belonging to different countries (*border*=1 if *m* and *n* are in different countries, *border*=0 otherwise). The categorical variable *area_s* identifies three kinds of links ($S=3$) according to the geographical area: links within the EU area, links within the non-EU area and the flows between the two areas. European countries are assigned to the EU area in case they have been formally within the EU for most of our sample period. Thus we consistently use EU15 as definition of EU in the empirical analysis and countries that joined EU between 2004 and 2009 are assigned to the non-EU area.¹¹

Given our set of independent variables, the linear indices $X\beta^0$ for the zero-generating process and $X\beta^1$ for the Negative Binomial process are modeled in parallel as

$$X\beta^j = \beta_0^j + \beta_1^j \textit{border} + \beta_2^j \textit{distance} + \beta_3^j \textit{size}_{m,t} + \beta_4^j \textit{size}_{n,t} + \sum_{s=2}^S \gamma_s^j \textit{area}_s + \sum_{t=2}^T \theta_t^j \textit{year}_t + \sum_{k=1}^K \psi_k^j Z_k, \quad (2)$$

where $j=0, 1$, \textit{year}_t are dummy variables that capture exogenous yearly shifts, and Z is a set of variables containing either further controls or interaction terms relevant to identify the evolution of the spatial effect of interest.

We make use of the general model highlighted in Eqs. (1) and (2) to perform three sets of estimates according to our research questions (i–iii). The linear indices modeled in Eq. (2) are adjusted in each of the three exercises to allow for interactions between the spatial measure of our interest and year dummies. This allows the impact of spatial measures to vary year-by-year. In particular,

for cases (i) and (ii) concerning the strength of distance and country borders over time, we add a set of interaction terms between year dummies and, respectively, *distance* and *border*. For case (iii) concerning the evolution of the border effect within Europe, we separate the data into EU and non-EU sets, removing flows between the two. As a consequence, the trinomial variable *area_s* collapses into a binary variable contrasting links within the EU area to links within the non-EU area, that we call *eu*. Then linear indices are modeled following Chessa et al. (2013) who apply a Difference-in-Difference-in-Differences (DiDiD) strategy to isolate the country border effect within EU. Technically, this is achieved by including in Eq. (2) a full set of double/triple interaction terms among the three key variables *eu*, *border* and \textit{year}_t (see Appendix B for a more detailed description of the methodology). By interacting the yearly border effect with the *eu* indicator, we can purge the border effect within EU of global trends that are common to developed countries. In fact, global forces drive a general increase in the propensity to collaborate over national borders that gross up the impact of EU-specific integration policies. Since EU and non-EU developed countries are similarly exposed to global trends that affect the integration of their R&D systems, we use the non-EU OECD members as control group to identify the integration effect that can be attributed to EU specific factors. EU recent members that have not been either in the EU or non-EU area during most of our sample period, are removed from the group of non-EU OECD members in case (iii).¹²

The parameters in the adjusted linear indices of the three cases are estimated through maximum likelihood methods. Then to obtain meaningful measures of the evolving spatial effects, parameter estimates are replaced in Eq. (1) for each case to compute the expected number of links and its variations. In particular, we test our research questions estimating yearly marginal changes in the expected value of y_i . Marginal changes of $E(y_i | X)$ are essentially absolute or relative differences between Y' and Y'' that represent respectively the expected number of links in the base status (Y') and in the status reached as result of the relevant spatial effect (Y''). Given that $E(y_i | X)$ is a function of X_i , the computation of Y' and Y'' requires to set specific values for each regressors. We define the change in the expected number of links as result of a given spatial effect, in a given year, for a given network, as

$$D_t = (Y'' - Y'). \quad (3)$$

For cases (i) and (ii) we report the percentage change to allow comparisons among networks, i.e. we divide D_t by Y' (see Appendix B for details).

4. Results

In this section we present the results of our three regression models which estimate the evolution of (i) the distance effect in global patent activity, (ii) the border effect in global patent activity, and (iii) the cross-border effect in EU countries relative to non-EU countries. Finally, we summarize and discuss the main results of our work.

4.1. Evolution of the distance effect

Fig. 1 shows the evolution of the average distance of R&D collaborations between NUTS3 regions. The mobility network shows the largest change over the entire period, as the average distance for inventor relocation was 1267 kilometers in 1986 and has almost

¹¹ Our analysis of European integration thus refers to the extent of cross-border collaboration between EU15 countries. We also tried in a separate regression to control for the impact of New Member States, which turned out to be negligible.

¹² Using the EU15 definition, we remove European countries that joined EU between 2004 and 2009, namely Bulgaria, Czech Republic, Cyprus, Hungary, Latvia, Lithuania, Slovakia, Slovenia, Romania, Malta, Estonia, Poland.

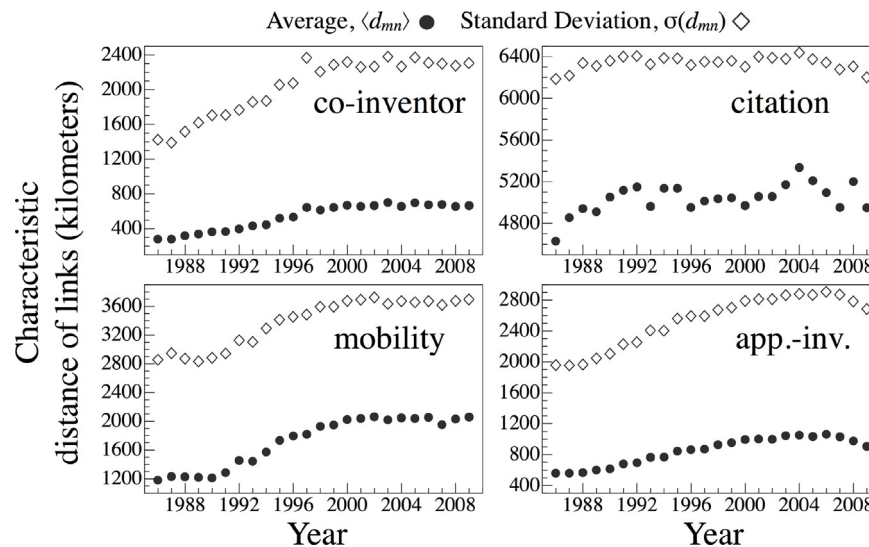


Fig. 1. Evolution of average distance of R&D collaborations. Notes: All links between NUTS3 regions are used to compute the average distance. For inventors mobility self-loops are removed as not meaningful.

doubled by 2009, increasing to 2051 kilometers. Note that we do not use inter-regional links ($d_{mn} = 0$) for the calculation of the average and standard deviation distance for the mobility network. As documented in previous literature for various collaboration measures, we also find that the average distance of R&D collaboration has increased for all four cases. However, we present evidence of saturation in recent years, mainly due to geographic upper-bounds on the largest collaboration distances.

Having demonstrated gradual shifts in the average of $d_{m,n}$ over time, we now shift our focus to the impact that distance has on the likelihood of collaboration. Fig. 2 shows network-by-network estimates of ϵ_t , the elasticity of $y_{m,n}$ with respect to $d_{m,n}$, as defined in Eq. (B.2). For each year we report the point elasticity evaluated at sample means and the 95% confidence interval. For example, for an “average” pair of NUTS3 region in 2008, a 1% increase in distance implies a 1.24% decrease in the expected number of links for the applicant-inventor network, a 1.13% decrease for the co-inventor network, a 0.94% decrease for the citation network, and a 0.27% decrease for the mobility network. As a computation example of these quantities, for the coinventor network in 2008, $\epsilon_{2008} = -1.13\%$ is obtained from $(Y'' - Y')/Y' = (0.11061 - 0.11187)/0.11187 = -0.0113$. These estimates demonstrate that distance is still a significant constraint on inter-regional connectivity for each network. The magnitude of the citation network elasticity is much less, indicating that distance impedes the flow of codified knowledge much less than the flow of tacit knowledge and people.

Regarding the time evolution of the distance effect in the co-inventor, applicant-inventor, and citations networks, we observe an overall positive trend corresponding to an increase over time in the magnitude of the distance effect (see Fig. 2). For the first two networks, the positive trend emerges around in mid 1990s, while for citations it starts earlier and is roughly stable from 2002 onwards. For mobility, there is a significant decrease from 1998–1997, then stabilizing afterward. Fig. 2 demonstrates overall that for three of the four networks, the magnitude of the distance effect is increasing over time, especially since mid 1990s, which is concurrent with the observed saturation in the average distance of links shown in Fig. 1.¹³

¹³ The increase in the magnitude of the distance effect is a general pattern among the developed countries we analyzed, and is not being driven by specific countries

Furthermore, Fig. 3 demonstrates that the positive trend in the magnitude of ϵ_t persists even when we remove regressors one by one until what remains is a basic gravity-like model, which includes only physical distance, its interactions with year dummies, year dummies and the size of nodes. However, it is crucial to note that if the size of nodes (regions) is omitted from the regression, resulting in a dependent variable that is regressed against only distance, its interaction with year dummies and year dummies, then the positive trend becomes a negative trend, as consistent with the observed positive trend in the average distance of links. The change in the trends suggests that the increase in the average distance of collaboration can be explained by the attraction premium of large regions (“hubs”) who gain an excess number of collaborations from afar. Following from general proportional growth models, as the central nodes continue to grow, the (new) peripheral nodes are more likely to connect to hubs, resulting in an increase of the average distance since there are a relatively small proportion of “hub” regions distributed throughout the globe. By accounting for this “attractive force” between two nodes by incorporating region size in the gravity approach, the net result is an overall increase in the magnitude of the distance effect with time.¹⁴

A closer look to the data reveals a significant shift in the increased likelihood of large distance $d_{ij} > 2000$ km collaborations over time. Hence, the increase in the average distance may be largely due to a “tail effect” in the extreme right tail of the distribution, as the distribution is relatively stationary in the range $50 < d_{m,n} < 1000$ km.

Summarizing this subsection, we find that, contrary to the popular belief, the role of distance is increasing in time with the only important exception of the inventor mobility network. Most of the changes in the distance effect occur before 2003, whereas there is no significant variation in the last period.

or groups of similar countries. We estimated several models using different subsamples and the results demonstrate that our general conclusions are in fact robust with respect to diverse subsets of countries. These results are made available by the authors.

¹⁴ Continuing to reintroduce the other controls back into our regression model, the overall trend in the magnitude of the distance effect remains the same, there is an overall increase in the magnitude of the distance effect. Fig. 3 refers to the applicant-inventor network. A similar plot for co-inventorships is available upon request.

