

RA **Economics and institutional change**

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

Andrea Morescalchi
Fabio Pammolli
Orion Penner
Alexander M. Petersen
Massimo Riccaboni

ISSN 2279-6894
IMT LUCCA EIC WORKING PAPER SERIES #01/2014
© IMT Institute for Advanced Studies Lucca
Piazza San Ponziano 6, 55100 Lucca

Research Area
Economics and institutional change

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

Andrea Morescalchi

IMT Institute for Advanced Studies Lucca;
European Commission Joint Research Centre, Ispra, Italy

Fabio Pammolli

IMT Institute for Advanced Studies Lucca

Orion Penner

IMT Institute for Advanced Studies Lucca

Alexander M. Petersen

IMT Institute for Advanced Studies Lucca

Massimo Riccaboni

IMT Institute for Advanced Studies Lucca;
Department of Managerial Economics, Strategy and Innovation, K.U. Leuven

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

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

^a*IMT Lucca, Lucca, Italy*

^b*DMSI, KU Leuven, Leuven, Belgium*

^c*European Commission JRC, Ispra, Italy*

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. We further investigate the role of large innovation “hubs” as attractors of new collaboration opportunities and the impact of region size and locality on the evolution of cross-border patenting activities. 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. Moreover, when analyzing networks of geographical mobility, multinational R&D activities and patent citations we cannot detect any substantial progress in European research integration above and beyond the common global trend.

Keywords: Geography of knowledge, Networks of Innovators, European integration, Spatial proximity, Cross-border collaboration, Gravity model.

*Corresponding author

Email address: andrea.morescalchi@imtlucca.it andrea.morescalchi@gmail.com (Andrea Morescalchi)

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 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 between inventors located in different countries and institutional settings.

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: (1) the network of individual collaborations among inventors (patent co-inventorship); (2) the internationalization of R&D activities by multinational firms (the patent applicant-inventor network), (3) knowledge spillovers as proxied by patent citations and (4) inventor mobility. Third, in our analysis we investigate, jointly, the physical distance effect and the country-border effect and control for other types of distance. Few previous studies have investigated the dynamics of the geographical distance and border-effects simultaneously and those that have focused 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

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

¹Overall, our approach poses some numerical challenges since we analyze a network of around 2.7 billion links: $5,552^2$ (potential links between 5,552 regional nodes) \times 4 (networks) \times 22 (years) = 2,712,573,952 possible connections.

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 analysis 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; Frenken et al., 2009b; Hoekman et al., 2009, 2010; Scherngell and Barber, 2011; Scherngell and Hu, 2011; Scherngell and Lata, 2012; Pan et al., 2012; Hoekman et al., 2013). 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; 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).

Table 1: Literature survey, selected contributions

author (year)	data source	time frame	network	method	distance	border	EU
Aldieri (2011)	USPTO	1979-2002	C	Poisson,li,cs	-, \nearrow
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. (2009b)	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	-, \nearrow	-	..
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	-	-	..
Maggioli and Uberti (2007)	EPO	1998-2002	A	OLS(g),NUTS2,s	-	-	..
Maurseth and Verspagen (2002)	EPO	1979-1996	C	Tobit/NB,NUTS1/2,s	-	-	..
Mignèlez and Moreno (2013)	EPO	1996-2001	A,M	OLS-ZINB,NUTS2,cs	-, \nearrow	-	..
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	-, \nearrow	-	..
Peri (2005)	USPTO	1975-1996	C	NB,NUTS1,s	-	-	..
Picci (2010)	EPO	1990-2005	A,F	OLS/Poisson(g),NUTS0,cs	-, \nearrow	-	+, \nearrow
Ponds et al. (2007)	WoS	1988-2004	A	NB(g),NUTS3,s	-
Ponds (2009)	WoS	1988-2004	A	Probit,,d
Scherngell and Hu (2011)	CNKI	2007	A	Poisson,NB(g),,,s
Singh and Marx (2013)	USPTO	1975-2004	C	WESML,CBSA,d	-, \nearrow	-	..

Notes: 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; li: longitude-latitude; M: mobility; NB: Negative Binomial; n.s.: not significant; NUTS: Nomenclature of Territorial Units for Statistics; OLS: Ordinary Least Square; OLS-ZINB=OLS,Poisson,ZINB; r.e.: rare event; s: statics; s.i.: spatial interaction; USPTO: US Patent and Trademark Office; WESML: weighted exogenous s ampling maximum likelihood; WoS: Web of Science; ZINB: Zero-Inflated Negative Binomial; ZIP: Zero-Inflated Poisson; \nearrow : increasing in absolute terms; \searrow : decreasing in absolute terms; =: approximately constant.

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; Ponds et al., 2007; 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 versus 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., 2009b) or a few countries (an exception being Hoekman et al., 2009 who analyze EU27 countries plus Norway

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

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

⁴The 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 nonEU member and partner states. For European countries, TL2 and TL3 are largely consistent with the Eurostat classifications NUTS2 and NUTS3 (Maraut et al., 2008).

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). In our analysis we investigate, jointly, the distance effect (Ponds et al., 2007) and the country-border effect (Ponds, 2009) in Europe and other OECD countries. Few previous studies have investigated the dynamics of the distance and border-effects simultaneously (Hoekman et al., 2010; Singh and Marx, 2013). Our analysis spans two decades from 1998 to 2009. Only a handful of very recent contributions adopted a long-term dynamic perspective to investigate the evolution of knowledge networks (see Table 1). 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 versus 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?
- 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?

To do that, we construct four networks (I-IV) using readily available information extracted from patents, which serve as representations of knowledge geography and provide quantitative structures for measuring knowledge diffusion.

(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; 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 as 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 firms scientist

⁵Beyond 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).

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 given that 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, 2013). 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, that account to a large extent for localization of co-inventions or citations (Miguèlez and Moreno, 2013).

Summarizing, patent citations offer evidence of the flow of codified knowledge. Mobile inventors, on the other hand, bring with them tacit and experience-based knowledge. Similarly, collaboration networks, either within or across organizational boundaries, imply sophisticated forms of coordination and knowledge transfer grounded in team and firm based interactions.

Based on the relative complexity of the knowledge being transferred in each of the innovation networks, the first hypothesis investigated in this paper is:

Hypothesis 1. R&D collaboration networks are the most constrained by distance, followed by inventor mobility and patent citations. Globalization forces will have an immediate effect on patent citations, followed by co-authorships. But it takes longer for inventors to move freely and firms to overcome home country bias through foreign R&D investment. Therefore, when the effect of globalization saturates, preferential attachment to leading regions will prevail, leading us to a second general hypothesis:

Hypothesis 2. The effect of geographical and institutional barriers is increasing over time in importance, particularly at national and regional levels and for small and mid-sized regions. The increase in strength of geographical and institutional barriers will be stronger and start earlier for patent citations followed by co-authorships, foreign R&D investments and eventually inventor mobility.

These effects have some compelling policy implications for the future of the European Research Area, and indicate priority should be given to unleashing the mobility of skilled labor force within Europe and boosting the global attractiveness of European research hubs.

3. Data and Methodology

3.1. Data

The data analyzed in this study are drawn from the OECD REGPAT database (Maraut 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 5,552 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 first names and initials have been dealt with, but have not been fully disambiguated. More precisely, our approach tracks the flow of names between regions which is a gross proxy of pure

⁶This 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.

⁷We 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.

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. 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. 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).

We describe our econometric approach by focusing on three key ingredients that are relevant to test our research hypotheses. 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 of interest.

3.2.1. Probability model for the intensity of R&D collaboration between regions

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 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

⁸This 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).

⁹We 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 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.

¹⁰See 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.

links, we opted for a Zero-Inflated Negative Binomial (ZINB) density, as consistent with Hoekman et al. (2009), Frenken et al. (2009b) 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

$$\begin{aligned} E(y_i|X_i) &= \left(1 - \frac{\exp(X_i\beta^0)}{1 + \exp(X_i\beta^0)}\right) * \exp(X_i\beta^1) = \\ &= \frac{\exp(X_i\beta^1)}{1 + \exp(X_i\beta^0)}, \end{aligned} \tag{1}$$

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 (see Appendix A for technical details on the ZINB model).

3.2.2. Independent variables

The selection of independent variables is relevant to 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. 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; Paci and Usai, 2009; Scherngell and Barber, 2011; Scherngell and Hu, 2011; Scherngell and Lata, 2012; Miguèlez and Moreno, 2013), and *area* (Peri, 2005; Frenken et al., 2009b; 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.¹¹

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, 2013). 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).¹²

3.2.3. Strategy to test the three research questions

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

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

¹¹Our 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.

¹²We 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.

where $j = 0, 1$ and $year_t$ are dummy variables that capture exogenous yearly shifts. 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 relevant spatial measure and year dummies. This allows the impact of spatial measures of our interest to vary year-by-year. Maximum likelihood estimates of parameters in the linear indices are then used in Eq. 1 to compute 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 the non-linear dependence of Y_i on X_i the computation requires to set specific values for each regressors. 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.¹³ 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. 2 and compute the elasticity of y_i with respect to *distance* over years.¹⁴ Specifically we estimate for each year the quantity

$$\begin{aligned} \epsilon_t &= \frac{(Y'' - Y')}{Y'}, \text{ where} \\ Y' &= E(y_i | X = \langle X \rangle, year_t = 1, distance = \langle distance \rangle), \\ Y'' &= E(y_i | X = \langle X \rangle, year_t = 1, distance = \langle distance \rangle * 1.01). \end{aligned} \quad (3)$$

Estimates of Y' and Y'' are obtained replacing in Eq. 1 parameter estimates from Eq. 2 and setting regressors values at sample means (*i.e.* $X = \langle X \rangle$) except for $year_t$ and $distance$.¹⁵

To test our second question (ii) we estimate the evolution of the country-border effect. To do this

¹³Robustness 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.

