Research Policy 45 (2016) 177–194

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Do inventors talk to strangers? On proximity and collaborative knowledge creation

Riccardo Crescenzi[a,b,*], Max Nathan[b,c], Andrés Rodríguez-Pose[a,b]

[a] Department of Geography & Environment, London School of Economics, United Kingdom
[b] SERC, London School of Economics, United Kingdom
[c] Birmingham Business School, University of Birmingham, United Kingdom

ARTICLE INFO

Article history:
Received 26 October 2013
Received in revised form 15 July 2015
Accepted 16 July 2015
Available online 22 October 2015

JEL classification:
O31
O33
R11
R23

Keywords:
Innovation
Patents
Proximities
Regions
Knowledge spillovers
Collaboration
Ethnicity

ABSTRACT

This paper examines the characteristics of the collaborations between inventors in the United Kingdom (UK) by looking at what types of proximities – geographic, organisational, cognitive, social, and cultural–ethnic – between inventors are prevalent in partnerships that ultimately lead to technological progress. Using a new panel of UK inventors this paper provides an analysis of associations between these 'proximities' and co-patenting. The results show that while collaboration within firms, research centres and universities remains crucial, external networks of inventors are key feature of innovation teams. The analysis shows that external networks are highly dependent on previous social connections, but are generally unconstrained by cultural or cognitive factors. Geographical proximity is also weakly linked with external networks. Our results suggest that innovation policies should, rather than focus on spatial clustering, facilitate the formation of open and diverse networks of inventors.

1. Introduction

The age of the lone researcher, of the quixotic 'basement tinkerer' (Rabinow, 1976), or of the 'garage inventor' (Seaborn, 1979) is receding. The romantic notion that a new Nikola Tesla will emerge from the lab with the next AC motor (or a death ray) increasingly belongs to a bygone era. While in the late 1970s around 75% of EPO patent applications in the United Kingdom (UK) were filed by individual inventors, nowadays that figure is below 15%. More than 80% of all patents are registered to more than one inventor, suggesting that collaboration in research and innovation has become the norm. Increasingly larger teams are formed within firms or research centres. Complex networks of researchers involving different firms, often in collaboration with universities, public agencies, and research centres drive the world of invention in the early 21st century. As Seaborn (1979:88) puts it, "big science [has] eclipsed the garage inventor [. . .] Edison has been superseded by a team of white-coated theoretical physicists".

While the trend towards the formation of ever-larger research teams and inventor networks has been well documented, we know much less about the features of these teams. What are the characteristics of the inventors that decide to work in a team? Is collaborative research produced by inventors that talk to colleagues, or to strangers? These are the questions at the heart of this paper, which aims to shed new light on the patterns of collaboration observed among UK inventors.

In the paper, collaboration by inventors is captured by means of co-patenting over the past three decades. We explore the individual circumstances that members of a co-patenting team may share and which, according to the literature on innovation and proximity (Boschma, 2005; Boschma and Frenken, 2009; Torre and Rallet, 2005), can be grouped into different types of proximities:

* Corresponding author at: Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Tel.: +44 2079556720.
E-mail address: r.crescenzi@lse.ac.uk (R. Crescenzi).

http://dx.doi.org/10.1016/j.respol.2015.07.003

(a) geographic (the physical distance between inventors); (b) organisational (whether the inventors share the same organisational context, such as the same firm, university or research centre); (c) social (whether inventors have co-invented in the past or share other co-inventors); (d) cultural–ethnic (whether co-patenting inventors share the same national, cultural, and/or ethnic background); and (e) cognitive (the distance between the technology fields of the co-patenting inventors) proximity.

To explore these linkages, we develop a new empirical strategy, which builds on ideas developed in seminal papers by Jaffe et al. (1993), Singh (2005), and Agrawal et al. (2008). We use this to build a new panel of EPO patents microdata from the KITES-PATSTAT resource and we then analyse the incidence of the different proximities considered in co-patenting teams, controlling for a broad set of observable and time-invariant unobservable characteristics (the former through a vector of individual, organisational, and environmental factors; the latter by means of fixed effects). The empirics also employ the innovative ONOMAP name classification system to ascribe inventor ethnicity – and thus, ethnic/cultural proximity. We use social network analysis to identify the position of each inventor in pre-existing collaboration networks.

The paper represents – to the best of our knowledge – the first empirical work assessing the incidence of such a large set of proximities in collaboration patterns, and contributes to the existing literature in several different ways. It finds that, for inventors as a whole, organisational proximity is a key feature of co-patenting teams together with cultural/ethnic diversity. Conversely, geographical proximity is linked to co-patenting in combination with other proximities.

For 'multiple patent' inventors, we find that organisational proximity remains highly relevant, while cultural/ethnic factors are not relevant. Social network and cognitive proximities are more important characteristics of the teams formed by these inventors. The analysis also confirms the incidence of 'unconstrained' (i.e. free from ethnic factors) social proximity and social networks in collaborative activity. For this category of inventors the importance of geographical proximity only emerges as well in interaction with other proximities.

Our results have important implications for the analysis of innovation dynamics and, possibly, for the targeting of innovation policies. The empirical analysis suggests that knowledge and key competences for innovation processes are combined (and recombined) within the organisational boundaries of firms, research centres and universities. These are the key units of analysis of innovation dynamics. Assets internal to the individual organisational unit are complemented by processes of external search that take place within existing social networks that 'bridge' various organisations. The formation of these networks remains largely unconstrained by cultural or cognitive proximity considerations in order to ensure variety and avoid lock-in situations. In this picture the direct contribution of geographical proximity and spatially mediated processes remains limited: it only emerges in the form of hyper-geographical proximity (inventors in the same organisation are likely to be co-localised in the same premises) and as a reinforcement (or facilitator) for network-based interactions. These results suggest that innovation policies should place less emphasis on spatial clustering and localised collaborations and focus more on the capabilities internal to each firm and its ability to access 'unconstrained', open and diverse external networks.

The paper is structured as follows. Section 2 reviews the existing literature on collaborative working among inventors and outlines a conceptual framework for the analysis of the drivers of collaborations among inventors. Section 3 introduces our data and gives some stylised facts. Section 4 sets out our empirical strategy and model. The empirical results with a number of robustness checks are discussed in Section 5. Section 6 concludes.