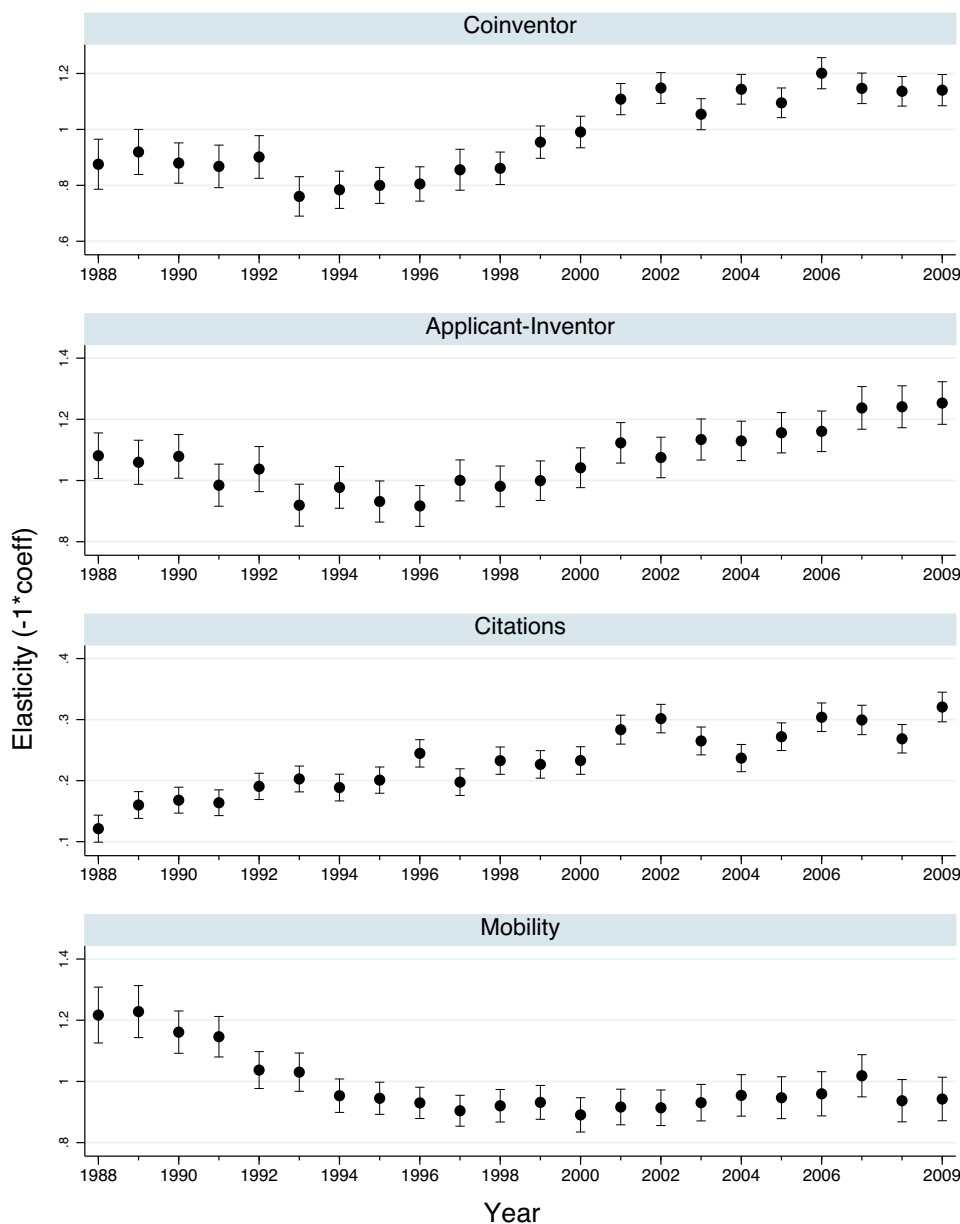


Fig. 2. Elasticity of distance over years. Notes: Estimates derive from four separate ZINB-gravity models for the count of links between NUTS3 regions. In the graph we report the year-elasticity and the corresponding 95% confidence interval constructed estimating standard errors through the Delta method. Year estimates of the distance effect are obtained including interaction terms between the continuous distance variable and year dummies. We report in Table C.3 in Appendix C regression estimates of parameters in Eq. (B.1) from which elasticities are calculated. Given these parameters estimates and sample averages of regressors, we calculate the expected value of y as in Eq. (1) for two different levels of distance, i.e. sample average (Y) and sample average increased by 1% (Y'). As a computation example, for the coinventor network in 2008, this results in $\epsilon_{2008} = (Y' - Y)/Y = (0.11061 - 0.11187)/0.11187 = -0.0113 = -1.13\%$.

4.2. Evolution of the border effect

In Fig. 4 we show the evolution of the border effect. These plots report the yearly semi-elasticity with respect to country-borders, i.e. the percentage change in the expected link count $y_{(m,n)}$ when the dummy *border* changes from 0 to 1. For example in 2008, taking an “average” pair of NUTS3 regions, the country border effect reduces the expected $y_{(m,n)}$ by 88.7% for the co-inventor network, 88.4% for the applicant-inventor network, 85.2% for the mobility network and 48.1% for the citation network.¹⁵ Thus, the effect of country

¹⁵ As a computation example of these quantities, for the coinventor network in 2008, $\pi_{2008} = -88.7\%$ is obtained from $(Y' - Y)/Y = (0.0688 - 0.6085)/0.6085 = -0.8869$.

borders is clearly quite strong. Similar to the distance effect, it is far less important for the citation network and more important in the co-inventor and applicant-inventor networks than for the mobility network.

Unlike the distance effect, we find some indication that the border effect is decreasing in magnitude for the all four networks. For the co-inventor, applicant-inventor and citation networks, the trend in the semi-elasticity is overall negative when considering the entire period of analysis. However, the trend is not continuous as there appear to be sub periods with positive trend. Notably, we observe an increase in the magnitude of the border effect starting around in the late 1990s or early 2000s for the co-inventor, applicant-inventor, and citation networks, and persisting until recent years. For the citation network the overall negative trend turns out to be not statistically significant. Nevertheless, the levels

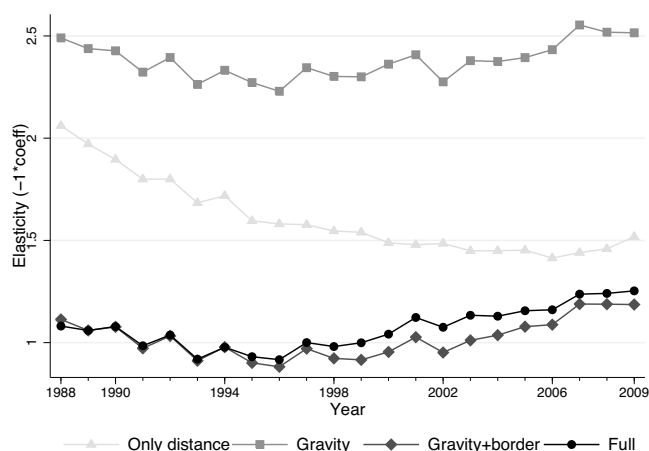


Fig. 3. Elasticity of distance with different models – applicant-inventor. Notes: Each line plot refer to estimates of different ZINB model (see notes to Fig. 2). We report below the set of regressors used in each case: “Full”: See Eq. (B.1); “Gravity+border”: $X^{\beta j} = \beta_0^j + \beta_1^j \text{border} + \beta_2^j \text{distance} + \beta_6^j \text{size}_{m,t} + \beta_7^j \text{size}_{n,t} + \sum_{t=2}^T \theta_t^j \text{year}_t + \sum_{t=2}^T \delta_t^j \text{distance} * \text{year}_t$; “Gravity”: $X^{\beta j} = \beta_0^j + \beta_2^j \text{distance} + \beta_6^j \text{size}_{m,t} + \beta_7^j \text{size}_{n,t} + \sum_{t=2}^T \theta_t^j \text{year}_t + \sum_{t=2}^T \delta_t^j \text{distance} * \text{year}_t$; “Only distance”: $X^{\beta j} = \beta_0^j + \beta_2^j \text{distance} + \sum_{t=2}^T \theta_t^j \text{year}_t + \sum_{t=2}^T \delta_t^j \text{distance} * \text{year}_t$.

still remain lower than 1988, the first year of analysis. For mobility network, the border effect resembles the pattern observed for the distance effect, there is a significant decline until 1995 followed by a relatively stable period until 2009.

To provide a better understanding of the kind of collaborations which happen more often across countries, we split regional links according to the size of regions. We identify the 100 top regions as those having the highest number of patents filed in the period 1986–2009 and distinguish three groups of links respectively between (1) large and large regions, (2) large and small regions, (3) small and small regions.¹⁶ In Fig. 5 we report the temporal evolution of the percentage of co-inventor links, subdivided among these three groups, that are also cross-border. An overall increase in the total share of cross-border links is apparent, with roughly half of the total owed to the cross-border activity between Top100 and NonTop100. Thus, the cross-border share is drastically larger for links between small and large regions, indicating that when small regions collaborate across borders they are significantly more likely to collaborate with large regions. The relative increase in the percentage of links that are cross-border is very similar across the three subsets, corresponding to approximately a 100% growth over the period analyzed. However, it is also clear that despite the same relative growth, the share of cross-border links that pair top regions with non-top regions is increasing over the other types in absolute terms. This suggests that, in the light of our regression results, an easing of the cross-border effect can be partially explained by the increased likelihood of small regions pairing with large regions.¹⁷ Nevertheless, the increase in cross-border links between top regions is also substantial, which reiterates the role of large hubs as global conduits of knowledge transfer.

¹⁶ The percentages of all links pertaining to the three groups are 44.1% (Top100–Top100), 18.2% (Top100–NonTop100), 37.7% (NonTop100–NonTop100) respectively.

¹⁷ A similar reaching out effect has been already noticed for the US Life Sciences patent network by Owen-Smith et al. (2002).

4.3. Evolution of the European research integration

Efforts to stimulate and reinforce R&D integration in the EU have been ongoing for decades, with the aim of developing an innovation system that can benefit from cross-border knowledge spillovers. Cross-border R&D activities are not only important to the EU, but reflect overall globalization trends. As the prevalence of multi-institutional and multi-national teams increases across science (Wutchy et al., 2007; Jones et al., 2008), the evaluation of the propensities and incentives to collaborate internationally will become an important and active area of research (European Commission, 2013, 2013; Science Europe, 2013).