¹⁴The elasticity is formally defined as $\frac{\partial E(y_i)}{\partial distance} \frac{distance}{E(y_i)}$. From Eq. 1 we can compute the derivative as $\frac{\partial E(y_i)}{\partial distance} = E(y_i) \left(\beta_2^1 - \beta_2^0 \frac{\exp(X_i \beta^0)}{1 + \exp(X_i \beta^0)} \right)$. Thus we obtain $\epsilon_t = distance \left(\beta_2^1 - \beta_2^0 \frac{\exp(X_i \beta^0)}{1 + \exp(X_i \beta^0)} \right)$.

¹⁵See Winkelmann (2008) for the computation of marginal effects for the ZINB model.

we modify Eq. 2, adding interactions of *border* with year dummies, resulting in

$$\begin{aligned}
X\beta^j &= \beta_0^j + \beta_1^j border + \beta_2^j distance + \beta_3^j techdist_t + \beta_4^j distance * techdist_t + \beta_5^j neighbour \\
&+ \beta_6^j size_{m,t} + \beta_7^j size_{n,t} + \sum_{s=2}^S \gamma_s^j area_s + \sum_{t=2}^T \theta_t^j year_t + \sum_{t=2}^T \delta_t^j distance * year_t \\
&+ \sum_{t=2}^T \omega_t^j border * year_t.
\end{aligned} \tag{4}$$

Given maximum likelihood estimates of parameters in the augmented Eq. (4) we compute the marginal effects of the *border* variable over years. 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

$$\begin{aligned}
\pi_t &= \frac{(Y'' - Y')}{Y'}, \text{ where} \\
Y' &= E(y_i | X = \langle X \rangle, year_t = 1, border = 0), \\
Y'' &= E(y_i | X = \langle X \rangle, year_t = 1, border = 1).
\end{aligned} \tag{5}$$

Estimates of Y' and Y'' are obtained replacing in Eq. 1 parameter estimates from Eq. 4 and setting regressors values at sample means (*i.e.* $X = \langle X \rangle$) except for $year_t$ and $border$.

To test our third question (iii) we change the specification of Eq. (2) and separate the data into EU and non-EU sets, removing flows between the two. Linear indices are modeled following Chessa et al. (2013) who apply a Difference-in-Difference (DiD) strategy to isolate the country border effect within EU.¹⁶ 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. 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. 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.¹⁷ 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.¹⁸ The linear indices in 2 are

¹⁶See Chessa et al. (2013) for a more detailed description of the methodology.

¹⁷Using 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.

¹⁸Here 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 nonEU) is important to take into account in estimation possible differences in the integration behavior of small regions.

now modeled as

$$\begin{aligned}
X\beta^j = & \beta_0^j + \beta_1^j border + \beta_2^j eu + \beta_3^j distance + \beta_4^j size_m + \beta_5^j size_n + \gamma^j border * eu + \\
& + \sum_{t=2}^T \theta_t^j year_t + \sum_{t=2}^T \delta_t^j border * year_t + \sum_{t=2}^T \zeta_t^j eu * year_t + \sum_{t=2}^T \eta_t^j border * eu * year_t, \tag{6}
\end{aligned}$$

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. In particular, *eu* flags pairs of NUTS3 regions that are within the EU ($eu = 1$) and pairs of NUTS3 regions for which neither are in the EU ($eu = 0$). The dummy variable *border* still flags pairs of NUTS3 regions within the same country but now links pertain always to the same area (EU or non-EU) whether or not they are cross-border or within-border. For example, Italy-France can be a valid cross-border link for EU and USA-Japan can be a valid cross-border link for non-EU. However Italy-USA, Italy-Japan, France-USA and France-Japan are excluded. In terms of the standard DiD formalism (Angrist and Krueger, 1999; Heckman et al., 1999; Athey and Imbens, 2006; Blundell and Costa Dias, 2009) the dummy *border* distinguishes treated units, *i.e.* regions belonging to different countries, from un-treated units, *i.e.* regions within the same country. The impact on cross-border collaborations arising only from EU factors is netted out by providing a further comparison group of links between non-European countries, that implies an extension of the standard DiD strategy of one state indicator (treatment vs control group) to the case of two state indicators. This is achieved by including the dummy *eu*. The inclusion of year dummies and its interactions with the two state indicators (*border* and *eu*) is relevant to embed the institutional comparison in a temporal perspective. Due to the addition of a second state indicator the approach turns out to be a Difference-in-Differences-in-Differences estimator (DiDiD). 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

$$\begin{aligned}
\tau_t = & \left(Y'' - Y' \right), \text{ where} \\
Y' = & E(y_i | X = \langle X_{border, eu, t^*} \rangle, border = 1, eu = 1, year_{t^*} = 1), \\
Y'' = & E(y_i | X = \langle X_{border, eu, t^*} \rangle, border = 1, eu = 1, year_t = 1). \tag{7}
\end{aligned}$$

Estimates of Y' and Y'' are obtained replacing in Eq. 1 parameter estimates from Eq. 6 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 $X = \langle X_{border, eu, t^*} \rangle$), 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^* .

3.3. Contextual decomposition of the coinventor data into size, distance, and time subsets

In order to provide further insight into the contextual evolution of the distance effect, we analyzed and compared the changes in the co-inventor network for specific non-overlapping subsets of the aggregate data.

Specifically, we first separated the network data into three time 8-year groups: $T_1 = 1986 - 1993$, $T_2 = 1994 - 2001$, and $T_3 = 2002 - 2009$. Secondly, we separated the coinventor pairs into 3 distance regimes: “small” ($d_{m,n} \leq 100$ km, corresponding to $g_d = 0$), “medium” ($100 < d_{m,n} \leq d_{EU}$ km, corresponding to $g_d = 1$), and “long” ($d_{m,n} > d_{EU}$ km, corresponding to $g_d = 2$). Our choice of the thresholds separating the three distance groups is motivated by the empirical distribution of $d_{m,n}$ values shown in Fig. 1(A), which appears to be a mixture of at least three underlying distributions: one representing the inter-regional scale, one representing a local intercontinental/national scale, and one representing a global scale. Hence, we choose a d_{EU} value to represent the characteristic distance of EU-EU links, which we estimate by calculating the average $d_{m,n}$ value for distinct region-region collaborations (not weighted by copatent link intensity $y_{m,n}$) over each period T_i and also for T_{all} : $\langle d_{m,n}^{EU}(T_1) \rangle = 376$, $\langle d_{m,n}^{EU}(T_2) \rangle = 478$, $\langle d_{m,n}^{EU}(T_3) \rangle = 531$, $\langle d_{m,n}^{EU}(T_{all}) \rangle = 562$. We choose to use the largest of these four values, $d_{EU} \equiv \langle d_{m,n}^{EU}(T_{all}) \rangle = 562$ km.

Finally, we separated regions according to patent productivity as a proxy for size S_n : “ $S = Small$ ”, “ $M = Medium$ ”, or “ $L = Large$ ”. Figure 1(B) shows the probability density function $P(S)$ calculated by aggregating S_n values into the three time periods defined above; the overall functional form of $P(S)$ appears to be stable over time, but with a slight shift in the range reflecting increasing patent activity over the entire period. The partitions in S_n were chosen such that regions with $S_n \geq 419$ patent counts (corresponding to the top 100 regions in the three-year time period 1986 – 1988) were defined as “Large”. These large “hubs” account for $f_L = 48\%$ of the total patents in period T_1 . The threshold between “Small” and “Medium” was defined such that regions in each of these two groups equally share the remaining $100 - 46.8 = 52\%$ of the net patents, hence $f_M = f_S = 0.26$. Using this definition, the cutoff between small and medium sized regions is $S_c = 136$ patent counts (in terms of region rank, this corresponds to the rank threshold $r_c = 321$, meaning that regions ranked 101-321 fall into the “medium” size group). Using these definitions, each collaboration count $y_{m,n}$ can be distributed into one of 6 partitions: SS, MS, MM, LS, LM, LL .

In summary, we disaggregated the data across 3 consecutive 8-year periods, 3 distance scales, and 6 permutations of region-region size pairing, for a total of 54 data subsets. Then, for each subset we measured the distance effect ρ by estimating the parameters of the basic gravity model,

$$\ln y_{(m,n)} = \ln \beta_0 + \rho \ln d_{m,n} + \beta^+ \ln \text{Max}[S_m, S_n] + \beta^- \ln \text{Min}[S_m, S_n]. \quad (8)$$

Hence, each data subset provides a comparable estimate of $\rho(T_i, g(S_m, S_n), g_d)$ which depends on the time period group T_i , the size pairing group $g(S_m, S_n)$, the distance group g_d .

The results of this contextual analysis are shown in Fig. 1(C). Each datapoint is colored according to the size partition, and the size of each datapoint is proportional to the average $\log d_{ij}$ of the data

entering into the calculation of a given ρ value. A decrease in the magnitude of ρ over time is consistent with distance becoming less of a collaboration barrier. To provide a measure for the change in ρ over time, we calculated the percent change between consecutive periods,

$$\% \Delta \rho(T_i) = 100 \times [\rho(T_i) - \rho(T_{i-1})] / \rho(T_{i-1}) , \quad (9)$$

which is shown for each subset in Figures 2 and 3. There are six types of cross-border links: EU-EU, EU-World, EU-USA, World-World, World-USA, and USA-USA, where World represents all regions not in either the EU or USA groups. Each subpanel is a pie-chart with the total number of link counts $y_{m,n}$ for each division in period T_i . The fraction f_{EU-EU} of the total links that are EU-EU links and the percent change in this quantity between consecutive periods,

$$\% \Delta f_{EU-EU}(T_i) = 100 \times [f_{EU-EU}(T_i) - f_{EU-EU}(T_{i-1})] / f_{EU-EU}(T_{i-1}) , \quad (10)$$

are listed at the top of each subpanel. For comparison, the fraction $f_{USA-USA}$ and percent change between periods is also listed at the bottom of each subpanel. The colored number listed in the center within each pie-chart is the percent change $\% \Delta \rho$ calculated between consecutive time periods. To guide the eye, the $\Delta \rho(T_i)$ values are colored: the the largest positive growth values (meaning that the ρ value increases in magnitude indicating an increasing collaboration barrier due to distance) are colored red and the largest negative values (meaning that the ρ value decreased in magnitude indicating a decrease in the collaboration barrier due to distance) are colored light blue. Figure 2 shows the values for T_1 and T_2 and Figure 3 shows the values for T_2 and T_3 .