2. Collaborative working among inventors and proximity relationships

Collaborative invention efforts have been on the rise for quite some time. The number of co-authored scientific publications, both international (Glänzel, 2001; Glänzel and Schubert, 2005) and within specific countries has been increasing in recent decades. In the US, for instance, Adams et al. (2005) find a 50% rise in the average number of authors per academic paper during the period 1981–1999. Similar shifts can be seen in patenting activity. In the UK 'co-invented patents' rose from around 100 in 1978 (24.2% of all patents) to over 3300 in 2007 (66.6% of all patents). Over the period as a whole, 57.3% of patents had more than one inventor. Co-patenting also increased across all major technology fields and the share of inventors working alone fell dramatically. During the period of analysis the mean size of patenting teams rose from under two to over four.

These trends are the result of the evolution of both public policy and corporate strategies. National governments have sought to develop the formation of innovation 'ecosystems'. This, in combination with the internationalisation of firms' activities and the tendency of multinational firms to couple with local partners in knowledge-intensive activities (Cantwell, 2005; Yeung, 2009), has encouraged the formation of research teams which expand well beyond the firm or their research centre. University-industry joint ventures and the growth of Triple Helix relationships involving firms, universities, and government (D'Este and Iammarino, 2010; Leydesdorff and Etzkowitz, 1998) have gradually become the norm.

These 'global' high-level trends have taken place in a context of in-depth change in the individual-level incentives for collaboration. The increasing sophistication of 'frontier' science reinforces the returns to specialisation and promotes collaboration as a means to handle a growing 'burden of knowledge' (Agrawal et al., 2014; Jones, 2009), as well as a form of 'risk sharing' for high-risk/high-gain projects. At the same time, the need to gain access to both highly complex research infrastructure and larger funding pools via collaborative grants (Freeman, 2014) strengthens the incentives for the formation of larger teams of researchers. Finally, research projects require increasingly more diverse sets of complementary skills and competences to be successful (Agrawal et al., 2008).

Scientists and researchers have responded to these changes by making collaborative research the norm both in the United States (Jones et al., 2008) and Europe (Brusoni et al., 2007; Giuri and Mariani, 2012). However, collaboration comes at a cost for all parties involved: the search for the best possible collaborator(s)/team members – whether this decision is taken by the inventors themselves or by managers within the boundaries of a firm – is an expensive process in terms of time and resources. Agents face benefits and costs when considering potential connection/collaboration (see Jackson (2006) for a recent review) and a number of studies have drawn on principal–agent theory to look at contract formation and partner selection at the individual level (Ackerberg and Botticini, 2002; Sedikes et al., 1999).

Moreover, collaboration is by definition a social act and, in addition to economic considerations, it is shaped by personal preferences and circumstances (Giuri and Mariani, 2012), an individual's position in an organisation, the nature and capacity of those organisations, the type of work they do, and a range of external circumstances – such as legal and funding frameworks, industry and policy trends.

Various 'proximities' assist in the creation of innovation networks by reducing team formation costs and overcoming coordination and control problems. Boschma (2005) distinguishes five types of proximity – cognitive, organisational, social, institutional, and geographic. He suggests first, that these factors may operate as substitutes or complements; and second, that proximities are not always beneficial (Boschma, 2005). Excessive proximity may cause 'lock-in' – a lack of openness and flexibility that inhibits innovation by leading to the formation of teams where redundant knowledge prevails.

Cognitive or 'technological' proximity allows agents to communicate in the same research field (Seely Brown and Duguid, 2002). Organisational and social proximities lower transaction costs via (respectively) contracts and social relationships (Kaiser et al., 2011). However, hierarchical organisational structures or supply chain relationships close firms off from the technological opportunities that 'open innovation' models provide (Von Hippel, 2005). And social networks based on 'strong ties' may be less effective than larger networks of 'weak ties', if they do not admit new members or new thinking (Granovetter, 1973). Geographical proximity reduces the cost of knowledge sharing, transmission and monitoring by enabling face-to-face contacts between the agents involved (Storper and Venables, 2004). However, the spatial clustering of innovative activities, if not constantly renewed by inflows and outflows of new resources by means of mobility, can easily lead to cognitive lock-in (Crescenzi et al., 2007). In addition, an emerging body of newer literature has shown that co-ethnic and diasporic networks (see Docquier and Rapoport (2012) for a recent survey) increase trust and lower transactions costs, assisting in the generation and diffusion of collaborative ideas (Kapur and McHale, 2005; Kerr and Lincoln, 2010; Saxenian and Sabel, 2008). Just as with other proximities, however, the capacity of different co-ethnic groups may vary substantially and is potentially limited by discrimination.

While there is consensus in economic geography, innovation studies and management on the importance of multiple conditions for knowledge-intensive collaborations, the empirical challenge remains how to operationalise, disentangle and 'weight' the relative importance of these – often coexisting and overlapping – proximities.

Until recently the general view was that geographical proximity trumped all others. Jaffe et al. (1993) were the first to use patent citations as a way to provide a 'paper trail' for knowledge flows. They showed that geographic proximity facilitated local knowledge exchange. More recent analyses have tended to re-state the relevance of physical proximity. Lobo and Strumsky (2008) found that spatial agglomeration of inventors in US MSAs is more important than the density of inventor connections (which are negatively correlated to patenting). Similarly, Fleming et al. (2007) have suggested that spatial proximity is a crucial enabling factor for the effective transmission of knowledge flows.

Other research has however suggested that the role of geographical factors may have been overstated (Thompson and Fox-Kean, 2005). In particular, the importance of *social networks* as a catalyst for inventor collaboration has come forward in recent years (Singh, 2005; Brusoni et al., 2007; Giuri and Mariani, 2012). From a social network perspective, Evans et al. (2011) have uncovered evidence that homophily explains co-authorship of scientific papers, but that institutional and geographic proximity play bigger roles in this respect. Cassi and Plunket (2010) suggest that spatial and social proximity are highly complementary, even if individual partners are in different types of organisations. Singh (2005) confirmed that the effect of geography and firm boundaries on knowledge flows diminishes substantially once interpersonal networks have been accounted for. Prior social relationships between investors also explain current citation patterns, even after spatial proximity is altered by mobility decisions (Agrawal et al., 2006). Overall, geographic and social proximity are considered to operate as substitutes (Agrawal et al., 2008).[1]

But, despite these efforts, our understanding of which are the most important types of proximities for the formation of collaborative inventor networks is still rather poor (Torre and Rallet, 2005). A number of under-explored areas remain. First, we still know relatively little about *individual innovative agents* – most studies aggregate outcomes to firms, cities and regions. Second, while many studies explore knowledge spillovers, or consequences of collaboration, much less work has been done on the intrinsic features of these collaborations (Boschma and Frenken, 2009). Third, due to data constraints, there are few studies that have been able to explore time periods above a decade. Fourth, the role of *cultural and ethnic proximity* has been particularly neglected in quantitative analysis outside the US (see Kerr (2013) for a review of the US literature). As far as we are aware there is only one relevant study for Europe (Nathan (2014), for UK 'minority ethnic inventors').