In this section we measure the role of borders in EU vis-à-vis non-EU collaboration networks by comparing two types of links: (a) links that are completely contained in the EU, (b) links that are completely external to the EU. Flows between EU and non-EU are removed for identification purpose. This analysis points out the impact of EU-specific factors (a systemic “treatment”) aimed at increasing cross-border connectivity, by measuring changes in the link counts relative to non-EU countries which did not receive the treatment.

Fig. 6 shows estimates of the absolute change in the number of cross-border links due to EU-specific integration forces operating in year t relative to a baseline year ($t^* = 2004$), indicating that there are some positive signs of integration in European patent activity. In the case of the co-inventor network, we find an increasing overall trend of cross-border collaboration between inventors in Europe vis-à-vis other OECD countries. This effect was remarkable since the mid-1990s, but has stalled since 2004.¹⁸ However, analogous trends in the other networks are difficult to detect.¹⁹ Apart from some positive effects for the applicant-inventor in 1986–1997, and citation network in 1996–2000 and 2005–2009, no significant trends can be depicted for the entire sample period.²⁰ We replicated the analysis including also new member states in the group of European NUTS3 regions. In the augmented group of 27 countries the role of new members in the R&D networks is anyways very small, accounting for a tiny percentage of the whole links. This is reflected in estimates as the evolution of the treatment effects is very similar to what we report.²¹

4.4. Conclusions

The main results of our econometric analysis are summarized in Table 2, which illustrates for each network trends in the dynamics of the three effects. The overall trends we find are consistent for all networks, except inventor mobility, and in line with previous results in the literature (see Table 1). However, by taking a closer look into the dynamics of innovator networks across geographical and institutional borders, we notice that there have been some relevant changes of the global tendency over time. First, results in Fig. 1 show that the increase of the average distance saturated across all networks. We also notice that the effect of geographical and institu-

¹⁸ As a computation example, for the coinventor network in 2008, this results in $\tau_{2008} = 0.2988 - 0.3210 = -0.0222$, i.e. no difference in the impact of EU-specific factors upon integration with respect to 2004, the baseline year. See Eq. (B.6).

¹⁹ In the case of inventor mobility, the number of non-zero link counts was too low to be modeled using ZINB, thus estimation was carried out aggregating the network at NUTS2 level.

²⁰ These trends are very similar to the findings of Chessa et al. (2013), who use the same methodological approach but follow a different sample selection rule and use the co-applicant network in place of the applicant-inventor network. In particular, they select an unbalanced panel of NUTS3 regional pairs removing all regions below a cutoff on the number of patents. We interpret the finding that our conclusions are qualitatively very similar as a sign of robustness of the methodological approach.

²¹ Figures are made available by the authors.

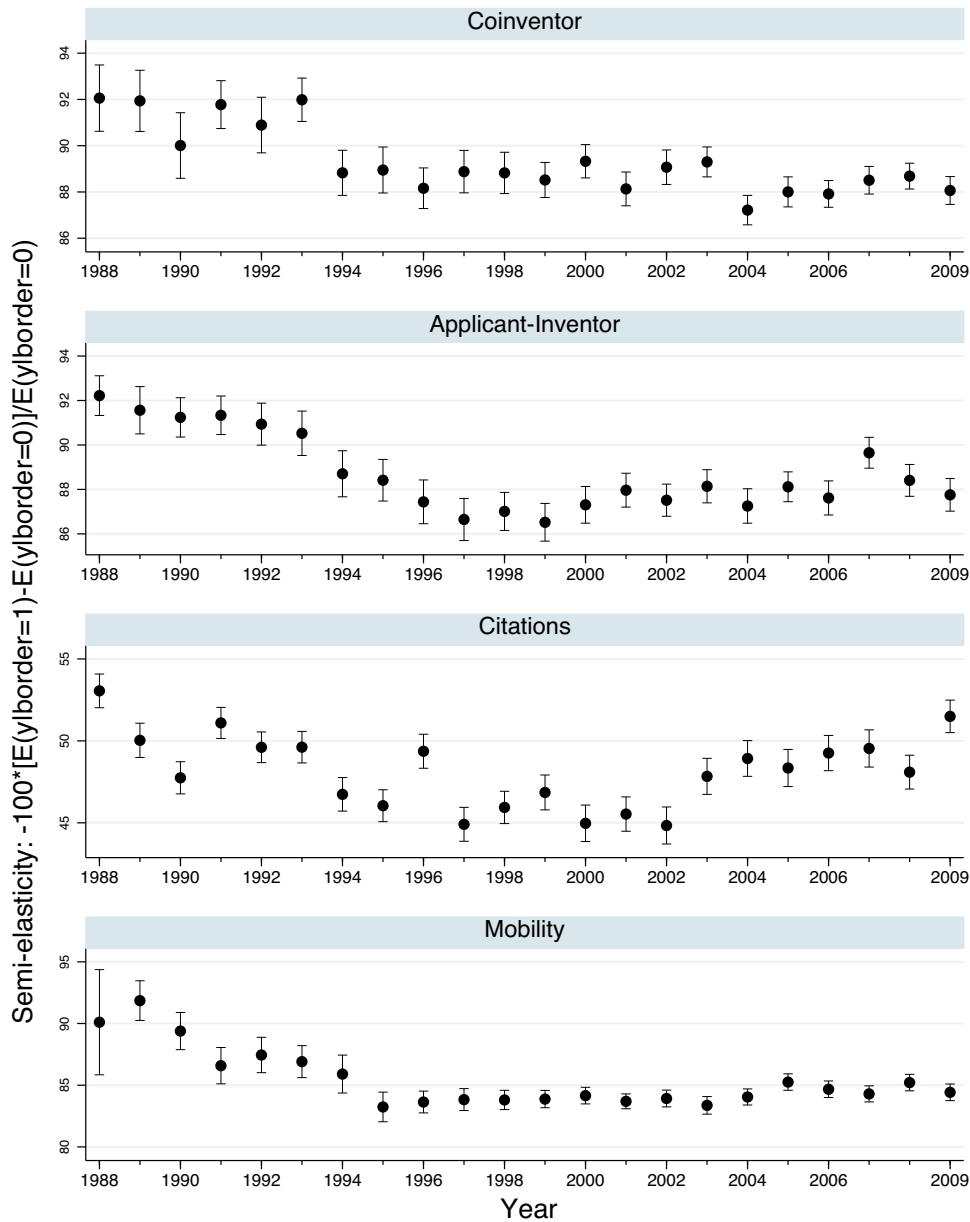


Fig. 4. Border effect over years. Notes: Estimates derive from four separate ZINB-gravity models for the count of links between NUTS3 regions. In the graph we report the yearly semi-elasticity and the corresponding 95% confidence interval constructed estimating standard errors through the Delta method. Year estimates of the country-border effect are obtained including interaction terms between the cross-border dummy and year dummies. We report in Table C.4 in Appendix C regression estimates of parameters in Eq. (B.3) from which semi-elasticities are calculated. Given these parameters estimates and sample averages of regressors, we calculate the expected value of y as in Eq. (1) for $border=0$ (Y') and $border=1$ (Y''). As a computation example, for the coinventor network in 2008, this results in $\pi_{2008} = (Y'' - Y')/Y' = (0.0688 - 0.6085)/0.6085 = -0.8869 = -88.7\%$.

tional barriers has been increasing, or at best has been stable, since mid-1990s, as shown in Figs. 2 and 4. The outcome of our analysis reveals that the slowdown of the globalization trend compounds with the emergence of innovation hubs, which attract connections on a global scale. Fig. 5 shows that hubs attract a large amount of connections from small and mid sized regions across borders. These connections are steadily growing faster than within borders connections, contributing to an easing of the cross-border effect. However, this trend in the connections between large and smaller regions has also contributed to the reverse in the role of distance since mid-Nineties, as most of the attracted nodes are in the close by countries. Overall, Table 2 shows that the increase in geographical and institutional barriers is stronger and starts before for codified knowledge (patent citations) while it is weaker and more recent

for tacit knowledge (inventor mobility) with other R&D networks as intermediate cases.

In particular, we notice that the inventor mobility network appears to be becoming less bound by borders and distance, while all others are becoming more constrained. For citations, it is likely that a saturation limit has been reached, whereas the span of collaboration networks is about a half of the average distance traveled by mobile inventors. Evidence of easing in institutional barriers to international mobility in 1988–2005 can be explained by globalization of labor markets rather than by EU policies. All in all, the sudden stop of the process of European integration since 2004 seems to be caused by an inversion of the global trends. This has some important implications for the Horizon 2020 agenda that we will discuss in the concluding policy-oriented discussion.

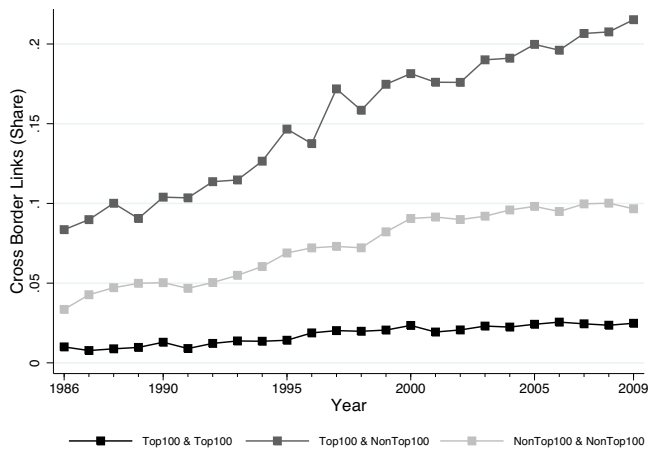


Fig. 5. Flows between top regions and small regions – evolution of the cross-border shares. Notes: These plots report the year-by-year % of cross-country regional links for three sub-samples. Regional links are grouped in links of (1) both top 100 regions, (2) one top 100 and one non-top region, (3) both non-top regions. Top 100 regions are selected counting the total number of patents over the period 1986–2009.

5. Final discussion

We analyzed the temporal evolution of spatial biases in the strength of inter-regional connectivity within regional patent networks. Focusing on a set of 50 developed nations over the period 1988–2009, using inter-regional links at the NUTS3 region level, we have contributed to the body of literature on the geography of knowledge by analyzing four different R&D networks: co-inventor, applicant-inventor, citations, inventor mobility. Making use of a gravity-like econometric approach and controlling for a number of separation effects, we estimated year-by-year effects of physical distance and country-borders, and the trend of integration in the European Union as compared to the other developed countries.