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 discuss the main conclusions of our work about the two main research hypotheses we stated at the end of Section 2.

4.1. Evolution of the distance effect

Distance is a crucial aspect of knowledge transfer when the flow of knowledge is based on human interactions (*i.e.* co-inventor, applicant-inventor) or mobility. While these networks require costs of communication or costs of moving, citations on the other hand benefit from the availability of online repositories of bibliometric records which only require an access cost. Figure 4 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 1,267 kilometers in 1986 and has almost doubled by 2009, increasing to 2,051 kilometers. Note that we do not use inter-regional links ($d_{nn} = 0$) 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.

Figure 1(A) shows for each year t the probability density function $P_t(d_{m,n})$ calculated by weighting the coinventor link distance $d_{m,n}$ between region m and n by the link occurrence, $y_{m,n}$. Note the significant shift in the increased likelihood of large $d_{m,n}$ over time which occurs especially in the right tail of the distribution. For example, comparing 1986 and 2009 indicates roughly a factor of 10 increase in the likelihood of large distance $d_{ij} > 2000$ km collaborations. 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.

Interestingly, in double logarithmic scale, it is evident that each $P_t(d_{m,n})$ can be characterized by at least three underlying distributions. Aggregated, these distributions result in an overall triple-humped distribution shape. For example, approximately 8% of the probability density belongs to the group $g_d = 2$ of long-distance collaborations spanning more than $d_{EU} = 562$ km. Notably, there is a sharp cutoff feature $\approx 6,000$ km, representing the trans-pacific/atlantic length scale. About 82% of the coinventor link counts belong to the small distance $g_d = 0$ group, with $d_{m,n} \leq 100$ km. The remaining 10% of link counts fall into the medium distance regime, chosen to represent the characteristic distance scale for Europe.

Having demonstrated gradual shifts in the frequencies of $d_{m,n}$ over time, we now shift our focus to the impact that distance has on the likelihood of collaboration. Figure 5 shows network-by-network estimates of ϵ_t , the elasticity of $y_{m,n}$ with respect to $d_{m,n}$, as defined in Eq. (3). 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 mobility network, and a 0.27% decrease for the citation 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, in line with our first research hypothesis.

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 Figure 5). 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.¹⁹ For mobility, there is a significant decrease from 1998-1997, then stabilizing afterward. Figure 5 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

¹⁹The temporal evolution of the distance effect is very similar when using the augmented Eq. 4, *i.e.* when the temporal evolution of the cross-border effect is accounted for in the model. Results are available upon request.

in the average distance of links shown in Figure 4.²⁰

Furthermore, Figure 6 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.²¹

Do large patent hubs attract collaborations differently than medium and small regions? In order to gain insight into the size dependency of the distance effect, and in particular the role that large “hub” regions play in the evolution of cross-border collaboration, we partitioned the co-inventor network data into 54 different datasets depending on the sizes of the regions collaborating (6 size-size groups), the distance over which the co-inventors collaborated (3 distance groups), and the time period (3 consecutive 8-year periods). The region size proxy S_n is defined as the number of patents from region n in each specified time period T_i .

For each subset we calculated the relative change in $y_{m,n}$ with respect to a relative change in distance, using the benchmark gravity model defined in Eq.(8) to estimate the distance effect parameter ρ . Figure 1(C) shows the 54 regression estimates for $\rho(d_{ij}, t)$, which are separated into 3 panels depending on the d_{ij} group. Each datapoint is colored according to the size partition, and the size of each datapoint is proportional to the average $\log d_{ij}$ of the data entering into the calculation of that given ρ value. For a specific data subset, a decrease in the magnitude of ρ over time is consistent with distance becoming less of a collaboration barrier for a given type of inventor collaboration. Next we list our main observations by distance group.

Short-distance collaborations ($g_d = 0$): For short-distance collaborations, the magnitude of ρ is increasing for all the groups (LS , LM , and LL) that involve large hub regions. Hence, at short distance, the relative effect of distance is becoming a stronger impediment to these intra-regional collaborations involving large hubs. The relatively large magnitude increase in ρ for the LL group from period T_1

²⁰The 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 or groups of similar countries. We estimated several models using different sub-samples 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.

²¹Continuing 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. Figure 6 refers to the applicant-inventor network. A similar plot for co-inventorships is available upon request.

to T_2 may be due to the competitive aspects of hub “attractor” regions, which attract a significant portion of their collaboration activities locally, and in doing so may shield small regions from potential collaborations with other hubs.

Within the subset of LL links representing collaboration between nearby hubs, an increasing share is being represented by EU-EU links (the percent change $\% \Delta f_{EU-EU} = 56.5\%$ and 28.6% for the T_1 to T_2 and T_2 to T_3 periods, respectively). However, this pattern of growth within the LL link subset was also present for USA-USA links, which dominates the EU fraction by roughly a factor of two, indicating that the EU is significantly behind the US in this category.

Medium-distance collaborations ($g_d = 1$): For medium-distance collaborations, corresponding to the characteristic scale of Europe, the magnitude of ρ decreased for the LL and LM groups from T_1 to T_2 , but then stagnated from T_2 to T_3 . Meanwhile, for the SS and MS groups, the ρ value increased from T_1 to T_2 and also from T_2 to T_3 , indicating that there is a decreasing likelihood of these types of collaborations, possibly due to crowding out by large regions. Interestingly, from T_2 to T_3 , the magnitude of ρ increased for all groups except for LL .

Concerning the relative number of EU-EU versus USA-USA links, the trends vary across time period. For MM , the EU increased its share two-fold relative to the USA, from $f_{EU-EU}/f_{USA-USA} = 2.2$ in T_2 to $f_{EU-EU}/f_{USA-USA} = 4.4$ in T_3 . For LS , the USA increased its share slightly with respect to the EU. For LM , both the EU and the USA increased their share, however the ratio sustained the value $f_{EU-EU}/f_{USA-USA} = 2.3$. For LL , again both regions increased their share, however the ratio decreased from $f_{EU-EU}/f_{USA-USA} = 1.3$ in T_2 to 0.9 in T_3 . This last feature demonstrates a trend in the US for large hubs to increasingly collaborate, showing a 1132% growth from period T_1 to T_2 and sustaining a 157% growth from period T_2 to T_3 , reflecting the competitive cross-border complementarity that the 2013 European Research Area Progress Report highlights as a key criteria for transnational development of the ERA (see European Commission (2013a,b)). It is possible that certain EU funding criteria that require cross-border collaborations may come at the expense of decreasing the complementarity of the match, and hence the competitiveness of the collective. These types of policies may be orthogonal to the market-based collaboration forces existing in the rest of the world, and may further delay or send the EU off course in its ambition to create an effective, efficient, and geographically distributed ERA.

Long-distance collaborations ($g_d = 2$): For large-distance collaborations, the magnitude of ρ significantly decreased for all groups from T_1 to T_2 , but much less so from T_2 to T_3 , suggesting a stagnation in the role of distance at the global scale, commiserate with the saturation demonstrated in Figures 1(A) and 4.

For both the LM and MM subsets, there is significant growth in f_{EU-EU} from T_1 to T_2 and also from T_2 to T_3 . There are also positive signs of global integration, as the LL link counts indicate a significant increase in the EU-USA share (27.8% growth) and also the EU-World share (21% growth) from T_2 to T_3 . There is also an increase in the share of EU-EU LM and MM links from T_2 to T_3 . Nevertheless, the USA-USA links dominate the network at the global scale, commiserate with the findings of the 2013 Science Europe and Elsevier SCIVAL Analytics report on inter-country/state

coauthorship (see Science Europe (2013)). Concerning the rest of the world, the fraction $f_{W-USA} > f_{EU-W}$ if any of the two regions is L, and $f_{W-USA} < f_{EU-W}$ otherwise, indicating that the USA has a higher outcome rate from teaming scenarios between large hubs and smaller regions. Such teaming schemes, also known as “research buddy” model, is at the core of EU Horizon2020 policies aimed at raising the quality and competitive level of low-performing member state regions by partnering them with established European hubs (see Institutional Consortium (2013)).

Summarizing this subsection on the the role of distance, the overall trend is that most of the changes in ρ occur from T_1 to T_2 , but there is stagnation in the final period T_2 to T_3 . Furthermore, by analyzing the trends in the specific subsets of link types, a question emerges concerning EU cross-border collaboration policies: are EU policy efforts consistent with external global trends? If not, then EU policies aimed at spurring ERA development may possibly be orthogonal to, and/or significantly lagging, the economic, knowledge, and labor market forces that influence cross-border collaboration patterns globally. We will address this issue in the results subsection 4.3.