These gaps raise three important research questions:

(1) What forms of proximity are associated with the incidence of collaborative knowledge creation at the individual level?
(2) What is the interaction between different proximities?
(3) How has the salience of these proximities changed over time?

Providing answers to these questions is of crucial importance to the understanding of the process of innovation and its diffusion: they shed new light on the physical or virtual milieu of the innovation process and on the relevant unit(s) of analysis for its understanding. If innovative collaborations are largely shaped by localised and spatially mediated processes, then cities (and their functional hinterland) are key units of analysis and possible targets for innovation policies. If, by contrast, the process of searching and matching of scientific competences for innovative projects takes place within the organisational boundaries of firms and research centres, the focus should move to microanalysis. Finally, if cultural/ethnic, social and cognitive channels are essential for the formation of innovation networks, then the emphasis should be put on how both localities and firms can be connected and re-connected beyond physical contiguity.

This paper aims to answer these questions by looking at what inventors' characteristics and relationships to each other prevail in collaborative knowledge creation (specifically co-patenting) teams. To do this, we need to disentangle relational factors from each other, and from other characteristics of collaboration teams (i.e. 'individual', 'institutional' and 'environmental' factors).

3. Data and stylised facts

Our dataset contains European Patent Office (EPO) patent microdata from the PATSTAT database, modified by the KITES team at Universita' Bocconi (hence 'KITES-PATSTAT'). The raw data runs from 1978 to 2010, comprising 116,325 patents with at least one UK-resident inventor. 173,180 inventors are associated with these patents, of whom 133,610 are UK residents. Unlike standard patent data, KITES-PATSTAT has been cleaned to allow robust identification of individual inventors, their spatial location and patenting histories, as well as the usual array of patent and applicant-level characteristics [see Lissoni et al. (2006) for details of the cleaning process].[2]

Patent data have a number of advantages for our purposes. They provide rich data over a long time period, as well as detailed

[1] Micro-level analyses have reached similar conclusions on the simultaneous interplay of a variety of drivers for knowledge exchange and cooperative innovation projects. See for example: Singh (2005), Agrawal et al. (2008), Paier and Scherngell (2008), Lychagin et al. (2010), Griffith et al. (2011), and D'Este et al. (2013).

[2] KITES-PATSTAT also provides extensive applicant-level information, particularly for corporate applicants: names/address details are matched to company information from Dun and Bradstreet.

information on individual inventors and their past/present collaborators, the type of research they work in (via detailed 'technology field' codes), and the organisations they work for (typically the patent applicant) (OECD, 2009). On the other hand, the data have two inherent limitations. Patents measure *invention* rather than *innovation*; and they tend to only observe some inventions and inventors (for instance, some members of a research team may be left off the patent application). Other limitations, such as patenting's manufacturing focus and vulnerability to policy shocks, are dealt with using appropriate industry controls and time trends, as discussed below.

We make some basic edits to the data to make it fit for purpose. First, there is typically a lag between the application and the granting of a patent. This means that in a panel of patents, missing values appear in final periods. Following Hall et al. (2001), we truncate the dataset by three years to end in 2007. Second, we geo-locate UK-resident inventors in UK Travel to Work Areas (TTWAs). TTWAs are designed to represent functional labour markets and offer a good proxy for the local spatial economy.

Third, EPO patent data gathers patent applications through EU countries' national patent offices, through European-wide applications to the EPO ('Euro-direct'), and international applications that have reached the European examination stage ('Euro-PCT') (OECD, 2009). As such, our dataset may not include all PCT patent applications, which cover international applications to multiple patent offices. Since PCT applications are increasingly the favoured route for inventors seeking to access international markets, there is a risk of selecting out some collaborative activity. However, the use of PCT applications has increased substantially since the early 2000s, meaning that this may affect only part of our sample. By truncating the end of the time series, we minimise the PCT selection issue.

We use the ONOMAP system (Mateos et al., 2011) to observe likely inventor origin/ethnicity from individual inventors' name information, building on work by Nathan (2014) and key US studies by Kerr (2010) and Agrawal et al. (2008). ONOMAP is developed from a very large names database extracted from Electoral Registers and telephone directories, covering 500,000 forenames and a million surnames across 28 countries. Classifying individuals according to most likely 'cultural–ethnic–linguistic' characteristics, ONOMAP also provides information on ONS ethnic groups, geographical origin and major language. More details on ONOMAP are given in Appendix A.1.

3.1. Stylised facts

Collaborative research and invention have progressively increased their importance over time and across technology fields (Fig. 1). Over the whole period, 57.3% of patents were 'co-invented'. Co-invention is the norm: 15.9% of inventors only work alone (i.e. never co-invent); 4.7% sometimes co-invent; 79.4% only co-invent. Patents with five or fewer inventors comprise over 95% of the sample, of which over half are co-invented. Most co-invented patents have two or three inventors, with two being – by some margin – the mode (26.2% of patents, versus 14.7% of patents with three inventors). Three distinct phases can be identified within the sample period: from the late 1970s to the late 1980s; the 1990s, with a peak in co-inventing in 2000; and then a plateau period, with a slight decline at the end of the panel (probably reflecting fewer granted patents).