Contrary to the widespread notion that the importance of distance has been decreasing over time due to globalization and technological advancement, our results show that the constraint imposed by geographical distance on R&D inter-regional links seem to have actually increased in three of the networks analyzed: co-inventor, applicant-inventor, and citations (see Table 2). On average, inter-regional links take place at a larger distance, although this trend has saturated in the last decade for each of the networks we analyzed. The greatest shift in the frequencies of observed inter-regional links has been in the extreme tail of the distance distribution. We observe roughly a factor of 10 increase in the likelihood of collaborations for distances above 2000 km when comparing 1986–2008. Although these long-distance collaborations are far less likely than the average, a factor of 10 increase can induce a “tail effect” which can account for the increase in the average distance of collaborations reported in the literature and confirmed here.

The overall increase in the average distance can intuitively be understood as the result of large “hub” regions attracting new peripheral regions as they enter the system and grow over time. However, *ceteris paribus*, for a pair of regions of a given size the strength of their connectivity gets more sensitive to physical distance with time. This means the cost of inter-regional collaboration at a given distance is still large, even increasing, but whenever a small region becomes connected to a hub the relevance of this cost is counterbalanced by the benefit of linking to a core region. Indeed, large and diversified regions tend to extend their basin of attraction across national borders, prevalently toward small regions. As the EU seeks to implement a “Teaming Scheme” (aka “Research Buddy” plan) whereby top research centers will pair with

“low-performing” member states to set up research centers, there should be a better understanding of the role of size on collaboration pairing (see Institutional Consortium (2013)). A possible explanation for the preferential attachment of big regions has to do with their capability to combine local proximity to attract socially distant groups with the capability of attracting socially close but geographically distant communities of inventors (Agrawal et al., 2008). It is worthwhile to further explore the sources of the preferential attachment mechanism across multiple R&D networks in future research.

The higher likelihood of small regions pairing with large regions across national borders contributes to an easing of the cross-border effect. Our estimates of the evolution of the cross-border effect indicate that national borders can be crossed more easily now than in the late 1980s, particularly due to a significant decrease up to the late 1990s or early 2000s depending on the network (see Table 2). Hence, when we observe collaborations to happen more frequently across borders, this is largely driven both by an erosion of institutional frictions that impede inter-national connectivity (Hoekman et al., 2010) and the “reaching-out” by international hubs, rather than a decrease in the costs associated with collaborating over distance. Part of this story can be explained in terms of the role played by the European Union in promoting inter-national connectivity within the area, though signs of integration are weak and overall significant only for collaboration between inventors (see Table 2). Moreover, while the contribution of EU-specific integration forces upon the reduction of the border effect on a global scale can be identified for some time frames, global trends often evolve independently, emphasizing the role played by the erosion of formal and informal institutional barriers not EU-specific. An outstanding example is the mobility network, which seems to have key relevance in driving cross-network interdependencies. In fact, we note that the time window where we observe a significant decrease in the border effect for mobility is somehow related to the period of decreasing border effect in the other R&D networks. This applies also for the distance effect, at least for co-inventor and the location of R&D activities. These results reinforce the view that individual mobility is the driving force of knowledge diffusion (Breschi and Lissoni, 2009). However, given no evidence of impact on international mobility arising from EU institutions, this potential seems largely unexploited in the EU area (European Commission, 2013).

In estimating the evolution of the distance effect we note that the mobility network stands out as the only network with a negative trend, though with no significant change after 1997. This suggests that the globalization of skilled-labor job markets which enabled a reduction in mobility costs have had a larger impact on the geography of knowledge than advances that favor a reduction in cost of communication. In particular, the result that the distance effect is steadily increasing in the network of citations despite well-known advances in technologies easing the codification of knowledge corroborates the notion that tacit and embodied knowledge still play a major role in diffusion. In particular, patents are pieces of codified knowledge building upon a stock of tacit knowledge that hinders its fruition (Breschi and Lissoni, 2009). Overall, the increase in distance effect supports the view that improvements in communication technologies, while on the one hand facilitating the substitution of face-to-face interactions with arm’s-length communication, on the other hand create a greater need for close interactions to exchange complex knowledge which is responsible for research activities to agglomerate rather than to disperse (Gaspar and Glaeser, 1998).

For the geographical dispersion of the network of R&D activities (applicant-inventor network) we observe a significant decrease of the border effect over the period 1988–1999 and a significant decline of the distance effect over a similar time window (1988–1996). Excessive geographical dispersion of learning centers

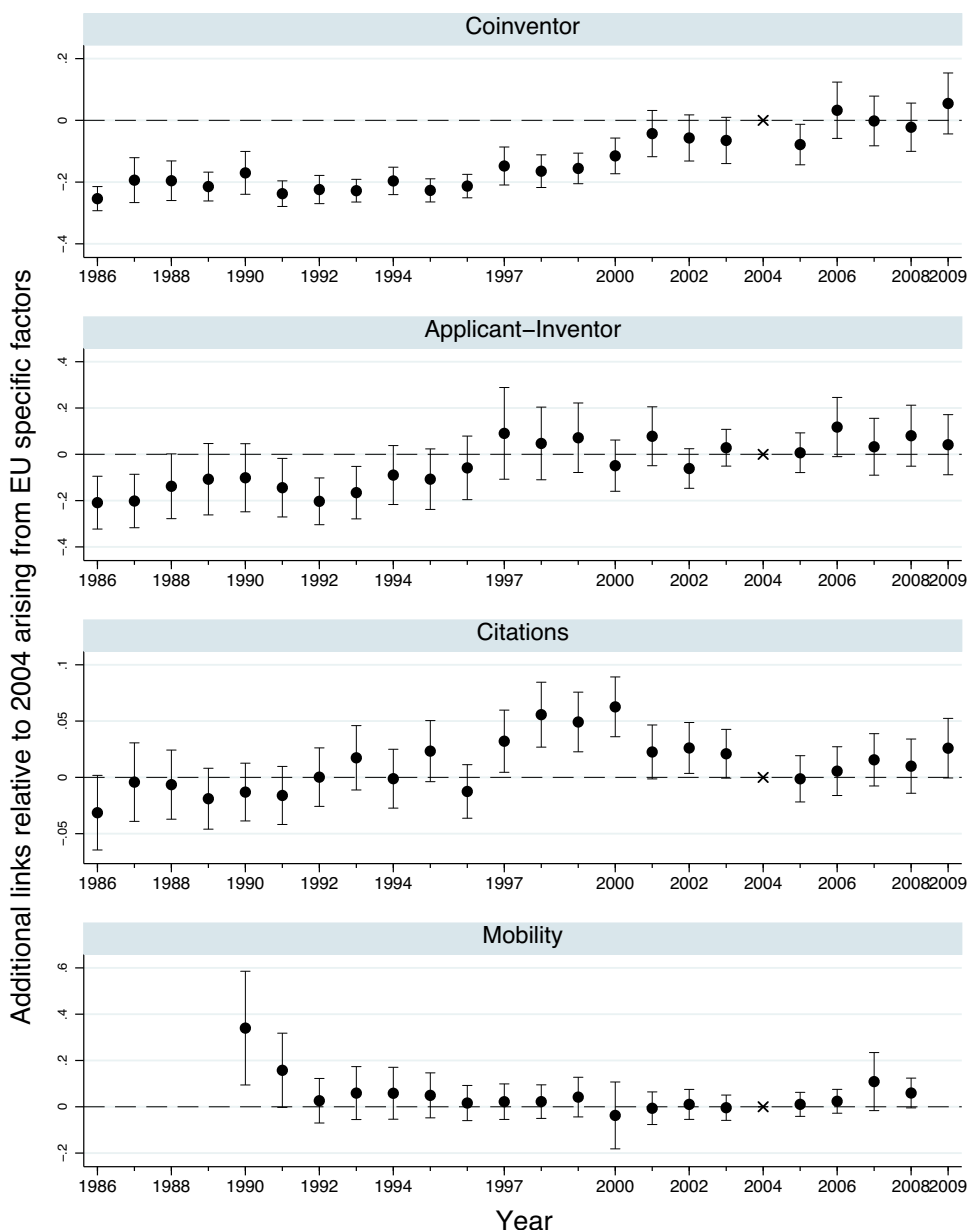


Fig. 6. Evolution of European integration. Notes: Estimates derive from four separate ZINB-DiDiD-gravity models for the count of links between NUTS3 regions. In the graph we report point estimates of the average treatment effects relative to 2004 and 95% confidence intervals. The y axis reports the additional number of cross-border links for an average pair of regions (i) relative to within-border links, (ii) due to EU-specific factors as compared with non-EU OECD countries, and (iii) relative to 2004 baseline year. We report in Table C.5 in Appendix C regression estimates of parameters in Eq. (B.5) from which treatment effects are calculated. Given these parameters estimates and specific values selected for regressors, we calculate the expected value of y as in Eq. (1) for $year = t' = 2004$ (Y') and $year = t$ (Y''). Values of regressors are computed for an average pair of NUTS3 regions belonging to two different EU countries in 2004. As a computation example, for the coinventor network in 2008, this results in $\tau_{2008} = (Y'' - Y') = (0.2988 - 0.3210) = -.0222$. In the case of inventor mobility the number of non-zero link counts was too low to be modeled using ZINB, thus estimation is carried out aggregating the network at NUTS2 level.

can lead to difficulties in controlling the generation and exploitation of knowledge, especially given its predominant content of tacitness. This argument has been invoked to explain a substantial change in international location decisions observed immediately after an opposite trend between 1985 and 1995. In fact, the strong movement to establish a transnational configuration of R&D observed between 1985 and 1995 has been blamed to result in overly complex and unmanageable organizational architectures (Gerybadze and Reger, 1999). In the light of the role played by knowledge-seeking reasons in the internalization of innovative activities, our results are coherent with these trends and point out that limits encountered by R&D internalization strategies in

controlling the accumulation of knowledge across geographical and institutional borders have not been reduced by globalization forces.

As concluding remarks concerning policy, we stress the importance of R&D clusters. Our evidence suggest that integration in research is being driven by the top regions reaching out to more peripheral regions and across borders. This trend in the evolution of R&D networks supports policies oriented to the exploitation of agglomeration economies in research clusters rather than targeting promotion of cross-border collaboration (Hoekman et al., 2010). However, the importance of investment in programs that incentivize mobility of researchers throughout Europe seems to be reaffirmed, even if we do not have explicit evidence of tangible

Table 2
Results summary of the three research questions.