4.2. Evolution of the border effect

In Figure 7 we report the evolution of the border effect. These plots report the yearly semi-elasticity π_t as defined in Eq. (5), *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.²² Thus, the effect of country borders is clearly quite strong. Similar to the distance effect and in line with our research hypothesis 1, 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 and coherently, 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 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

²²As a computation example of these quantities, for the coinventor network in 2008, $\pi_{2008} = -88.7\%$ is obtained from $(Y'' - Y')/Y' = (0.0688 - 0.6085)/0.6085 = -0.8869$.

small regions.²³ In Figure 8 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-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 Top100-NonTop100 share also shows the most significant increase of the three subgroups, suggesting 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. However, links between top regions (Top100-Top100) rarely occur across borders, with only a mild increase over the period we examine.²⁴ This suggests that borders have a context-dependent size-dependency, one which shields large competitive hubs from each other, but is less of a barrier for small regions seeking to pair with a hub.

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. By way of example, consider the COST (European Cooperation in Science and Technology) programs that were initiated in the 1970s to spur intergovernmental European cooperation, and the Cooperation Programme that was embedded within recent Framework Programmes (FPs) to make explicit the emphasis on cross-border activities. Moreover, cross-border R&D activities are not only important to the EU, but reflect overall globalization trends. As the prevalence of multi-institutional and multinational 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 (2013b,a); Science Europe (2013)).

In the previous sections we aggregated all OECD countries, focusing on the evolution of the border effect on a global scale. However, the 21 year time span of our networks provides the added opportunity to observe, analyze, and interpret long-term trends in the context of European research integration policies as well as other global factors. Hence, 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.

Figure 9 shows estimates of τ_t as defined in Eq. (7), 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

²³The percentages of all links pertaining to the three groups are 44.1% (Top100-Top100), 18.2% (Top100-NonTop100), 37.7% (NonTop100-NonTop100) respectively.

²⁴A similar reaching out effect has been already noticed for the US Life Sciences patent network by Owen-Smith et al. (2002)

OECD countries. This effect was remarkable since the mid Nineties, but has stalled since 2004. 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. However, analogous trends in the other networks are difficult to identify.²⁵ 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.²⁶ 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.²⁷

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 has been some relevant changes of the global tendency over time. First, results in Figures 1 and 4 show that the increase of the average distance saturated across all networks. The outcome of our contextual analysis, as depicted in Figures 1, 2 and 3, reveals that the slowdown of the globalization trend compounds with the emergence of innovation hubs, which attract connections on a global scale. The second effect has induced an increase in the effect of geographical and institutional barriers since mid-Nineties, as shown in Figures 5, 6 and 7. Coherently with our second research hypothesis this effect appears to be stronger for small and mid-sized regions on national and regional geographical scales. As predicted, Table 2 shows that the increase in geographical and institutional barriers is stronger and starts before for codified knowledge (patent citations) whereas it is weaker and more recent for tacit knowledge (inventor mobility) with other R&D networks as intermediate cases.

Looking jointly at the trends in Figure 7 and Figure 9 for the country-border effect in the global patent activity, we can draw interesting insights about the impact that EU-specific integration has on the global evolution of the border effect. Positive trends in Figure 7 can be driven by: stronger integration in type (a) links (EU-EU), stronger integration in type (b) links (nonEU-nonEU), larger weight of type (c) links (EU-nonEU). Though the individual contribution of these links on the global patterns cannot be quantified from Figure 9, we can grasp whether the relative contribution of (a) versus (b) links reflects into the sign of global trends. In particular, we note that when the trend

²⁵In 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.

²⁶These 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.

²⁷Figures are made available by the authors.

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	=	
				1999-2009	↗				
(III) Citations	1988-2002	↗↗	↗↗	1988-2002	↘↘	↘	1986-1996	=	↗
	2002-2009	=		2002-2009	↗↗		1996-2000	↗↗	
							2000-2005	↘	
						2005-2009	↗		
(IV) Mobility	1988-1997	↘↘	↘↘	1988-1995	↘↘	↘↘	1990-1992	↘	=
	1997-2009	=		1995-2009	=		1992-2008	=	

Notes: ↗↗ significantly increasing, ↗ mildly increasing, = stable, ↘ mildly decreasing, ↘↘ significantly decreasing. This Table summarizes the main trends identified for the three sets of estimates reported in Figures 5, 7 and 9.

in Figure 9 becomes positive, *i.e.* integration in (a) is faster than (b), this typically reflects into movement of the opposite sign in Figure 7. In the co-inventor network, faster EU relative integration in 1995-2004 is consistent with a decline in the global trend between 1993 and 2004. In the applicant-inventor network, a relative reduction in the border effect in EU in 1986-1997 pairs a global reduction in 1988-1997. In the citation network, a significant relative EU integration in 1996-2000 can be associated to a significant decline in the global effect in the same period. However, we can also identify significant declines in the global country-border effect that are not associated to faster integration in EU, suggesting that the impact of EU-specific institutional factors can provide only a partial explanation of the lessening on a global scale. This leaves some scope for the role played by the erosion of formal and informal institutional barriers at a global level. This interpretation seems to apply specifically to the mobility network for early years of the sample. In particular, 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 Section.

5. Final Discussion

We analyzed the temporal evolution of spatial biases in the strength of inter-regional connectivity within a set of regional patent networks. Focusing on a set of fifty 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 in the frequencies of observed inter-regional links distance has been in the extreme tail of the inter-regional distance distribution. We observe roughly a factor of 10 increase in the likelihood of collaborations for distances above 2,000 km when comparing 1986 to 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. Hence, in light of these insights, it is possible to say that distance is not necessarily “dead”, but that distance has at the least “saturated”. Nevertheless, the role of distance is still a rather important factor that distinguishes the size groups of regions we analyzed, and provides insight into the role of large “hubs” which can readily attract over long distance, but also shield each other competitively over short distances. In light of our decomposition of the coinventor data into size, distance, and time subsets, we urge policy makers to consider the subtle but important roles of context when evaluating progress in cross-border collaboration and mobility. Within large datasets, there can be niche patterns that are overlooked when data are pooled without controls and/or decomposed into conditional subsets. Our decomposition method sheds light on the size-dependent mechanisms of preferential growth and attachment in dynamic innovation networks.

To this end, by controlling for numerous variables and analyzing conditional subsets of the network data we were able to gain new insight in to the mechanisms underlying the evolution of inter-regional connectivity. 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. 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.

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)). Indeed, we find that the distance effect is becoming larger in magnitude for medium-distance collaborations between small/medium regions pairing with large regions (see Figure 3, where all $\% \Delta \rho$ values are positive in the $g_d = 1$ column, with the exception of the *LL* group).

The smart specialization policies should take into account for the size-dependency of cross-border collaboration. We find evidence at the short distance scale that large regions may competitively shield surrounding smaller regions, meaning that a beneficial interaction between a small/medium sized region and a large region may not develop because of the competition between large hubs over short scales. Hence, investing in the development of large attractive hubs may lead to inefficient dilemmas within regions with a high density of hubs, an issue that should be considered as the EU encourages smart specialization policies in order to effectively, efficiently, and homogeneously develop the ERA.

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 spread out (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, 2013b).

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). Also, both effects increase almost hand-in-hand since the late 1990s. An increase in these effects means that, once we account for the effect of size and other variables, inventors are more likely to be located nearby and in the same country as their institution (patent applicant). Excessive geographical dispersion of learning centers 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). This trend is in line with smart specialization strategies as they can be a valuable asset to speed up the creation and consolidation of a European Research Area. 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 benefits in the European Union as opposed to the rest of the developed world.

However, it is also important to stress that policies embedded within EU funding programs aimed at transnational cooperation (see European Commission (2013b)) 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 *LS* and *LM* size pairing over medium distances ($g_d = 1$), the EU has significantly larger fraction of EU-EU links than the US has US-US links (roughly twice as many, see Figs 2 and 3). Conversely, for the large distance ($g_d = 2$) group, the EU-W link counts are significantly outnumbered by the USA-W 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. 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] and the Research Foundation of Flanders (FWO) (G073013N). O.P. acknowledges funding from the Social Sciences and Humanities Research Council of Canada.

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.
- Angrist, J., Krueger, A., 1999. Empirical strategies in labor economics, in: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Elsevier. volume 3A, pp. 1277–1366.

- Athey, S., Imbens, G., 2006. Identification and inference in nonlinear difference-in-differences models. *Econometrica* 74, 431–497.
- 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., Krieckemeier, U. (Eds.), *Palgrave Handbook of International Trade*. Palgrave Macmillan Press, New York.
- Blundell, R., Costa Dias, M., 2009. Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources* 44, 565–640.
- 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 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, 2013a. *European Research Area Facts and Figures 2013*. Technical Report. European Commission.
- European Commission, 2013b. *European Research Area Progress Report 2013*. Technical Report. European Commission.
- Feldman, M.P., 1994. *The geography of innovation*. volume 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., 2009b. 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.
- Heckman, J., LaLonde, R., Smith, J., 1999. The economics and econometrics of active labor market programs, in: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*. Elsevier. volume 3A, pp. 1865–2097.
- 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.
- 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. chapter 10, pp. 321–348.
- 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.