Co-invention trends vary substantially across patent fields. Fig. 2 shows the trends in co-invented patents across seven aggregated technology fields (using the OST reclassification).[3] At the start of

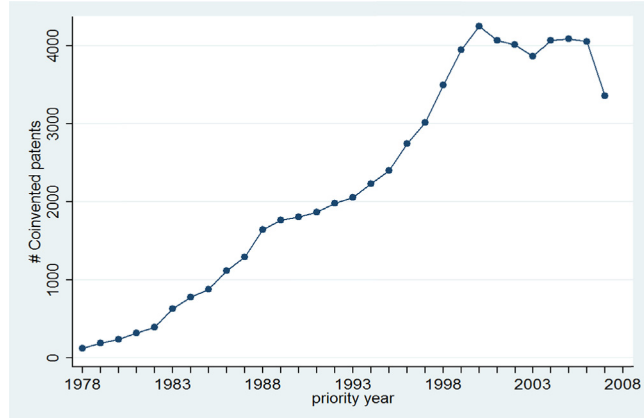

Fig. 1. Co-invented patents, 1978–2007.

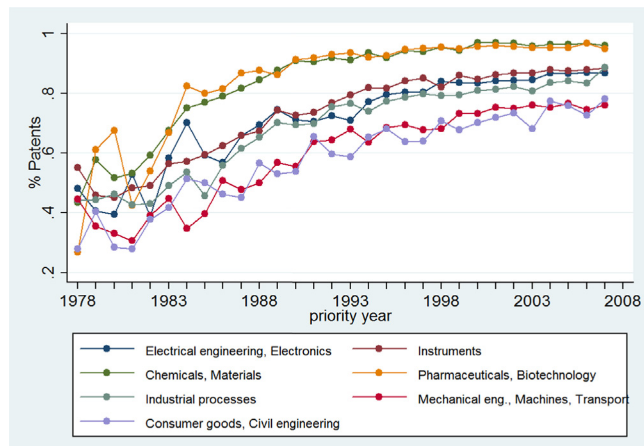

Fig. 2. Co-invented patents by technology field, 1978–2007.

the sample period, shares of co-inventing were low (from 0.14 in consumer goods to 0.22 in electrical engineering). By the end of the period shares were higher, although the variation across sectors increased: from 0.49 in consumer goods to 0.83 in chemicals and materials. In six out of seven sectors, co-invention shifted from minority to majority type.

Inventors' behaviour has also changed over time. Fig. 3 shows that these aggregates hide large changes within the sample. Counts and shares of 'only co-inventing inventors' rose substantively; in contrast, while counts of 'only solo' inventors rose slightly, their relative shares underwent an extensive decline.

Finally, we look at inventor team composition. Fig. 4 shows the trend in average team size. The trend line is spikier than the co-invention trend, but the general shape is the same. We can see that the average patent in 1978 had 1.73 inventors; by 2007 this had risen to just over four inventors.

Taken together, these stylised facts suggest a substantial rise in co-patenting across technology fields between the late 1970s and the late 2000s, involving significant changes in inventor behaviour. In terms of the institutional and organisational context in which the patenting process takes place, applicant-level information reveals that private firms and their research labs have been the leading actors behind patenting (with 55% of the total patents in the sample) while universities, other public research centres and NGOs

[3] This classification is used for illustrative purposes only. In the regression analysis the 30-fold typology is used, alongside more detailed typologies of 121 IPC three-digit sub-classes and 1851 six-digit main classes to generate technology field fixed effects.

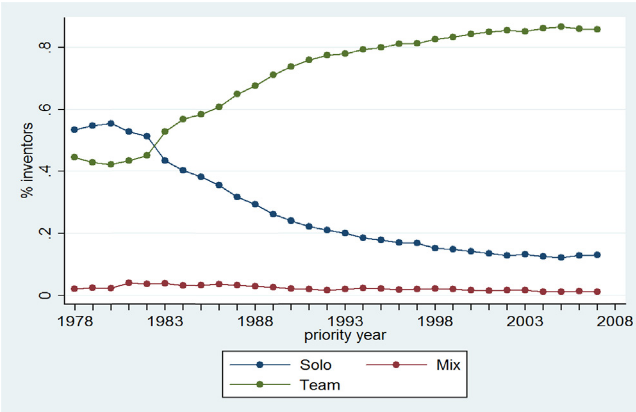

Fig. 3. Inventor behaviour, 1978–2007.

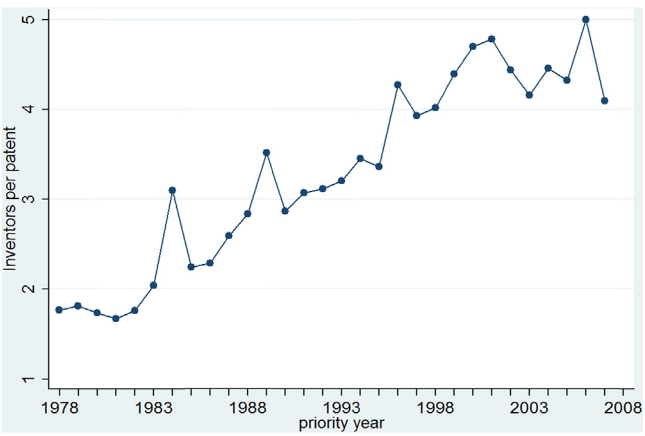

Fig. 4. Inventor team size, 1978–2007.

account for 4.2% of the total patents, and individual applicant for 6.7% (see Table B.1 in the Appendix).

4. Empirical strategy

4.1. Counterfactual build

The paper aims to compare the characteristics of inventors working together with those who do not, to identify what factors are systematically correlated with the incidence of co-invention. As we are interested in collaborations between inventors, we make the inventor pair the unit of observation.[4]

We therefore need to consider both a plausible set of *possible pairs* – inventors who *might have* worked together – and the set of *actual pairs* (those who did collaborate). To do this, we build a 'synthetic counterfactual' consisting of a feasible set of potential collaborators. We then use a case–control strategy, in which we disentangle the characteristics of actual pairs from the feasible possible pairs that might have existed. That is, we uncover the factors associated with collaboration, and control for its underlying incidence. To model inventor collaboration (as opposed to citation patterns) we make some changes to the case–control methods deployed by Jaffe et al. (1993), Thompson and Fox-Kean (2005), Agrawal et al. (2006) and others.