	(i) Distance effect			(ii) Border effect			(iii) Integration EU vs. non-EU		
	Subperiod	Trend	Overall trend	Subperiod	Trend	Overall trend	Subperiod	Trend	Overall trend
(I) Coinventor	1988–1993	↘		1988–1993	=		1986–1995	=	
	1993–2002	↗↗	↗↗	1993–2004	↘↘	↘↘	1995–2004	↗↗	↗↗
	2002–2009	=		2004–2009	↗		2004–2009	=	
(II) Applicant-inventor	1988–1996	↘↘		1988–1993	↘		1986–1997	↗	
	1996–2009	↗↗	↗↗	1993–1999	↘↘	↘↘	1998–2009	=	↗
(III) Citations	1988–2002	↗↗		1999–2009	↗		1986–1996	=	
	2002–2009	=	↗↗	1988–2002	↘↘		1996–2000	↗↗	↗
				2002–2009	↗↗	↘	2000–2005	↘	
(IV) Mobility	1988–1997	↘↘		1988–1995	↘↘		2005–2009	↗	
	1997–2009	=		1995–2009	=		1990–1992	↘	
							1992–2008	=	=

Note: ↗↗ significantly increasing, ↗ mildly increasing, = stable, ↘ mildly decreasing, ↘↘ significantly decreasing. This table summarizes the main trends identified for the three sets of estimates reported in Figs. 2, 4 and 6.

benefits in the European Union as opposed to the rest of the developed world.

It is also important to stress that policies embedded within EU funding programs aimed at transnational cooperation (see European Commission (2013)) may run counter to global trends. By way of example, while the overall intensity of cross-border activity is increasing in Europe, it may be evolving in a way that is orthogonal to the US/World. Specifically, for the cases in which large regions collaborate with small and mid-sized regions over medium distances, we find that the EU has significantly larger fraction of EU–EU links than the US has US–US links (roughly twice as many). Conversely, for the long distance group, the EU–World link counts are significantly outnumbered by the USA–World link counts for pairings involving large regions. Hence, the policy implication is that while encouraging intra-EU collaboration is good for developing and sustaining the ERA, it may come at the cost of missing out on competitive global market forces which match collaboration partners according to “best-with-best” principles, independent of region (Boyle, 2013). To this end, EU policy makers might consider the pros and cons of a competitive funding system wherein the most competitive EU grants do not have any EU-collaboration criteria.

As a final remark, we point out some limitations in our analysis, which could be addressed in future research. We do not consider scientific publications or R&D projects and collaborative agreements in our analysis. Further investigation is needed to assess whether similar trends are present for basic research and other networks of innovators. Another extension maybe to explicitly test for the dynamic interplay between different R&D networks focusing in particular on the relationship between inventors' mobility and collaboration. Finally, the increasing availability of large data sets of bibliometric information should encourage the application of new quantitative methods to assess the efficacy of the European R&D policies for smart specialization and integration.

Acknowledgements

We gratefully acknowledge feedbacks from Michael Hopkins, Daniele Rotolo, Ed Steinmueller and other seminar participants at the SPRU Friday seminar, 19 April 2013, University of Sussex. We also acknowledge Stefano Breschi and other seminar participants at the ENIC Workshop, July 2013, IWH Halle. Authors acknowledge funding from the National Research Program of Italy (PNR) project “CRISIS Lab”, the Italian Ministry of Education, University, and Research (MIUR) [PRIN project 2009Z3E2BF], the project SWITCH

and the Research Foundation of Flanders (FWO) (G073013N). O.P. acknowledges funding from the Social Sciences and Humanities Research Council of Canada.

Appendix A. Independent variables

We discuss the set of variables used for the general econometric model described by Eqs. (1) and (2). We focus here on the variables we add to the baseline gravity specification comprising the physical distance and the size of the two nodes. First, we add three spatial measures that account for different dimensions of distance: *border*, *neighbour* (Peri, 2005; Maggioni and Uberti, 2007; Scherngell and Barber, 2009, 2011; Paci and Usai, 2009; Scherngell and Hu, 2011; Scherngell and Lata, 2012; Miguèlez and Moreno, 2013a), and *area* (Peri, 2005; Frenken et al., 2009; Chessa et al., 2013). The dummy variable *border* flags pairs of NUTS3 belonging to different countries (*border*=1 if *m* and *n* are in different countries, *border*=0 otherwise). This variable is almost always present in gravity equations and takes into account that a common institutional framework eases coordination and interactions among individuals and groups (Gertler, 1995; Edquist and Johnson, 1997; Boschma, 2005). Indeed, the relevant institutional elements, either formal such as intellectual property rights, funding schemes and labor markets, or informal such as the cultural background, social norms and the language, have a strong national component (Hoekman et al., 2009). For this reason, the knowledge transfer is expected to be larger in regions that belong to the same country. However, the borders of common institutional frameworks do not necessarily correspond to national borders as other settings can overlap both at higher and lower divisions. At a lower level, geographical contiguity of regions surrounds cultural, social and language similarities that can facilitate knowledge transfers independently of national borders and physical proximity. For example, contiguous regions that share a national border may be as proximate in terms of informal institutions as contiguous regions that do not share a national border. Also, taken two pairs of regions with the same physical distance, say (*m*, *n*) and (*m*, *k*), region *m* could be more inclined to interact with *k* if *m* and *k* are contiguous while *m* and *n* are not. We thus include the dummy variable *neighbour* that flags pairs of adjacent NUTS3 (*neighbour*=1 if adjacent, *neighbour*=0 otherwise). To control for institutional proximity at a higher macro-political division, we split the networks in three kinds of links (*S*=3) according to the geographical area: links within the EU area, links within the non-EU area and the flows between the two areas. These macro-area effects are captured by the categorical variable *area*.

European countries are assigned to the EU area in case they have been formally within the EU for most of our sample period. Thus we consistently use EU15 as definition of EU in the empirical analysis and countries that joined EU between 2004 and 2009 are assigned to the non-EU area.²²

To identify the impact on knowledge flows that can be genuinely attributed to the spatial measures of our interest, it is important to control for non-spatial measures that can be related to spatial measures and affect the diffusion of knowledge as well. We therefore include as further separation variable the continuous variable *techdist* that measures the technological distance between regions in a given year (Peri, 2005; Maggioni and Uberti, 2007; Paci and Usai, 2009; Hoekman et al., 2010; Scherngell and Barber, 2009, 2011; Scherngell and Hu, 2011; Scherngell and Lata, 2012; Miguèlez and Moreno, 2013a). This variables proxy for cognitive proximity between regions since reflects the extent to which they share a common, related, or complementary technological knowledge base. Regions closely located may have comparable technological background thus the effect of physical distance may be overestimated if we omit to control for this (Hoekman et al., 2010). *techdist* is constructed using patent classes according to the International Patent Classification (IPC). In particular, for each region *m* we compute the vector $t(m)$ that measures the share of patenting in each of the technological subclasses for a given year. Technological subclasses correspond to the third-digit level of the IPC systems. We define the technological distance between regions *m* and *n* as $techdist_{m,n} = 1 - r^2$ where $r^2 = corr[t(m), t(n)]^2$ is the Pearson correlation coefficient between the technological vectors $t(m)$ and $t(n)$ (see Moreno et al., 2005 and Scherngell and Barber, 2009). The possibility that *distance* and *techdist* are complements or substitutes is accounted for by including also the interaction term $distance * techdist$ (Agrawal et al., 2008).²³

Appendix B. Strategy to test the three research questions

We describe here in more details the specification of linear indices (see Eq. (2)) and the measures we use to estimate the temporal evolution of the spatial effects. The linear indices $X\beta^0$ for the zero-generating process and $X\beta^1$ for the Negative Binomial process are modeled in parallel as

$$\begin{aligned} X\beta_i^j &= \beta_0^j + \beta_1^j border + \beta_2^j distance + \beta_3^j techdist_t \\ &+ \beta_4^j distance * techdist_t + \beta_5^j neighbour + \beta_6^j size_{m,t} \\ &+ \beta_7^j size_{n,t} + \sum_{s=2}^S \gamma_s^j area_s + \sum_{t=2}^T \theta_t^j year_t \\ &+ \sum_{t=2}^T \delta_t^j distance * year_t, \end{aligned} \quad (B.1)$$

where $j = 0, 1$. The linear indices modeled in Eq. (B.1) are adjusted in each of the three exercises to allow for interactions between the relevant spatial measure and year dummies. Maximum likelihood

²² Our analysis of European integration thus refers to the extent of cross-border collaboration between EU15 countries. We also tried in a separate regression to control for the impact of New Member States, which turned out to be negligible.

²³ We thank an anonymous referee for pointing out the role of interactions between proximity measures. We estimated the elasticity for *distance* over different levels of *techdist*. Our results are in line with (Agrawal et al., 2008) as show for the citation network that physical proximity and technological proximity are substitutes, i.e. the marginal benefit of physical distance is larger the lower the technological proximity. We do not report here this analysis to avoid overloading of the analytical content, but figures can be provided upon request.

estimates of parameters in the linear indices are then used in Eq. (1) to compute yearly marginal changes in $E(y_i | X)$.

The sample used in estimation is always a balanced panel of regional pairs. In case of unbalanced panels, some regions may appear or disappear over the sample period causing attrition bias. For cases (i) and (ii) we run estimates on a balanced panel of data by networks for the period 1988–2009. The sample used for estimation is constructed from regions with at least 20 patents in every year. We chose to set a threshold on patents for two reasons. First, there is a large concentration of NUTS3 regions pairs with no links, since many regions have very few patents. Second, our measure of technological distance requires a reasonable number of patents to be statistically reliable.²⁴ Given this rule, the estimation sample is identical in case (i) and (ii) and is constant across networks.