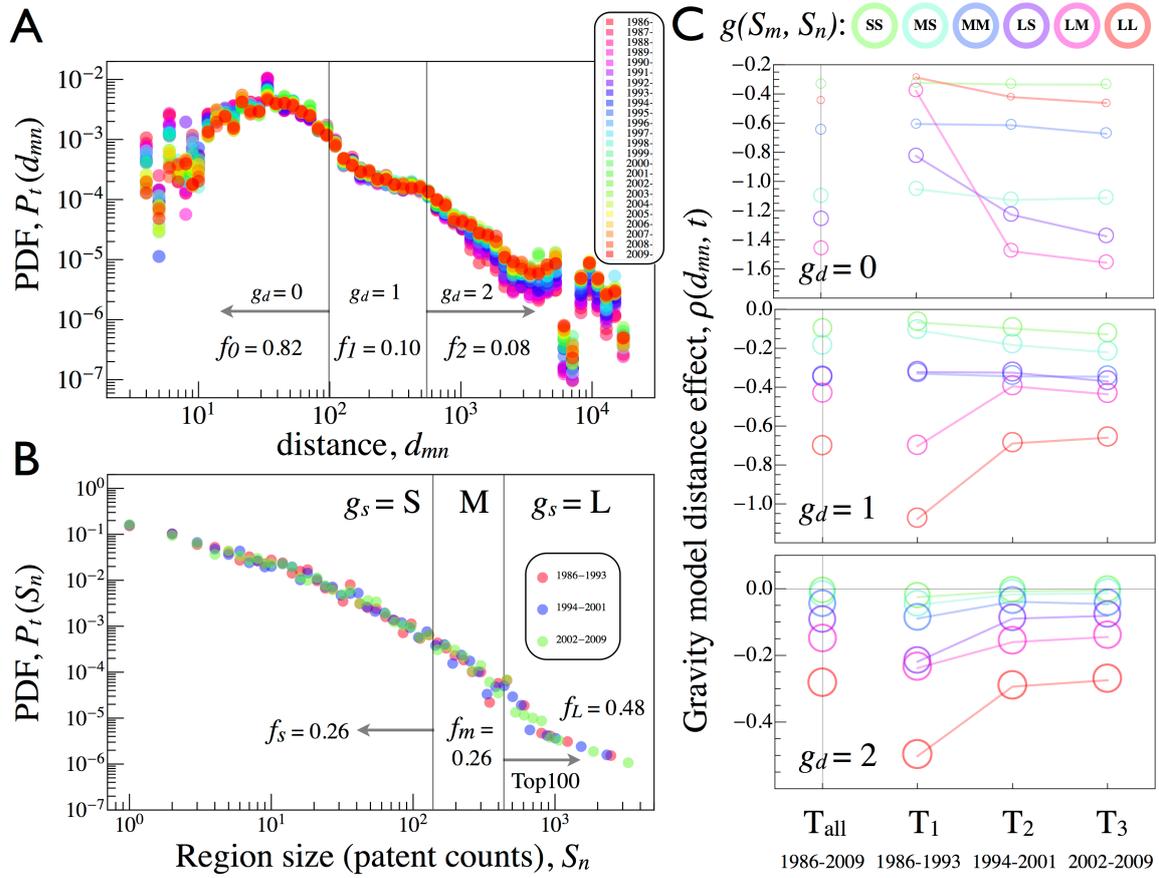
- LeSage, J.P., Fischer, M.M., Scherngell, T., 2007. Knowledge spillovers across europe: Evidence from a poisson spatial interaction model with spatial effects*. *Papers in Regional Science* 86, 393–421.
- Long, S., Freese, J., 2006. *Regression Models for Categorical Dependent Variables Using Stata* (2nd edition). Stata Press, College Station Texas, US.
- 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.
- Maraut, 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.
- Miguèlez, E., Moreno, R., 2013. Do labour mobility and technological collaborations foster geographical knowledge diffusion? the case of european regions. *Growth and Change* 44, 321–354.
- 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. *Economic 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.
- 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., Oort, F.V., Frenken, K., 2007. The geographical and institutional proximity of research collaboration. *Papers in Regional Science* 86.
- 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. chapter 3, pp. 56–85.
- 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.
- 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* (forthcoming) .
- Science Europe, 2013. Comparative benchmarking of European and US research collaboration 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 diffusion: Political borders vs. spatial proximity. *Management Science* 59, 2056–2078.

- 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., Reinstaller, 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.

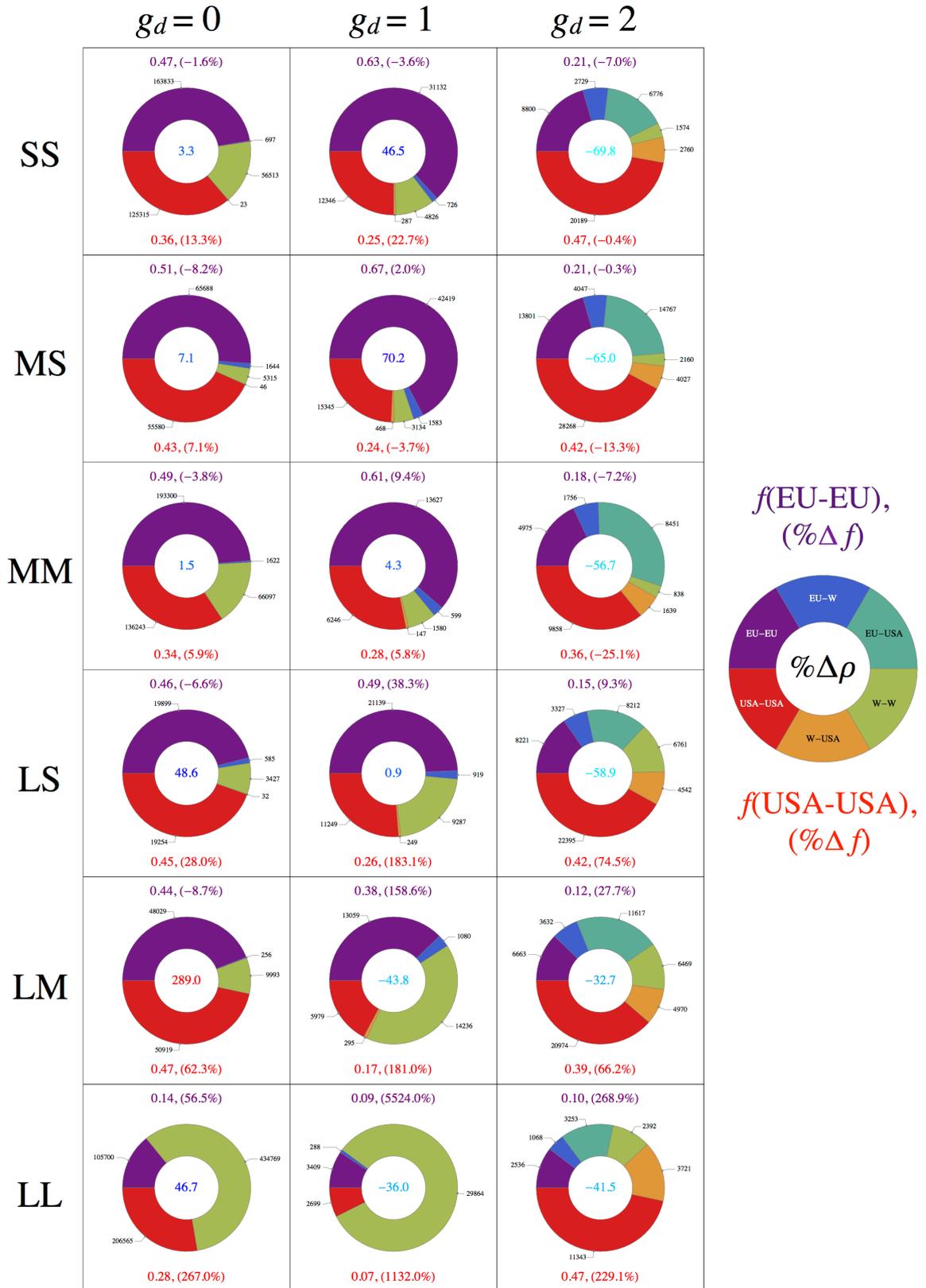
Figures

Figure 1



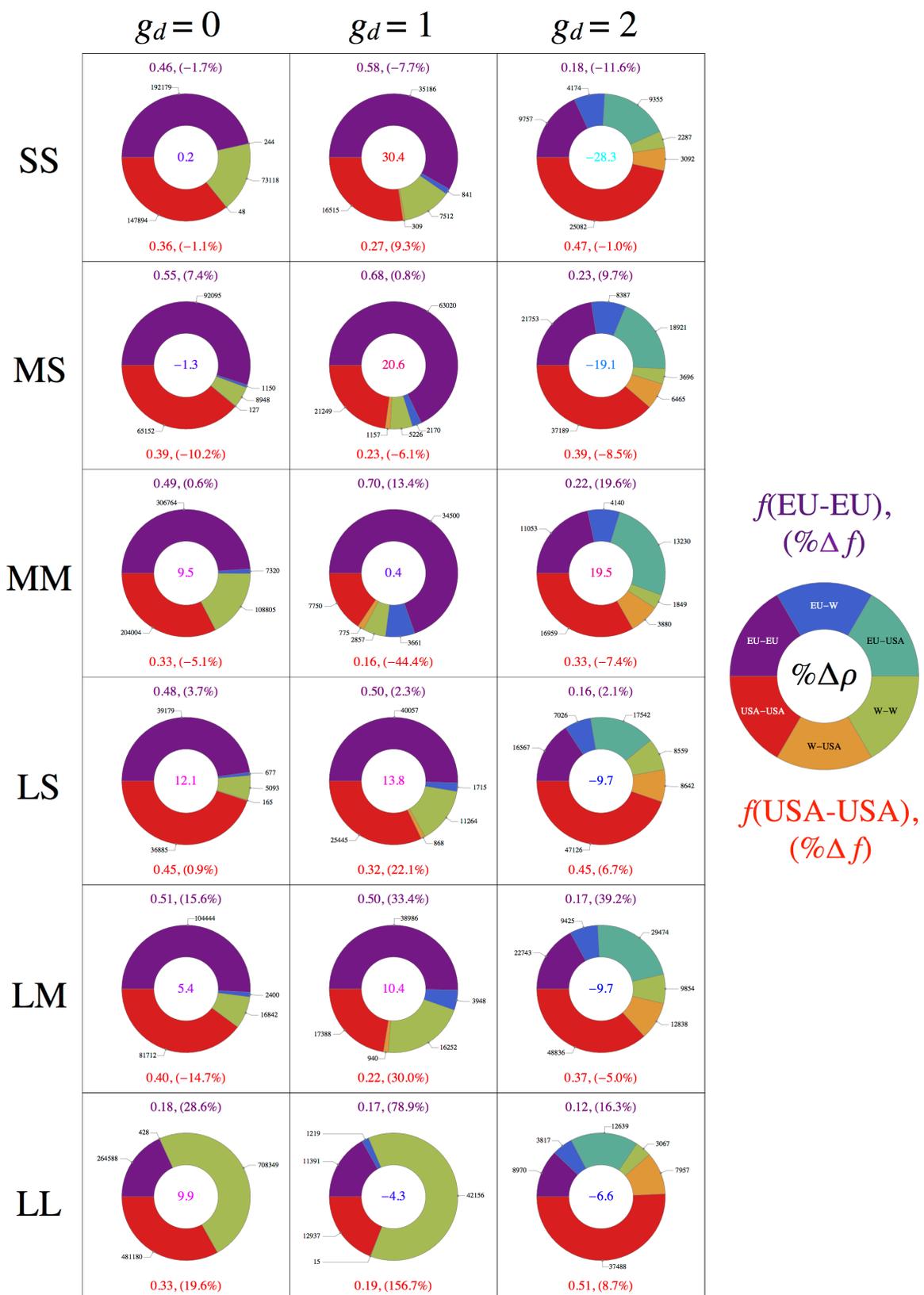
Notes: 3 distance regimes were defined: “small distance” ($d_{m,n} \leq 100$ km, corresponding to $g_d = 0$), “medium distance” ($100 < d_{m,n} \leq d_{EU}$ km, corresponding to $g_d = 1$), and “large distance” ($d_{m,n} > d_{EU}$ km, corresponding to $g_d = 2$). The value d_{EU} is chosen to represent the characteristic distance of EU-EU links, using $\langle d_{m,n}^{EU}(T_{all}) \rangle = 562$ km.