The feasible set of potential collaborators is constructed as follows. We first specify a feasibility framework. For example, it is highly unlikely that inventors active in the first year of our data (1978) will be collaborating with those active in the last year (2007). Moreover, collaborations across entirely unrelated technology sectors are implausible.[5] We thus assume that the set of potential collaborators is broadly delineated by time period (year) and some high-level IPC technology field (IPC1, 2 or 3). We then use our main variables of interest – proximities variables – as binary 'treatments' and look at the time-by-IPC field combinations that create broadly 'balanced' 'treatment' and 'control' groups among the set of actual inventors. This suggests year-by-IPC3 field is the most feasible combination.[6]

To make the data tractable, we randomly sample 5% of patents. This allows to fuzzily observe the real-world incidence of actual and feasible collaborations.[7] Given the boundaries of the feasible collaborator set, this implies stratifying the sample by year, 121 three-digit technology fields, and inventor team size.[8]

For the actual inventor pairs in each year-by-IPC3 cell in this 5% sample, we generate the set of feasible possible pairs *in that* cell. That is, each set of feasible collaborators is directly defined from the relevant set of actual collaborators. For each year, we append the resulting 'cell-pairs' to form a cross-section. We then append the resulting yearly cross-sections to create an unbalanced panel.

Each inventor in a pair is separately coded by address and by patent applicant, permitting the specification of controls and fixed effects on each side of the pair. We build a 16-year panel for the years 1992–2007 inclusive, reserving the period 1978–91 to provide historic information on inventors' patenting activity (see Section 5). This gives us a panel of 190,313 observations, covering 187,997 actual and possible inventor pairs, of whom 3857 (2.05%) are actual pairs.

We also look at the subset of 'multiple patent' inventors (who invent more than once in the observed time span). In the raw data there are 17,764 of these individuals, active on 62,339 patents. Multiple patent inventors are worth greater scrutiny for two reasons. First, the vast majority of inventors only patent once, so the 'multiple' group is an unusual minority. The inventor lifecycle literature suggests multiple patent inventors are likely to be highly productive individuals, often in senior scientific positions (Azoulay et al., 2007; Lee and Bozeman, 2005). It is therefore interesting to see if these inventors' collaborations have different features when compared to the pooled sample. Second, looking at multiple patent inventors lets us look at a broader set of proximities.

[4] To assess why inventor-level analysis is appropriate, consider an opposite scenario: working at the patent level. For example, we could estimate a model where the dependent variable would be a dummy taking the value 1 for a co-invented patent; independent variables would cover characteristics for the inventor/set of inventors involved. This has some desirable characteristics – not least, allowing to examine all patents in our sample – but discards crucial inventor-level information. An additional important challenge for patents with more than one inventor is where to locate the patent in space.

[5] This is also computationally intensive – for example, there are 173,180 inventors in our full sample 1978–2007, which makes for 1.499^{10} possible pairs.

[6] In principle possible pairs could be also restricted to those working in very detailed technology fields (say IPC6 and above). However, diagnostics suggest increasing incidence of individuals patenting *across* technology fields. More importantly, we want to explore whether this 'cognitive distance' affects levels of co-inventing, reserving the variation for regressions rather than build into the matching process.

[7] It is possible that some inventor pairs who did not patent in our sub-sample – who we designate 'feasible' pairs – did in fact patent in the rest of the dataset, forming 'actual' pairs. If this process was non-random it would bias our estimates. However and given our sampling procedure, as there is no reason to expect this to be anything other than random, we treat this issue as generating noise only. In principle, an alternative approach would have been to build a complete counterfactual from all patents and then sample, but, as we point out in footnote 5, this is not computationally feasible.

[8] Given the size of the original population, we sample without replacement. We seed the sample so that regression results are reproducible. We relax this in robustness checks, to test whether sample construction affects our findings.

Specifically, social and cognitive proximity measures need to be based on historic behaviour in order to avoid a mechanical link between dependent and independent variables in the model. In turn, this requires that we observe more than one patenting event. Including these proximities thus involves restricting our sample to the set of patents with (a) multiple patent inventors, and (b) only multiple patent inventors: there are 27,315 of these patents, involving 11,754 individuals. However, the smaller n allows a higher sampling rate. We sample 25% of these patents and stratify as before, then build a counterfactual group using the same approach as for the pooled sample. This results in a second unbalanced panel of about 40,638 observations.

4.2. Model specification

In order to explore the characteristics of co-inventor pairs, we look at links between collaborative activity in an inventor pair *ij* and the relational characteristics of *ij*, while controlling for individual, institutional and environmental factors affecting each inventor in that pair. We therefore estimate the following empirical model, for inventor pair *ij* in applicant *o*, area *a*, year group *t* and technology field *f*:

$$Y_{ijoatf} = a + \mathbf{PROX}\boldsymbol{b}_{it,jt} + \mathbf{CTRLS}\boldsymbol{c}_{(ioat,joat)f} + e_{ijoatf} \quad (1)$$

Here, Y is either a collaboration dummy (taking the value 1, if *ij* is an actual pair), or a continuous variable giving the count of collaborations for *ij*. The dummy variable allows us to study the characteristics of individual collaborations, while we use the inventor pair activity count in order to capture the features of repeated collaborations. **CTRLS** are either vectors of observables, or fixed effects (see below).

Our variables of interest are given by **PROX**, a vector of proximities covering spatial, cultural–ethnic and organisational proximities (for the pooled sample), plus cognitive and social proximities (for multiple patent inventors). The estimated coefficients (*b-hats*) of each PROX variable indicate the salience of a particular proximity in actual rather than possible inventor pairs, after controlling for individual, institutional and macro-environmental factors. Note that these are associations, not causal links.

Spatial proximity is calculated as the inverse of the linear distance between TTWA centroids where each inventor is located; for inventors in different countries this is set as zero. For robustness checking, we also construct a more sophisticated inverse linear distance function, with a threshold function to capture knowledge spillovers decay. We also build a simple dummy taking the value 1 if a pair is resident in the same TTWA.

Organisational proximity is captured by means of a dummy, which takes the value 1 if *i* and *j* share the same applicant and 0 otherwise. The computation of this variable is only problematic where the number of inventors exceeds the number of applicants in the patent. In this case it is more difficulty to establish a clear inventor-applicant association. As a consequence, where there is more than one applicant to which an inventor could be assigned (12.77% of the patents in the sample), we probabilistically assign inventors and applicants. We do this in two ways: (a) we use inventors' patenting history (whenever available) to assign inventors to their modal applicant if there is an ambiguous patent, then drop duplicates from the dataset (this works for 40% of ambiguous cases); (b) in those cases where a patenting history is unavailable, we randomly assign the inventor to one applicant or the other and run a robustness check, repeating the re-assignment procedure to confirm that this does not change the results. We further compute an additional proxy for 'scaled' organisational proximity. This variable takes the value of 0 (as above), if the applicant is not the same; 1, if the applicant is the same, but there is more than one applicant on the patent; 2, if the applicant is the same and there is only one applicant on the patent. The intuition here is that in a given collaboration, organisational proximity between two people in the same company is stronger than their collaboration if another company is also involved.