To test our research question (i) concerning the strength of distance over time, we make use of maximum likelihood estimates of parameters in Eq. (B.1) and compute the elasticity of y_i with respect to *distance* over years.²⁵ Specifically we estimate for each year the quantity

$$\begin{aligned} \epsilon_t &= \frac{(Y'' - Y')}{Y'}, \text{ where} \\ Y' &= E(y_i | X = (X), year_t = 1, distance = (distance)), \end{aligned} \quad (B.2)$$

$$Y'' = E(y_i | X = (X), year_t = 1, distance = (distance) * 1.01).$$

Estimates of Y' and Y'' are obtained replacing in Eq. (1) parameter estimates from Eq. (B.1) and setting regressors values at sample means (i.e. $X = (X)$) except for $year_t$ and *distance*.²⁶

To test our second question (ii) we estimate the evolution of the country-border effect. To do this we modify Eq. (B.1), adding interactions of *border* with year dummies, resulting in

$$X\beta_{ii}^j = X\beta_i^j + \omega_t^j border * year_t. \quad (B.3)$$

Given maximum likelihood estimates of parameters in the augmented Eq. (B.3) we compute the marginal effects of the *border* variable over years.²⁷ We report percentage changes as in case (i) to allow comparisons among networks. In particular we compute for each year the semi-elasticity defined as

$$\begin{aligned} \pi_t &= \frac{(Y'' - Y')}{Y'}, \text{ where} \\ Y' &= E(y_i | X = (X), year_t = 1, border = 0), \end{aligned} \quad (B.4)$$

$$Y'' = E(y_i | X = (X), year_t = 1, border = 1).$$

Estimates of Y' and Y'' are obtained replacing in Eq. (1) parameter estimates from Eq. (B.3) and setting regressors values at sample means (i.e. $X = (X)$) except for $year_t$ and *border*.

To test our third question (iii) we change the specification of Eq. (B.1) and separate the data into EU and non-EU sets, removing flows between the two. Linear indices are modeled following Chessa et al.

²⁴ Robustness checks were performed using different thresholds, both lower and higher than 20. Results hold very similar to those reported in this article. These are made available by the authors.

²⁵ The elasticity is formally defined as $(\partial E(y_i) / \partial distance)(distance / E(y_i))$. From Eq. (1) we can compute the derivative as $(\partial E(y_i) / \partial distance) = E(y_i) (\beta_2^j - \beta_2^0 (\exp(X_i \beta^0) / (1 + \exp(X_i \beta^0))))$. Thus we obtain $\epsilon_t = distance (\beta_2^j - \beta_2^0 (\exp(X_i \beta^0) / (1 + \exp(X_i \beta^0))))$.

²⁶ See Winkelmann (2008) for the computation of marginal effects for the ZINB model.

²⁷ As concerns case (i), the temporal evolution of the distance effect is very similar when using the augmented Eq. (B.3), i.e. when the temporal evolution of the cross-border effect is accounted for in the model. Results are available upon request.

Table C.3
Distance effect – regression estimates.

$y \equiv N_{(m,n)}$	Coinventor		Applicant-inventor		Citations		Mobility	
	$y (\geq 0)$	ZI	$y (\geq 0)$	ZI	$y (\geq 0)$	ZI	$y (\geq 0)$	ZI
Area (base = non – EU ↔ non – EU)								
area ₂ (EU ↔ EU)	–0.0501 (0.0350)	0.4398** (0.0313)	–0.0218 (0.0516)	0.2447** (0.0353)	0.3926** (0.0191)	0.6922** (0.0162)	0.4028** (0.0345)	1.1543** (0.0408)
area ₃ (EU ↔ non – EU)	0.2130* (0.0394)	–0.2868** (0.0314)	0.3384* (0.0717)	–0.3628* (0.0547)	0.0905* (0.0164)	–0.0638* (0.0149)	0.7929* (0.0947)	–0.3476* (0.1249)
border	–0.9220** (0.0319)	1.9549** (0.0273)	–1.1117** (0.0627)	1.9128** (0.0429)	–0.3436** (0.0153)	0.7242** (0.0132)	–1.3374** (0.1376)	1.6427** (0.1764)
year dummies		✓		✓		✓		✓
year_distance		✓		✓		✓		✓
distance	–0.0001** (0.0000)	0.0001** (0.0000)	–0.0001** (0.0000)	0.0000** (0.0000)	0.0000 (0.0000)	0.0000** (0.0000)	–0.0002** (0.0000)	–0.0000* (0.0000)
techdist	–2.0163** (0.0536)	0.8756** (0.0417)	–1.7550** (0.0767)	1.0735** (0.0539)	–1.6014** (0.0292)	0.4708** (0.0270)	–0.9833** (0.0606)	0.0458 (0.0700)
distance_techdist	0.0001** (0.0000)	0.0001** (0.0000)	0.0000* (0.0000)	0.0000 (0.0000)	0.0000** (0.0000)	0.0000** (0.0000)	–0.0001** (0.0000)	–0.0002** (0.0000)
neighbour	2.4128** (0.0387)	–26.8920** (0.5206)	2.0788** (0.0524)	–26.9908** (0.1068)	1.6756** (0.0421)	–4.0495** (0.2340)	1.0503** (0.0426)	–2.4044** (0.0814)
size _m	0.0007** (0.0000)	–0.0019** (0.0001)	0.0002* (0.0000)	–0.0058** (0.0001)	0.0001** (0.0000)	–0.0023** (0.0000)	0.0028** (0.0001)	–0.0401** (0.0008)
size _n	0.0001** (0.0000)	–0.0003** (0.0000)	0.0009** (0.0000)	–0.0001** (0.0000)	0.0001** (0.0000)	–0.0023** (0.0000)	0.0022** (0.0000)	–0.0438** (0.0009)
Obs.		3,213,540		6,415,200		6,415,200		6,403,320
No. of region pairs		146,070		291,600		291,600		291,060
No. of regions		540		540		540		540
ln(α)		1.9004**		1.8684**		1.3393**		0.1199*
Vuong test ~ N(0, 1)		125.53**		157.40**		322.19**		132.19**
l-Ratio test ~ χ ²		6,758,968.3**		6,024,960.4**		8,624,418.7**		197,858.4**

Note: Robust standard errors in brackets. The table reports estimates of four separate ZINB-gravity models for the count of links between NUTS3 regions (from these estimates we compute yearly elasticities of the distance effect reported in Fig. 2). For each model we report estimates of parameters in the two ZINB parts vectors, namely β¹ (pure count part vector) and β⁰ (zero-inflation part vector) as defined in Eq. (B.1). Parameters in β¹ capture the regressor effect on the number of links, provided the zero-generating process did not produce a zero. Parameters in β⁰ capture the regressor effect on the probability of observing a zero. Cross-sections of region pairs are pooled over years and estimation is carried out on the whole sample clustering standard errors at region pairs. size_m refers to the smaller of the two regions for co-inventor, while refers, respectively, to citing, to applicant's and to exit region for citations, applicant-inventor and inventor mobility. Vuong test statistics support the choice of the ZINB over a pure version NB2 (ψ_i = 0, ∀ i) (Vuong, 1989; Long and Freese, 2006) and likelihood ratio tests support the choice of ZINB vs. the ZIP (Long and Freese, 2006).

* Significant at 5%.
** Significant at 1%.

(2013) who apply a Difference-in-Difference (DiD) strategy to isolate the country border effect within EU.²⁸ Non-EU OECD members are used as control group to identify the integration effect that can be attributed to EU specific factors. In the impact evaluation jargon, this group acts as a counterfactual, in the sense that it is used to proxy EU trends that would have emerged in the absence of efforts devoted to boost integration in EU. The sample used in the estimation is a balanced panel of data for the period 1986–2009. The sample is restricted to those pairs for which at least one link is registered in the time period.²⁹ The linear indices in B.1 are now modeled as

$$\begin{aligned}
 X\beta^j &= \beta_0^j + \beta_1^j border + \beta_2^j eu + \beta_3^j distance + \beta_4^j size_m + \beta_5^j size_n \\
 &+ \gamma^j border * eu + \sum_{t=2}^T \theta_t^j year_t + \sum_{t=2}^T \delta_t^j border * year_t \\
 &+ \sum_{t=2}^T \xi_t^j eu * year_t + \sum_{t=2}^T \eta_t^j border * eu * year_t, \quad (B.5)
 \end{aligned}$$

where the trinomial variable *area* collapses in the binomial variable *eu* as links between the EU area and non-EU area are removed for identification purpose.

The identification of yearly treatment effects is achieved by including a full set of double/triple interaction dummy variables among the three dimensions ($eu = \{0, 1\}$, $border = \{0, 1\}$, $year_t = \{0, 1\}$ for $t = 2, \dots, T$) (Wooldridge, 2010).

Denoting the actual and counterfactual outcomes of our dependent variable as Y'' and Y' respectively and taking into account our DiDiD extension, we define the yearly treatment effect as

$$\begin{aligned}
 \tau_t &= (Y'' - Y'), \text{ where} \\
 Y' &= E(y_i | X = (X_{border,eu,t^*}), border = 1, eu = 1, year_{t^*} = 1), \\
 Y'' &= E(y_i | X = (X_{border,eu,t^*}), border = 1, eu = 1, year_t = 1). \quad (B.6)
 \end{aligned}$$

Estimates of Y' and Y'' are obtained replacing in Eq. (1) parameter estimates from Eq. (B.5) and setting specific values for regressors. We refer all regressors values to a generic pair of cross-border EU regions in the baseline year (i.e. we average over the sub-sample of these pairs obtaining $X = (X_{border,eu,t^*})$), except for *border*, *eu*, and $year_t$. Relative to the baseline year t^* (we use the arbitrarily chosen year 2004), τ_t reflects the impact of changes in institutional factors specific to the EU which have taken place in a given year t with respect to t^* .

Appendix C. Tables of estimates

See Tables C.3–C.5.

²⁸ See Chessa et al. (2013) for a more detailed description of the methodology.
²⁹ Here we do not include in the regressions the variable *techdist* which require to set a threshold on the number of patents. For the comparison in integration trends between the two areas (EU and non-EU) is important to take into account in estimation possible differences in the integration behavior of small regions.

Table C.4
Border effect – regression estimates.