Figure 2



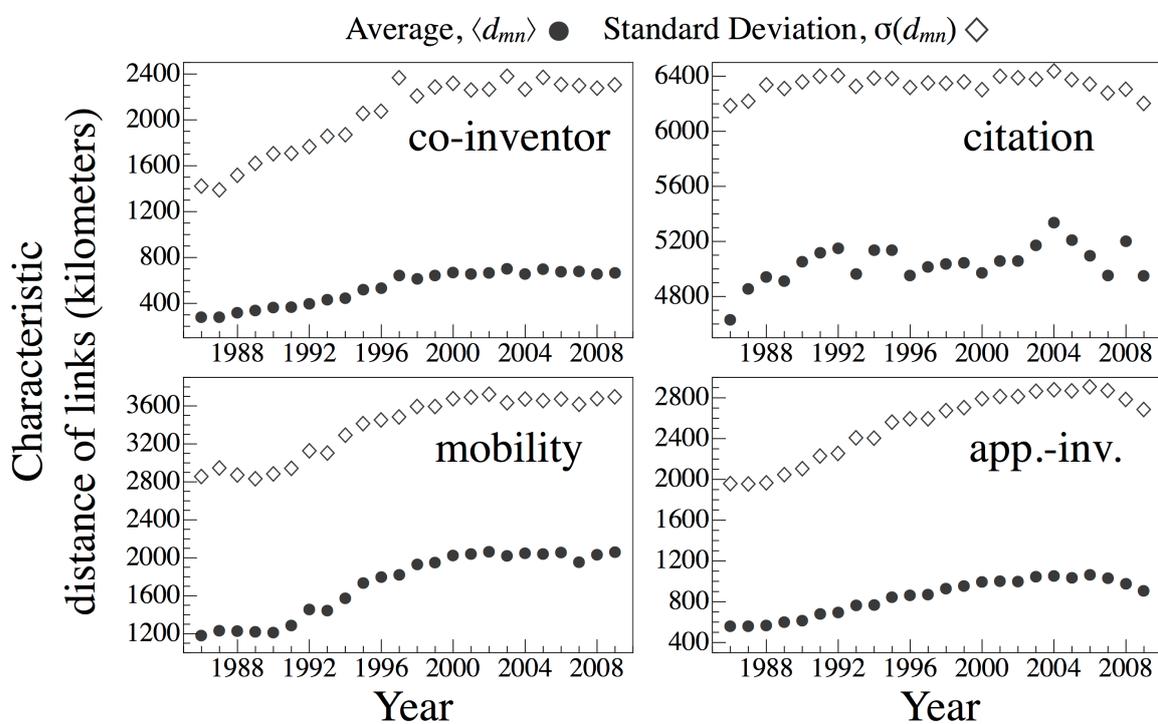
Notes: % changes calculated over period T1 to T2. Shown are link distributions for T2.

Figure 3



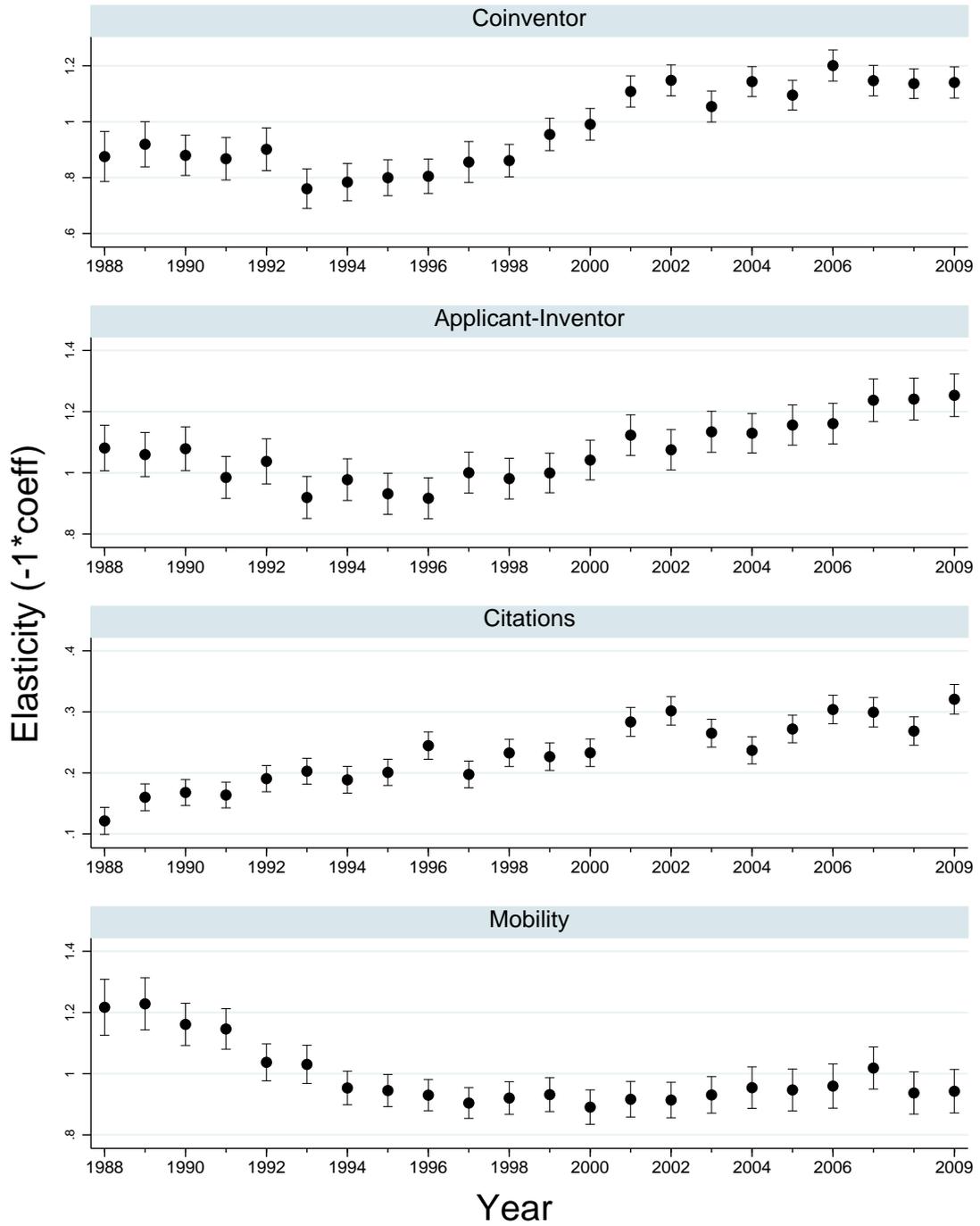
Notes: % changes calculated over period T2 to T3. Shown are link distributions for T3.