Ethnic proximity is developed using ONOMAP (see Appendix A.1). Our preferred measure is a dummy, which takes the value 1 if *i* and *j* share the same 'cultural–ethnic–linguistic' subgroup, of which there are 67 in the ONOMAP classification. In robustness checks we use dummies based on nine ethnicity groups based on the official UK government typologies, 13 geographical origin categories (covering the UK, Ireland, European zones and other continents), and a number of major languages.

For multiple patent inventors, we are also able to observe cognitive and social proximities. Cognitive proximity is set up as a dummy taking the value 1 if *ij* have previously both patented in the same IPC technology field. Following Thompson and Fox-Kean (2005) and Singh (2005), our preferred measure uses each of 1581 6-digit IPC fields. We use a less detailed IPC3 version in robustness checks.

Social proximity is defined as the inverse social distance between *i* and *j*, based on whether they have co-invented in the past, have co-authors in common, or more indirect links to actual/possible partners. We assume that ties decay after five years. Following Singh (2005), for a given year, we then measure the number of 'steps' between inventors *i* and *j* based on their activity in the previous five-year period. This is the social distance between *i* and *j*. We then take the inverse distance to generate the social proximity between the two. For example, if *i* and *j* have co-invented together in the past, the number of steps between them is 0. If *i* and *j* have not collaborated directly, but have both worked with *k*, then there is one step between them; if *i* is connected to *j* through *k* and *l*, there are two steps; and so on. Respective degrees of social proximity are then 0, −1, and −2, through to minus infinity (no link).

4.2.1. 'Observables' versus fixed effects models

We first estimate the model with **PROX** alone, followed by the model with vectors of observable characteristics, then with a battery of fixed effects. The observables model is:

$$\begin{aligned} Y_{ijoatf} &= a + \mathbf{PROX}\boldsymbol{b}_{it,jt} + \mathbf{IND}\boldsymbol{c}_{it,jt} + \mathbf{INST}\boldsymbol{d}_{io,jo} \\ &\quad + \mathbf{ENV}\boldsymbol{e}_{(iat,jat)f} + e_{ijoatf} \end{aligned} \quad (2)$$

The vectors **IND**, **INST** and **ENV** respectively cover observable characteristics of inventors (both *i* and *j* in the pair), applicant organisations, and time, technology field and TTWA characteristics. More detail on the controls is given in Appendix A.2. This setup handles a number of individual, institutional and macro features of collaborations. However, it does not deal with unobservables, and thus our preferred model is the fixed effects specification below, which handles time-invariant factors:

$$Y_{ijoatf} = a + \mathbf{PROX}\boldsymbol{b}_{it,jt} + I_{ioa} + J_{joa} + T_t + TF_f + u_{ijoatf} \quad (3)$$

This specification does not control for selection issues or time-varying unobservables (see Section 4.4): so, as before, we interpret *bs* as associations, not causal effects.

Note also that a fixed effects specification separately controlling for individuals, organisations, areas, time and technology field would be unwieldy. To keep regressions tractable, therefore, we set up inventor-applicant-area 'spells' for each inventor *i* and *j* in a pair, following Abowd et al. (2002:48) and Andrews et al. (2006:49). These 'spell' variables take unique values for each inventor-applicant-TTWA combination in the panel. Since our variables of interest are at the pair level, this allows controlling for non-time-varying unobservables on each side of the pair, while keeping in pair-level characteristics. In addition, we fit year dummies and an IPC6 technology field 'grouping variable' to cover time trends and technology field-specific shifts.

4.3. Estimation

We have a 'limited dependent variables' setting. In principle we should adopt non-linear estimators, since linear regression would assume the wrong functional form. As Angrist and Pischke (2009) point out, in certain cases non-linear models should be preferred on 'curve-fitting' grounds. For interpretation, however, results need to be converted into marginal effects, and when this is done, slope coefficients are very similar to OLS estimates. Practically, linear applications are also more parsimonious and easier to handle, especially when using panel data and when large numbers of fixed effects are required (Schmidheiny, 2013). In our case, the estimation of the coefficients of interest is conditional on being able to observe time-varying and invariant pair-level characteristics. Fitting pair-level fixed effects is therefore not feasible, and fitting individual-level dummies as fixed effects would tend to yield biased estimates as well as being computationally challenging (over 150,000 variables). That is, 'correct' functional form comes at a substantial cost.

Our preferred approach is to fit a linear model with robust standard errors, which for the binary outcome variable is simply a linear probability model. We then compare estimates from non-linear estimators in robustness checks.[9] For the counts model, we compare OLS with negative binomial estimates (we have over 90% zeroes and Poisson conditions are not met). We are only able to achieve convergence with a **PROX** plus observables model, as noted above. Results are given in Table C.4. For the binary model, excess zeroes mean that a standard logit model will not converge. We are able to achieve convergence using a rare events model (Boschma et al., 2013; King and Zeng, 1999): in this case, convergence is only feasible for a reduced-form observables model without **ENV**, with results very similar to equivalent OLS estimates.

4.4. Other issues

We briefly review four other issues that might affect our results. The first is endogenous partner selection (Ackerberg and Botticini, 2002). Consider a contract decision between a principal P and an agent A. Ideally P and A observe everything about each other, reaching the optimal contract. In reality, there are unobservable qualities of P and A which affect type of contract chosen. Our 'observables' specification faces a version of this problem; the fixed effects specification handles non-time-varying unobservable characteristics, but dynamic unobservable factors are left unobserved. As we set out above, this means our results are descriptive associations.

A second issue concerns the presence of third parties. So far we have looked at the characteristics of co-inventing teams assuming that A's decision to co-invent with B (or not) is not affected by the presence of C or D. But as Sedikes et al. (1999) point out, this assumption may not hold. A decision to partner with A rather than B may be affected by the presence of C, which shifts relative positions of A-B-C on specific decision axes. In this case, patent data gives us a limited view on collaboration structure by allowing us to see inventor team size. We use this to generate a team size count variable. We also fit a further variable giving the average number of co-inventors for each inventor in a given year: the intuition here is that inventors may seek out teams of a given size.