$y_i \equiv N_{(m,n)}$	Coinventor		Applicant-inventor		Citations		Mobility	
	$y(\geq 0)$	ZI	$y(\geq 0)$	ZI	$y(\geq 0)$	ZI	$y(\geq 0)$	ZI
<i>eu</i>	0.0150 (0.0311)	-0.1587** (0.0240)	-0.0988* (0.0472)	-0.0680* (0.0335)	-0.2473** (0.0161)	-0.3090** (0.0139)	-0.4945** (0.0369)	-0.8811** (0.0510)
<i>border</i>	-0.7671** (0.1070)	2.1157** (0.0738)	-1.6142** (0.0935)	1.3643** (0.0816)	-0.5355** (0.0319)	0.4916** (0.0348)	-2.2973** (0.3151)	0.0361 (0.3746)
<i>year dummies</i>		✓		✓		✓		✓
<i>year_distance</i>		✓		✓		✓		✓
<i>year_border dummies</i>		✓		✓		✓		✓
<i>distance</i>	-0.0001** (0.0000)	0.0000** (0.0000)	-0.0001** (0.0000)	0.0000** (0.0000)	0.0000 (0.0000)	0.0000** (0.0000)	-0.0001** (0.0000)	0.0000 (0.0000)
<i>techdist</i>	-1.8516** (0.0510)	1.1228** (0.0355)	-1.7566** (0.0779)	1.0096** (0.0542)	-1.6005** (0.0288)	0.4055** (0.0278)	-0.9651** (0.0614)	-0.0247 (0.0715)
<i>distance_techdist</i>	0.0001** (0.0000)	0.0001** (0.0000)	0.0000** (0.0000)	0.0000** (0.0000)	0.0000** (0.0000)	0.0000** (0.0000)	-0.0001** (0.0000)	-0.0002** (0.0000)
<i>neighbour</i>	2.3341** (0.0356)	-26.4159** (0.1233)	2.1003** (0.0529)	-30.4591** (0.0951)	1.6666** (0.0424)	-4.1917** (0.2656)	1.1036** (0.0434)	-2.4532** (0.0774)
<i>size_m</i>	0.0007** (0.0000)	-0.0020** (0.0000)	0.0002** (0.0000)	-0.0057** (0.0001)	0.0001** (0.0000)	-0.0022** (0.0000)	0.0027** (0.0001)	-0.0394** (0.0008)
<i>size_n</i>	0.0001** (0.0000)	-0.0003** (0.0000)	0.0008** (0.0000)	-0.0001** (0.0000)	0.0001** (0.0000)	-0.0023** (0.0000)	0.0022** (0.0000)	-0.0432** (0.0009)
Obs.		3,213,540		6,415,200		6,415,200		6,403,320
No. of region pairs		146,070		291,600		291,600		291,060
No. of regions		540		540		540		540
$\ln(\alpha)$		0.950**		1.872**		1.353**		0.160**
Vuong test $\sim N(0, 1)$		119.41**		156.67**		320.34**		130.87**
l -Ratio test $\sim \chi^2$		6,725,599.4**		6,022,369**		8,702,921**		105,927.8**

Note: Robust standard errors in brackets. The table reports estimates of four separate ZINB-gravity models for the count of links between NUTS3 regions (from these estimates we compute yearly semi-elasticities of the border effect reported in Fig. 4). For each model we report estimates of parameters in the two ZINB parts vectors, namely β^1 (pure count part vector) and β^0 (zero-inflation part vector) as defined in Eq. (B.3). Parameters in β^1 capture the regressor effect on the number of links, provided the zero-generating process did not produce a zero. Parameters in β^0 capture the regressor effect on the probability of observing a zero. Cross-sections of region pairs are pooled over years and estimation is carried out on the whole sample clustering standard errors at region pairs. *size_m* refers to the smaller of the two regions for co-inventor, while refers, respectively, to citing, to applicant's and to exit region for citations, applicant-inventor and inventor mobility. Vuong test statistics support the choice of the ZINB over a pure version NB2 ($\psi_i = 0, \forall i$) (Vuong, 1989; Long and Freese, 2006) and likelihood ratio tests support the choice of ZINB vs. the ZIP (Long and Freese, 2006).

* Significant at 5%.

** Significant at 1%.

Table C.5
Border effect EU vs. non-EU – regression estimates.

$y_i \equiv N_{(m,n)}$	Coinventor		Applicant-inventor		Citations		Mobility	
	$y(\geq 0)$	ZI	$y(\geq 0)$	ZI	$y(\geq 0)$	ZI	$y(\geq 0)$	ZI
<i>border</i>	0.7176** (0.1355)	-0.0607 (0.2968)	1.7120** (0.1623)	1.4961** (0.1618)	0.0901 (0.0500)	-0.2019** (0.0741)	-0.9399** (0.1133)	-20.3427** (1.8511)
<i>eu</i>	0.1006** (0.0313)	-0.0291 (0.0764)	0.2268** (0.0442)	-0.1014 (0.0763)	-0.3748** (0.0564)	-0.5394** (0.0585)	0.1091 (0.0732)	-0.2752 (0.3600)
<i>border_eu</i>	-2.3355** (0.1361)	-0.5995 (0.4006)	-3.0723** (0.1657)	-0.8334** (0.1982)	-0.9464** (0.0636)	0.4161** (0.0904)	-1.1694** (0.1447)	19.4609** (2.8939)
<i>year dummies</i>		✓		✓		✓		✓
<i>border_year dummies</i>		✓		✓		✓		✓
<i>eu_year dummies</i>		✓		✓		✓		✓
<i>border_eu_year dummies</i>		✓		✓		✓		✓
<i>distance</i>	-0.0003** (0.0000)	0.0001** (0.0000)	-0.0003** (0.0000)	0.0001** (0.0000)	-0.0001** (0.0000)	0.0001** (0.0000)	-0.0001** (0.0000)	0.0004** (0.0001)
<i>size_m</i>	0.0029** (0.0000)	-0.0363** (0.0020)	0.0003** (0.0000)	-0.0100** (0.0006)	0.0002** (0.0000)	-0.0038** (0.0001)	0.0015** (0.0001)	-0.0283** (0.0051)
<i>size_n</i>	0.0001** (0.0000)	-0.0001** (0.0000)	0.0020** (0.0001)	0.0002** (0.0000)	0.0001** (0.0000)	-0.0018** (0.0000)	0.0020** (0.0002)	-0.0220** (0.0047)
Obs.		3,847,704		3,151,440		7,266,600		127,015
No. of region pairs		160,321		131,310		302,775		6,685
No. of regions		4,039		3,152		3,003		287
$\ln(\alpha)$		2.1240**		2.2804**		2.2226**		0.3010**
Vuong test $\sim N(0, 1)$		152.34**		172.13**		106.8**		33.9**
l -Ratio test $\sim \chi^2$		21,128,342**		12,890,196**		12,890,196**		62,773.76**

Note: Robust standard errors in brackets. The table reports estimates of four separate ZINB-DiDiD-gravity models for the count of links between NUTS3 regions (from these estimates we compute yearly treatment effects reported in Fig. 6). For each model we report estimates of parameters in the two ZINB parts vectors, namely β^1 (pure count part vector) and β^0 (zero-inflation part vector) as defined in Eq. (B.5). Parameters in β^1 capture the regressor effect on the number of links, provided the zero-generating process did not produce a zero. Parameters in β^0 capture the regressor effect on the probability of observing a zero. Cross-sections of region pairs are pooled over years and estimation is carried out on the whole sample clustering standard errors at region pairs. *size_m* refers to the smaller of the two regions for co-inventor, while refers, respectively, to citing, to applicant's and to exit region for citations, applicant-inventor and inventor mobility. In the case of inventor mobility the number of non-zero link counts was too low to be modeled using ZINB, thus estimation is carried out aggregating the network at NUTS2 level. Vuong test statistics support the choice of the ZINB over a pure version NB2 ($\psi_i = 0, \forall i$) (Vuong, 1989; Long and Freese, 2006) and likelihood ratio tests support the choice of ZINB vs. the ZIP (Long and Freese, 2006).

* Significant at 5%.