Figure 4



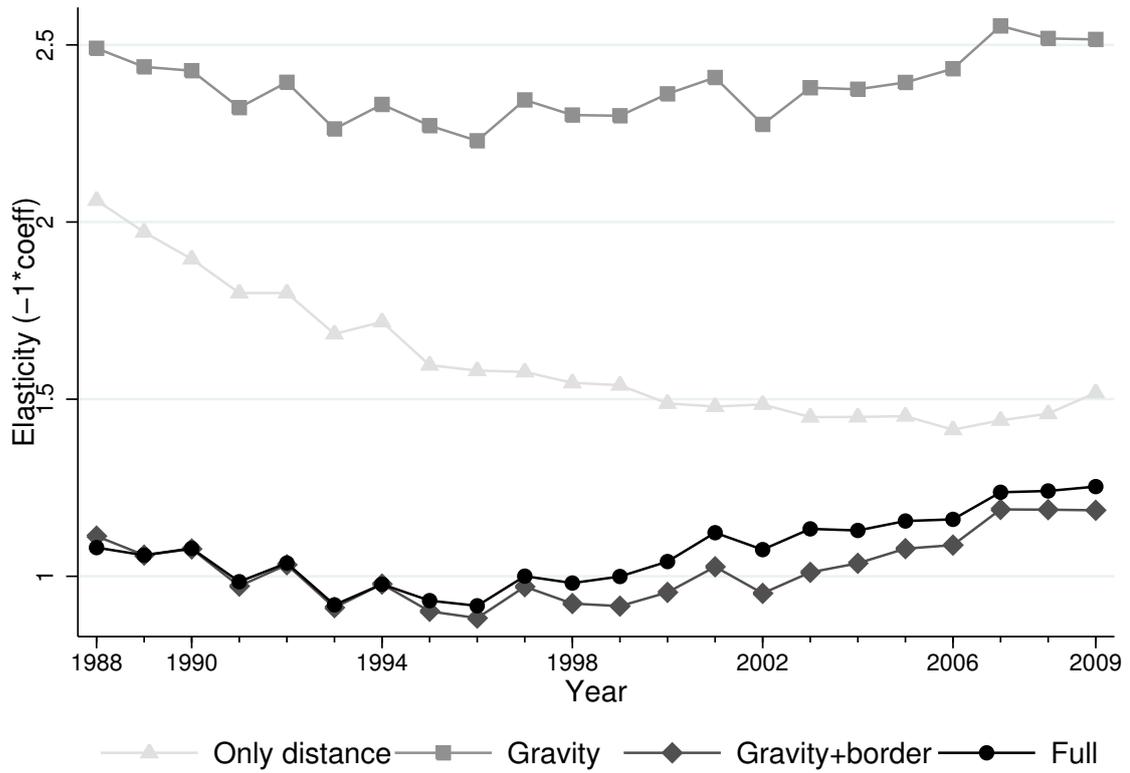
Notes: All links between NUTS3 regions are used to compute the average distance. For inventors mobility self-loops are removed as not meaningful.

Figure 5



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 B.3 in Appendix B regression estimates of parameters in Eq. 2 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 $\epsilon_{2008} = (Y'' - Y')/Y' = (0.11061 - 0.11187)/0.11187 = -0.0113 = -1.13\%$.

Figure 6



Notes: Each line plot refer to estimates of different ZINB model (see notes to Figure 5). We report below the set of regressors used in each case:

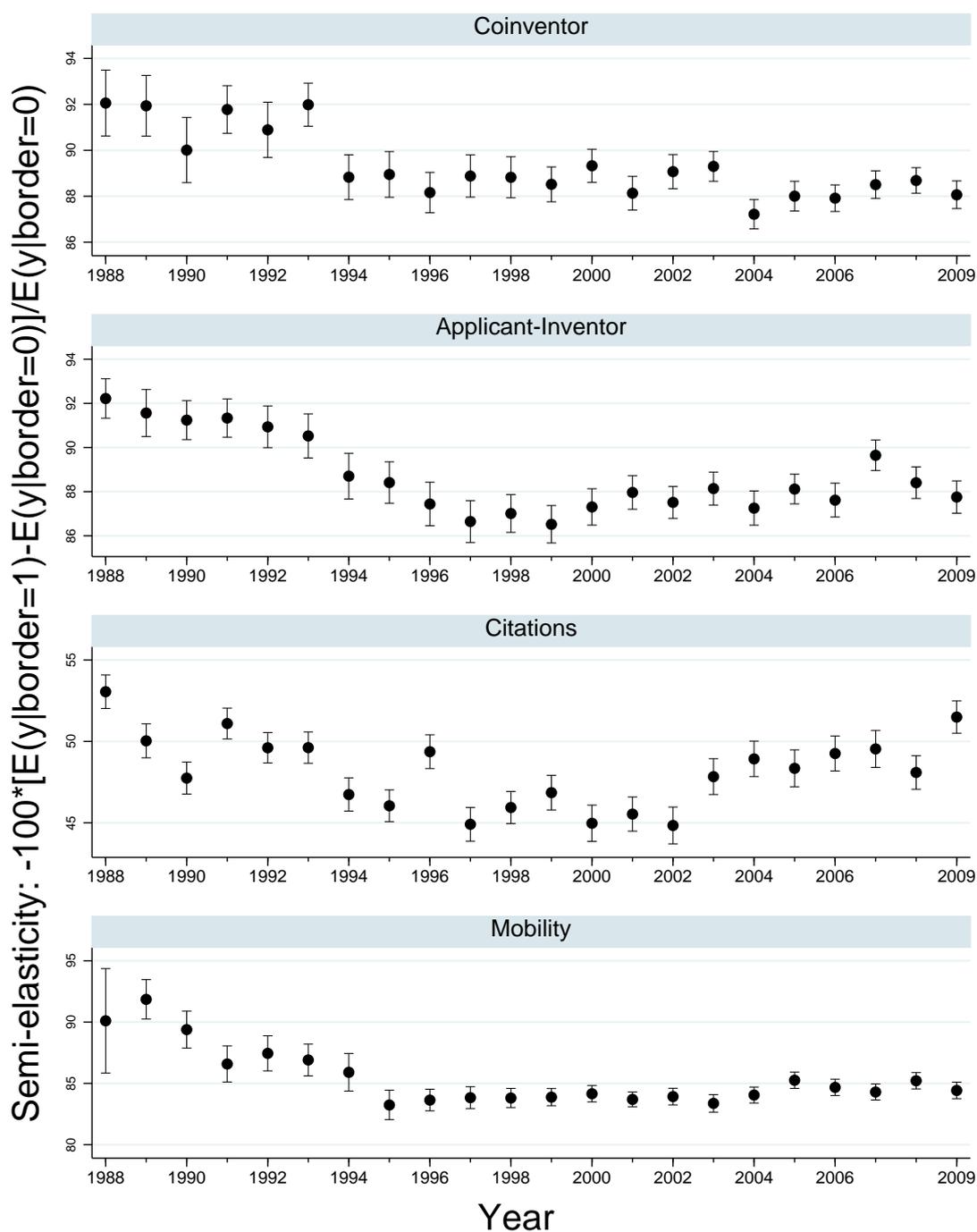
“Full”: See Eq 2;

“Gravity+border”: $X\beta^j = \beta_0^j + \beta_1^j border + \beta_2^j distance + \beta_6^j size_{m,t} + \beta_7^j size_{n,t} + \sum_{t=2}^T \theta_t^j year_t + \sum_{t=2}^T \delta_t^j distance * year_t$;

“Gravity”: $X\beta^j = \beta_0^j + \beta_2^j distance + \beta_6^j size_{m,t} + \beta_7^j size_{n,t} + \sum_{t=2}^T \theta_t^j year_t + \sum_{t=2}^T \delta_t^j distance * year_t$;

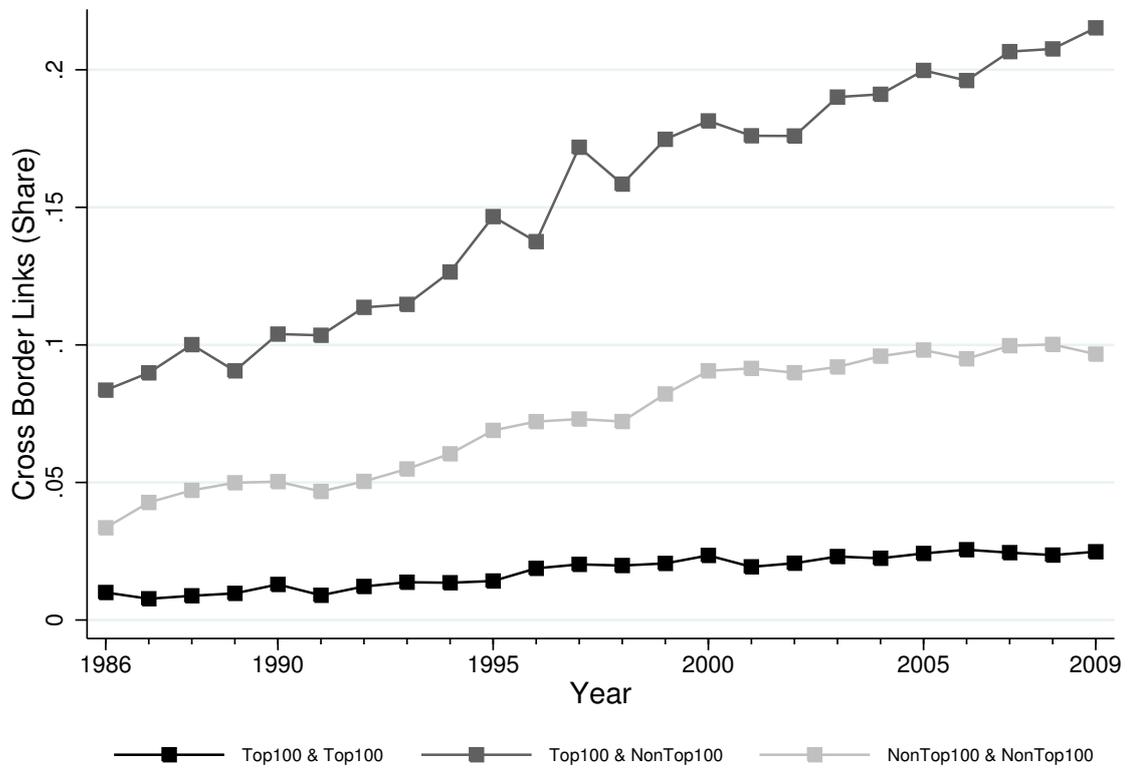
“Only distance”: $X\beta^j = \beta_0^j + \beta_2^j distance + \sum_{t=2}^T \theta_t^j year_t + \sum_{t=2}^T \delta_t^j distance * year_t$.