Third, our data structure implies that we do not actually *observe* inventors when they are not filing patents: they may be working on other inventions, or may be inactive. If there are structural patterns here, this may lead to omitted variable bias. We could therefore set inventor pair activity to zero in cells where patenting does not occur; or, more conservatively, *blank* all cells in which inventor pairs are not active. In a related paper, Nathan (2014) tests both approaches on a subset of multiple patent inventors – both deliver identical results. We therefore feel confident with a zero-basing assumption for our analysis.

Fourth, it is always possible that co-patenting is driving people to move in closer geographical proximity or to become part of the same firm. Without an exogenous source of variation in these proximities, selection issues are harder to deal with. We test the incidence of physical movement following the procedure in Agrawal et al. (2006), who identify 14.2% of inventors as *likely* movers across TTWAs, as well as using a more cautious strategy developed by Crescenzi and Gagliardi (2015), who report 5% of likely movers among UK inventors during the 1992–2007 period. This implies spatial movement is unlikely to affect results, but we are unable to easily develop parallel tests for other selection channels. For example, according to the PatVal survey, 34.7% of UK inventors who had patented changed job at least once, albeit during the whole length of their careers (Brusoni et al., 2007), making job-to-job mobility a potential additional channel of selection before researchers can observe any patenting activity. This implies that, once again, our results are treated as associations, rather than causal effects.

5. Results

The regression analysis is organised into three sections. The first section looks at the results for all inventors (larger sample, but more limited set of explanatory variables in terms of proximities), while the second section looks at the sub-sample of multiple patent inventors. The third section includes a number of robustness checks.

5.1. All inventors

For the full sample of inventors it is possible to explore the incidence of geographic, organisational and cultural/ethnic proximities. Results for the collaboration dummy are given in Table 1, and for co-invention counts in Table 2. Columns 1 to 3 include the basic specification with geographic, cultural/ethnic, and organisational proximities alone (Column 1), plus observables (Column 2), and plus fixed effects (Column 3). Columns 4 and 5 look at the time split for the 1990s and 2000s. In the interpretation, we focus on the relative sign and significance of the proximities variables, rather than unpacking specific point estimates.[10]

5.1.1. Collaboration

The results presented in Table 1 point to diversified associations between proximities and co-patenting. Co-patenting teams are formed predominantly within the boundaries of the same organisation in culturally and ethnically diverse groups. Teams are formed by preferably relying on long-distance means of communication in order to include collaborators with the necessary skills and possibly tap into the knowledge of remote locations.

[9] We estimate using the user-written Stata command reg2hdfe: this uses an iterative procedure on each coefficient to achieve convergence, following Guimaraes and Portugal (2010). Importantly, this approach provides a clear interpretation of the fixed effects, avoiding arbitrary 'holdouts'/reference categories.

[10] For the full sample, Table B.1 in Appendix B gives summary statistics (first panel) and correlations matrices of the proximities variables (second panel). Correlation matrices indicate that our variables of interest are free of collinearity problems. The results are confirmed in VIF tests. Other model fit statistics are also satisfactory, with R^2 around 0.38.

Table 1
All inventors, co-invention dummy, 1992–2007.

	(1) Proximities	(2) Proximities with controls	(3) Proximities with fixed effects	(4) Proximities with fixed effects 1990s	(5) Proximities with fixed effects 2000s
Geographic proximity	−0.00612*	−0.0207***	−0.0320***	−0.0526***	−0.0236***
	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)
Cultural/ethnic proximity	−0.00426***	−0.00922***	−0.0158***	−0.0209***	−0.0137***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Organisational proximity	0.304***	0.118***	0.120***	0.152***	0.107***
	(0.004)	(0.004)	(0.001)	(0.002)	(0.001)
IND, **INST**, **ENV** controls for *ij*	N	Y	N	N	N
Spells fixed effects for *ij*	N	N	Y	Y	Y
Observations	190,313	117,456	190,313	54,415	135,898
F	1729.627	39.960	27.508	19.286	32.871
R^2	0.271	0.526	0.662	0.660	0.666

All models use time dummies. Robust standard errors in parentheses.
IND vector includes pre-1992 patent count and dominant pre-92 patenting style for *i,j*.
INST vector includes applicant type dummies for *i,j*.
ENV vector includes IPC6 grouping variable, TTWA dummies for *i,j* and year dummies.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Column 1 suggests a robust positive association between organisational proximity and inventor pairs' tendency to co-invent. By contrast, cultural/ethnic proximity is strongly negatively connected to the propensity to co-invent and geographical proximity is only marginally significant. The introduction of controls for individual, institutional, and environmental conditions (Column 2), as well as fixed effects in Column 3 reduces the magnitude of the organisational proximity coefficient, while increasing the negative coefficient of cultural ethnic proximity. These controls also make the coefficient for geographical proximity highly significant. The results also indicate no particular change in the incidence of all proximities between the 1990s (1992–1999, Column 4) and the 2000s (2000–2007, Column 5), except for a marginal decrease in their size.

5.1.2. Inventor pair activity

In Table 2 we look at whether the results attained when assessing the characteristics of individual collaborations stand when considering the number of patents actually generated by the 'actual' inventors' pairs. The structure of the regressions presented in Table 2 is the same as that described for Table 1. The results in Table 2 are also broadly similar to the co-invention dummy models, suggesting that organisational proximity and cultural diversity are essential elements for collaboration between inventor pairs. The coefficient for the variable depicting geographical proximity turns negative only after the introduction of fixed effects: geographical proximity is correlated with a number of unobservable factors that, if not properly accounted for, are likely to bias estimated coefficients. Once again, no noticeable differences are evident between the 1990s and 2000s.

Overall, these results suggest that firms and research centres are the loci of knowledge combination and re-combination, the fundamental units of analysis of the innovation process. It is when inventors are within the same organisational boundaries that problems of signalling (e.g. in terms of the quality of the possible collaborators), free riding and procrastination (once teams are formed), and secrecy are successfully dealt with. By sharing the same set of internal routines, rules and norms inventors are able to work in teams in the most productive fashion. At the same time, within the boundaries of the same organisation, inventors can more

Table 2
All inventors, co-invention counts, 1992–2007.