** Significant at 1%.

References

- Agrawal, A., Cockburn, I., McHale, J., 2006. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* 6, 571–591.
- Agrawal, A., Kapur, D., McHale, J., 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics* 64, 258–269.
- Aldieri, L., 2011. Technological and geographical proximity effects on knowledge spillovers: evidence from the US patent citations. *Economics of Innovation and New Technology* 20, 597–607.
- Almeida, P., Kogut, B., 1999. Localisation of knowledge and the mobility of engineers in regional networks. *Management Science* 45, 905–917.
- Anderson, J.E., 2011. The gravity model. *Annual Reviews of Economics* 3, 133–160.
- Anderson, J.E., Van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. *American Economic Review* 93, 170–192.
- Anderson, J.E., Van Wincoop, E., 2004. Trade costs. *Journal of Economic Literature* 42, 691–751.
- Belderbos, R., Leten, B., Suzuki, S., 2013. How global is R&D? Firm-level determinants of home-country bias in R&D. *Journal of International Business Studies* 44, 765–786.
- Bergstrand, G.H., Egger, P., 2011. Gravity equations and economic frictions in the world economy. In: Bernhofen, D., Falvey, R., Greenaway, D., Kriekemeier, U. (Eds.), *Palgrave Handbook of International Trade*. Palgrave Macmillan Press, New York.
- Boerner, Penumarthy, S., Meiss, M., Ke, W., 2006. Mapping the diffusion of scholarly knowledge among major US research institutions. *Scientometrics* 68, 415–426.
- Boschma, R., 2005. Proximity and innovation: a critical assessment. *Regional Studies* 39, 61–74.
- Boyle, P., 2013. A single market for European research. *Nature* 501, 157–158.
- Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9, 439–468.
- Breschi, S., Malerba, C., 2005. *Clusters, Networks, and Innovation*. Oxford University Press, Oxford.
- Broekel, T., Balland, P., Burger, M., Van Oort, F., 2013. Modeling Knowledge Networks in Economic Geography: A Discussion of Four Empirical Strategies. *Papers in Evolutionary Economic Geography (PEEG)* 1325. Utrecht University, Section of Economic Geography.
- Brun, J., Carrère, C., Guillaumont, P., De Melo, J., 2005. Has distance died? Evidence from a panel gravity model. *World Bank Economic Review* 19, 99–120.
- Cairncross, F., 1997. *The Death of Distance: How the Communications Revolution will Change Our Lives*. Harvard Business School Publishing, Cambridge, MA.
- Cameron, A.C., Trivedi, P.K., 1998. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.
- Cantwell, J., 1989. *Technological Innovation and Multinational Corporations*. Basil Blackwell, Oxford.
- Cassi, L., Plunket, A., 2012. Co-patenting and inventive performance: in search of the proximity paradox. In: *Paper for the DRUID Conference*, p. 2012.
- Castells, M., 1996. *The Rise of the Network Society*. Blackwell Publishers, Oxford.
- Chessa, A., Morescalchi, A., Pammolli, F., Penner, O., Petersen, A.M., Riccaboni, M., 2013. Is Europe evolving toward an integrated research area? *Science* 39, 650–651.
- Coe, D., 2002. *The Missing Globalization Puzzle*. IMF Working Paper.
- Dunning, J.H., 1992. *Multinational Enterprises and The Global Economy*. Addison-Wesley, Wokingham, England.
- Dunning, J.H., Lundan, S.M., 2009. The internationalisation of corporate R&D: a review of the evidence and some policy implications for home countries. *Review of Policy Research* 26, 13–33.
- Edquist, C., Johnson, B., 1997. Institutions and organisations in systems of innovations. In: Edquist, C. (Ed.), *Systems of Innovation. Technologies Institutions and Organizations*. Pinter, pp. 41–63.
- European Commission, 2013. *European Research Area Facts and Figures 2013*. Technical Report. European Commission.
- European Commission, 2013. *European Research Area Progress Report 2013*. Technical Report. European Commission.
- Feldman, M.P., 1994. *The Geography of Innovation*, vol. 2. Kluwer Academic Publisher.
- Fischer, M.M., Scherngell, T., Jansenberger, E., 2006. The geography of knowledge spillovers between high-technology firms in Europe: evidence from a spatial interaction modeling perspective. *Geographical Analysis* 38, 288–309.
- Florida, R., 1997. The globalization of R&D: results of a survey of foreign-affiliated R&D laboratories in the USA. *Research Policy* 26, 85–103.
- Freeman, C., 1991. Networks of innovators: a synthesis of research issues. *Research Policy* 20, 499–514.
- Freeman, C., 1995. The “national system of innovation” in historical perspective. *Cambridge Journal of Economics* 19, 5–24.
- Frenken, K., 2002. A new indicator of European integration and an application to collaboration in scientific research. *Economic Systems Research* 14, 345–361.
- Frenken, K., Hardeman, S., Hoekman, J., 2009a. Spatial scientometrics: towards a cumulative research program. *Journal of Informetrics* 3, 222–232.
- Frenken, K., Hoekman, J., Kok, S., Ponds, R., Van Oort, F., Van Vliet, J., 2009. Death of distance in science? A gravity approach to research collaboration. In: Pyka, A., Scharnhorst, A. (Eds.), *Innovation Networks*. Springer-Verlag, New York, pp. 43–57.
- Gaspar, J., Glaeser, E.L., 1998. Information technology and the future of cities. *Journal of Urban Economics* 43, 136–156.
- Gertler, M., 1995. “Being there”: proximity, organization, and culture in the development and adoption of advanced manufacturing technologies. *Economic Geography* 71, 1–26.
- Gerybadze, A., Reger, G., 1999. Globalization of R&D: recent changes in the management of innovation in transnational corporations. *Research Policy* 28, 251–274.
- Giroud, X., 2013. Proximity and investment: evidence from plant-level data. *The Quarterly Journal of Economics* 128, 861–915.
- Gomes-Casseres, B., Jaffe, A., Hagedoorn, J., 2006. Do alliances promote knowledge flows? *Journal of Financial Economics* 80, 5–33.
- Granstrand, O., 1999. Internationalisation of corporate R&D: a study of Japanese and Swedish corporations. *Research Policy* 28, 275–302.
- Harhoff, D., Thoma, G., 2010. *Inventor Location and The Globalization of R&D*.
- Hoekman, J., Frenken, K., Tijssen, R., 2010. Research collaboration at a distance: changing spatial patterns of scientific collaboration in Europe. *Research Policy*, 39.
- Hoekman, J., Frenken, K., Van Oort, F., 2009. The geography of collaborative knowledge production in Europe. *The Annals of Regional Science* 43, 721–738.
- Hoekman, J., Scherngell, T., Frenken, K., Tijssen, R., 2013. Acquisition of European research funds and its effect on international scientific collaboration. *Journal of Economic Geography* 13, 23–52.
- Howells, J., 2012. The geography of knowledge: never so close but never so far apart. *Journal of Economic Geography* 12, 1003–1020.
- Institutional Consortium, 2013. *Teaming for excellence: building high quality research across Europe through partnership*. Max-Planck-Gesellschaft White Paper, 1–13.
- Jaffe, A., Trajtenberg, M., 2002. *Patents, Citations and Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge, MA.
- Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108, 578–598.
- Jones, B.F., Wutchy, S., Uzzi, B., 2008. Multi-university research teams: shifting impact, geography, and stratification in science. *Science* 322, 1259–1262.
- Keller, W., 2004. International technology diffusion. *Journal of Economic Literature* 42, 752–782.
- Kim, J., Lee, S.J., Marschke, G., 2009. International knowledge flows: evidence from an inventor-firm matched data set. In: Freeman, R.B., Goroff, D.L. (Eds.), *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*. University of Chicago Press, pp. 321–348, chapter 10.
- Kogut, B., Zander, U., 1993. Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies* 24, 625–645.
- LeSage, J.P., Fischer, M.M., Scherngell, T., 2007. Knowledge spillovers across Europe: evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science* 86, 393–421.
- Long, S., Freese, J., 2006. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd ed. Stata Press, College Station, Texas, USA.
- Lundvall, B., 1992. *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter, London.
- Maggioni, M., Uberti, T., 2007. Inter-regional knowledge flows in Europe: an econometric analysis. In: Frenken, K. (Ed.), *Applied Evolutionary Economics and Economic Geography*. Edward Elgar, Cheltenham, pp. 230–255.
- Maraud, S., Dernis, H., Webb, C., Spiezia, V., Guellec, D., 2008. *The OECD REGPAT Database: A Presentation*. Working Paper 2. OECD.
- Maurseth, P.B., Verspagen, B., 2002. Knowledge spillovers in Europe: a patent citations analysis. *The Scandinavian Journal of Economics* 104, 531–545.
- McCallum, J., 1995. National borders matter: Canada–U.S. regional trade patterns. *American Economic Review* 85, 615–623.
- Miguélez, E., Gomez-Miguélez, I., 2011. Singling Out Individual Inventors From Patent Data. IREA Working Paper No. 201105. University of Barcelona, Research Institute of Applied Economics.
- Miguélez, E., Moreno, R., 2013a. Do labour mobility and technological collaborations foster geographical knowledge diffusion? The case of European regions. *Growth and Change* 44, 321–354.
- Miguélez, E., Moreno, R., 2013b. Mobility, Networks and Innovation: The Role of Regions’ Absorptive Capacity. IREA Working Paper 2013/16. University of Barcelona, Research Institute of Applied Economics.
- Montobbio, F., Sterzi, V., 2013. The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations. *World Development* 44, 281–299.
- Moreno, R., Paci, R., Usai, S., 2005. Spatial spillovers and innovation activity in European regions. *Environment and Planning A*, 1793–1812.
- Narula, R., Zanfei, A., 2005. Globalization of innovation: the role of multinational enterprises. In: Fagerberg, J., Nelson, R.R., Mowery, D.C. (Eds.), *Handbook of Innovation*. Oxford University Press, Oxford, chapter 12, pp. 99–99.
- Nelson, R., 1993. *National Innovation Systems: A Comparative Analysis*. Oxford University Press, Oxford.
- Owen-Smith, J., Riccaboni, M., Pammolli, F., Powell, W.W., 2002. A comparison of U.S. and European university–industry relations in the life sciences. *Management Science*, 48.
- Ozman, M., 2009. Inter-firm networks and innovation: a survey of literature. *Economics of Innovation and New Technology* 18, 39–67.
- Paci, R., Usai, S., 2009. Knowledge flows across European regions. *The Annals of Regional Science* 43, 669–690.

- Pan, R.K., Kaski, K., Fortunato, S., 2012. World citation and collaboration networks: uncovering the role of geography in science. *Scientific Reports* 2, 1–7.
- Patel, P., Vega, M., 1999. Patterns of internationalisation of corporate technology: location vs. home country advantages. *Research Policy* 28, 145–155.
- Peri, G., 2005. Determinants of knowledge flows and their effect on innovation. *The Review of Economics and Statistics* 87, 308–322.
- Phene, A., Almeida, P., 2003. How do firms evolve? The patterns of technological evolution of semiconductor subsidiaries. *International Business Review* 12, 349–367.
- Picci, L., 2010. The internationalization of inventive activity: a gravity model using patent data. *Research Policy* 39, 1070–1081.
- Ponds, R., 2009. The limits to internationalisation of scientific research collaboration. *Journal of Technology Transfer* 34, 76–94.
- Ponds, R., Van Oort, F., Frenken, K., 2007. The geographical and institutional proximity of research collaboration. *Papers in Regional Science* 86, 423–443.
- Powell, W.W., Grodal, S., 2005. Networks of innovators. In: Fagerberg, J., Mowery, D.C., Nelson, R.R. (Eds.), *The Oxford Handbook of Innovation*. Oxford University Press, Oxford, pp. 56–85 (Chapter 3).
- Scherngell, T., Barber, M.J., 2009. Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. *Papers in Regional Science* 88, 531–645.
- Scherngell, T., Barber, M.J., 2011. Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU framework programme. *The Annals of Regional Science* 46, 247–266.
- Scherngell, T., Hu, Y., 2011. Collaborative knowledge production in China. Regional evidence from a gravity model approach. *Regional Studies* 45, 755–772.
- Scherngell, T., Lata, R., 2012. Towards an integrated European research area? Findings from eigenvector spatially filtered spatial interaction models using European framework programme data. *Papers in Regional Sciences* 92, 555–577.
- Science Europe, 2013. Comparative Benchmarking of European and US Research Collaborations and Researcher Mobility. Technical Report. Science Europe and Elsevier's Scival Analytics.
- Singh, J., 2007. Asymmetry of knowledge spillovers between MNCs and host country firms. *Journal of International Business Studies* 38, 764–786.
- Singh, J., Marx, M., 2013. Geographic constraints on knowledge spillovers: political borders vs. spatial proximity. *Management Science* 59, 2056–2078.
- Sonn, J., 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, 1020–1039.
- von Zedtwitz, M., Gassmann, O., 2002. Market versus technology drive in R&D internationalisation: four patterns of managing research and development. *Research Policy* 31, 569–588.
- Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57.
- Webb, C., Dernis, H., Harhoff, D., Hoisl, K., 2005. Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks. Working Paper 9. OECD.
- Whittington, K.B., Owen-Smith, J., Powell, W.V., 2009. Networks, propinquity, and innovation in knowledge-intensive industries. *Administrative Science Quarterly* 54, 90–122.
- Winkelmann, R., 2008. *Econometric Analysis of Count Data*. Springer-Verlag, Berlin.
- Wolfmayr, Y., Christen, E., Falk, M., Hollenstein, H., Knell, M., Pfaffermayr, M., Reinthaller, A., Unterlass, F., 2013. The Role and Internationalisation Strategies of Multinational Companies in Innovation. INNO-Grips II Report. European Commission, DG Enterprise and Industry, Brussels.
- Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge.
- Wutchy, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316, 1036–1038.