Figure 7



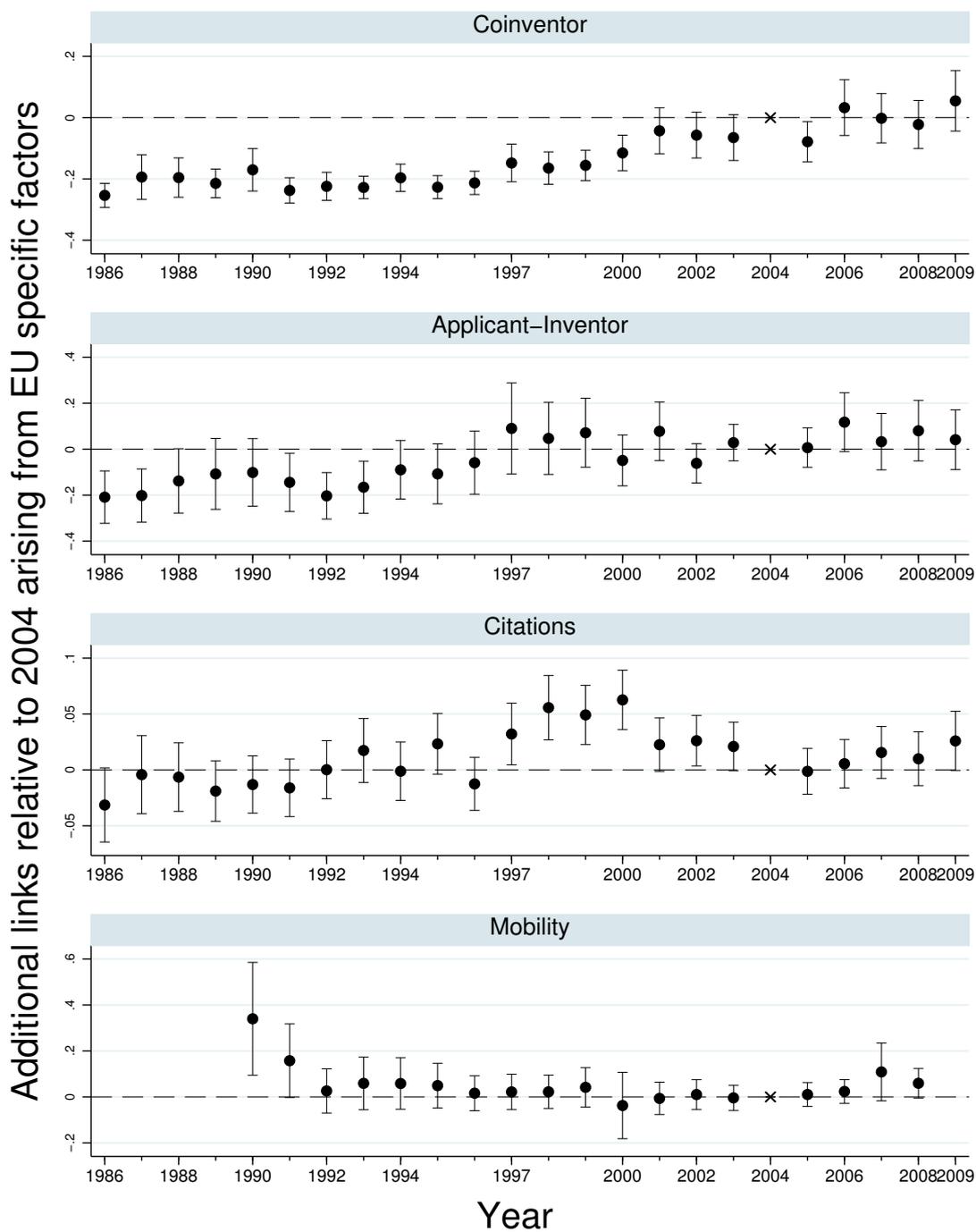
Notes: Estimates derive from four separate ZINB-gravity models for the count of links between NUTS3 regions. In the graph we report the yearly semi-elasticity and the corresponding 95% confidence interval constructed estimating standard errors through the Delta method. Year estimates of the country-border effect are obtained including interaction terms between the cross-border dummy and year dummies. We report in Table B.4 in Appendix B regression estimates of parameters in Eq. 4 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 $\pi_{2008} = (Y'' - Y')/Y' = (0.0688 - 0.6085)/0.6085 = -0.8869 = -88.7\%$.

Figure 8



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.

Figure 9



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 B.5 in Appendix B regression estimates of parameters in Eq. 6 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.

Appendix A. The ZINB model for the count of links

Zero-inflated models allow zeros to be generated by two distinct processes and are generally used when data exhibits “excess zeros” (Cameron and Trivedi, 1998), that is larger number of zero observations than what expected with the Poisson distribution. The ZINB model supplements a count density, \tilde{P} , with a binary zero generating process ψ . This allows a zero count to be produced in two ways, either as an outcome of the zero generating process with probability ψ , or as an outcome of the count process \tilde{P} provided the zero generating process did not produce a zero ($\psi_i = 1$).

The density distribution for the pair count y_i is then given by

$$P(y_i) = (1 - \psi_i) * \tilde{P}(y_i), \quad (\text{A.1})$$

where the zero generating process ψ_i is parameterized as a logistic function of the regressors in Z_i , with parameter vector β^0 :

$$\psi_i = \frac{\exp(Z_i \beta^0)}{1 + \exp(Z_i \beta^0)}. \quad (\text{A.2})$$

The count process $\tilde{P}(y_i)$ is modeled as Negative Binomial of the second kind (NB2):

$$\tilde{P}(y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1) + \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i}, \quad (\text{A.3})$$

where the conditional mean μ_i is parameterized as an exponential function of the linear index $X\beta^1$ ($\mu_i = \exp(X_i \beta^1)$), and $\alpha \geq 0$ is the overdispersion parameter.

In our estimation procedure we assume $X_i = Z_i$ because there is no reason to expect some variables would be relevant only in one of the two processes. However, individual regressors can impact the y_i estimator differently through the two distinct processes and their separate parameter vectors, β^0 and β^1 . Thus, drawing together Eqs. A.1, A.2, and A.3, and setting $X_i = Z_i$ our model for the expected count is what we report in Eq. 1.

Appendix B. Tables of estimates

Table B.3: Distance effect — Regression estimates

	Coinventor		Applicant-Inventor		Citations		Mobility	
	$y (\geq 0)$	ZI	$y (\geq 0)$	ZI	$y (\geq 0)$	ZI	$y (\geq 0)$	ZI
$y \equiv N_{(m,n)}$								
area (<i>base = nonEU ↔ nonEU</i>)								
area ₂ (<i>EU ↔ EU</i>)	-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 (<i>EU ↔ nonEU</i>)	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	6,415,200	6,415,200	6,415,200	6,403,320	
nr. of region pairs	146,070	291,600	291,600	291,600	291,600	291,600	291,060	
nr. of regions	540	540	540	540	540	540	540	
ln(α)	1.9004**	1.8684**	1.8684**	1.8684**	1.3393**	1.3393**	0.1199*	
Vuong test $\sim N(0, 1)$	125.53**	157.40**	157.40**	157.40**	322.19**	322.19**	132.19**	
l-ratio test $\sim \chi^2$	6,758,968.3**	6,024,960.4**	6,024,960.4**	6,024,960.4**	8,624,418.7**	8,624,418.7**	197,858.4**	

Notes: * significant at 5%; ** significant at 1%. 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 Figure 5). 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. 2. 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 versus the ZIP (Long and Freese, 2006).

Table B.4: Border effect — Regression estimates

	Coinventor		Applicant-Inventor		Citations		Mobility	
	$y_i \geq 0$	ZI	$y_i \geq 0$	ZI	$y_i \geq 0$	ZI	$y_i \geq 0$	ZI
$y_i \equiv N_{(m,n)}$								
<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	
nr. of region pairs	146,070		291,600		291,600		291,060	
nr. 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**	

Notes: * significant at 5%; ** significant at 1%. 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 Figure 7). 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. 4. 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 versus the ZIP (Long and Freese, 2006).

Table B.5: Border effect EU vs. nonEU — Regression estimates

$y_i \equiv N_{(m,n)}$	Coinventor		Applicant-Inventor		Citations		Mobility	
	$y (\geq 0)$	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	
nr. of region pairs	160,321		131,310		302,775		6,685	
nr. 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**	

Notes: * significant at 5%; ** significant at 1%. 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 Figure 9). 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. 6. 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 versus the ZIP (Long and Freese, 2006).

Captions of Figures

Figure 1: (A) Co-inventor network probability density function $P_t(d_{m,n})$. (B) Size distribution of total patent counts S_n aggregated over three consecutive 8-year time periods. (C) Gravity model distance elasticity ρ for 72 size-size, distance, and time-period subgroups.

Figure 2: Pie charts for 18 size and distance subsets indicating the fraction and number of links between regions in period T_2 (1994-2001). Percent changes ($\% \Delta$) in f_{EU-EU} and ρ , over the consecutive periods T_1 to T_2 , are listed within each subpanel.

Figure 3: Pie charts for 18 size and distance subsets indicating the fraction and number of links between regions in period T_3 (2002-2009). Percent changes ($\% \Delta$) in f_{EU-EU} and ρ , over the consecutive periods T_2 to T_3 , are listed within each subpanel.

Figure 4: Evolution of average distance of R&D collaborations

Figure 5: Elasticity of distance over years

Figure 6: Elasticity of distance with different models — Applicant-Inventor

Figure 7: Border effect over years

Figure 8: Flows between top regions and small regions — Evolution of the cross-border shares

Figure 9: Evolution of European integration



INSTITUTE FOR ADVANCED STUDIES LUCCA

2014 © IMT Institute for Advanced Studies, Lucca
Piazza San ponziano 6, 5100 Lucca, Italy. www.imtlucca.it