	(1) Proximities	(2) Proximities with controls	(3) Proximities with fixed effects	(4) Proximities with fixed effects 1990s	(5) Proximities with fixed effects 2000s
Geographic proximity	0.0197**	0.000539	−0.0180***	−0.0349***	−0.0118***
	(0.009)	(0.007)	(0.003)	(0.007)	(0.003)
Cultural/ethnic proximity	−0.00517***	−0.00964***	−0.0190***	−0.0263***	−0.0160***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.002)
Organisational proximity	0.343***	0.0996***	0.116***	0.169***	0.0927***
	(0.009)	(0.014)	(0.003)	(0.007)	(0.003)
IND, **INST**, **ENV** controls for *ij*	N	Y	N	N	N
Spells fixed effects for *ij*	N	N	Y	Y	Y
Observations	190,313	117,456	190,313	54,415	135,898
F	584.899	22.622	5.893	3.805	7.471
R^2	0.090	0.169	0.296	0.277	0.311

All models use time dummies. Robust standard errors in parentheses.
IND, **INST** and **ENV** control vectors as in Table 1.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

easily solve confidentiality problems, facilitating knowledge circulation.

Conversely, the results signal that the direct incidence of geographical proximity is complex and diversified. Given that inventors in the same organisation are likely to be co-localised in the same premises, organisational proximity also sheds light on the relevance of hyper-geographical proximity within the firms' boundaries. UK inventors first seek the collaboration of – or are coerced to collaborate with – other inventors within the boundaries of their organisations. Inventor teams are then complemented by researchers and specialists who can provide an additional edge to the team. This search for adequate co-inventors is by no means constrained by geographical boundaries and tends to follow non-spatially mediated channels. Given the higher costs of searching outside the firms' boundaries as well as the other frictions that affect collaborations across organisations, the search for the optimal partner tends not to be limited by spatial proximity (Fitjar and Rodríguez-Pose, 2011).

Contrary to some recent US studies (Kerr and Lincoln, 2010; Saxenian and Sabel, 2008), our main results also suggest that, on average, UK-based inventors seek cultural/ethnic/linguistic diversity rather than the proximity, trust and lower transaction costs which may be possibly associated with collaborating with co-nationals or with individuals of the same cultural/ethnic origin. The typical inventor seems less likely to work with people from the same national and/or cultural background even when they are physically close.[11] Ethnic proximity is likely to 'influence' the optimal search process of inventors beyond geographic proximity, eventually reducing the incentives to collaborate and increasing the risk of cognitive lock-in, in particular when this is coupled by geographic proximity. These results echo the evidence on the negative influence of 'bonding social capital' on innovation (Crescenzi et al., 2013): 'strong ties' increase the risk of exchanging redundant knowledge simply because they connect knowledge seekers with other individuals that are more likely to deal with 'known'/familiar information and knowledge (Laursen and Masciarelli, 2007; Levin and Cross, 2004; Ruef, 2002). This evidence effectively complements our previous results: once the search process overcomes firms' organisational boundaries, it follows trajectories that are not constrained by other proximities. On the contrary, it seems that team formation tends to privilege diversity in terms of non-local knowledge/skills/competences, as well as a variety of cultural and ethnic backgrounds.

5.2. Multiple patent inventor analysis

Do these patterns stand when we consider the subset of multiple patent inventors in detail? Are various proximities balanced in the same way by these highly innovative individuals? The focus on multiple patent inventors makes it possible to reduce the potential 'white noise' generated in patent data by 'hobbyist inventors', who patent only once in their life with limited impact on their technology field (Lettl et al., 2009). In addition, focusing on multiple patent inventors allows us to test different types of proximities – social and cognitive proximity – impossible to operationalise in the single inventor analysis.[12] The possibility to explore a wider set of non-spatial proximities facilitates the comparison of the incidence of networks based on cultural/ethnic proximity with potentially less 'constrained' and more 'diverse' collaboration-search patterns based on social and cognitive proximity. The results are given in Tables 3 and 4.

5.2.1. Collaboration

Collaboration results for multiple patent inventors are covered in Table 3. Column 1 introduces fixed effects, as in Column 3 in the previous tables. Social and cognitive proximity are included in Column 2, while Columns 3 to 7 interact each proximity with geographical distance. Finally, Columns 8 and 9 split the sample into the 1990s and 2000s time periods.

The most striking results from the analysis of multiple patent inventors are that organisational proximity remains main feature of co-patenting collaborations while cultural/ethnic proximity loses significance and geographical proximity emerges as a relevant characteristic for co-patenting teams, when interacted with other proximities (Table 3). The coefficient for organisational proximity is always positive and significant at the 1% level. This implies that multiple patent inventors do not become less constrained by organisational boundaries than occasional inventors. Indeed, their collaborations tend to have precisely the same characteristics, suggesting that they prefer to work with colleagues from the same organisation when deciding to collaborate or create networks. Potentially more 'senior' and experienced individuals with longer patenting experience still leverage their organisations as the fundamental milieu for their innovative activities. This reliance on organisational proximity increases in the 2000s (Table 3, Column 9). The sharing of highly sophisticated and confidential material via information and communication technologies has reinforced the importance of platforms and systems internal to the firm.

If the formation of teams takes place largely within the same organisational unit, external search patterns are also important. Social proximity is always positively and significantly associated with repeated invention. UK multiple patent inventors more frequently patent with other inventors with whom they had previously established direct or indirect collaborations (Table 3, Column 2). Cognitive proximity, measured on its own, by contrast seems to be negatively associated with collaboration by serial inventors (Table 3, Column 2). Multiple patent inventors work together with individuals with whom they had developed pre-existing connections via social networks, rather than with those in similar cognitive areas. Connections mediated by social proximity are stronger than those 'constrained' by cultural/ethnic factors and cognitive proximity. Indeed, serial inventors' teams are associated with cognitive diversity and complementary knowledge bases. The need to avoid lock-in and search for cognitive diversity and complementary knowledge drive the selection of collaborators within 'known' networks, so as to minimise search costs, while preserving diversity and access to non-redundant knowledge.

The analysis of the interaction effects of the different proximities considered in relation to geographical distance provides some interesting nuances to this picture. On the one hand, the relevance of cultural/ethnic proximities seems to emerge only at close quarters. Multiple patent inventors in the UK tend to work more with colleagues from the same cultural and/or ethnic background, only if they are based in the same or in nearby locations (positive interaction term in Table 3, Column 3). A similar tendency emerges for both organisational and social proximity: the incidence of organisational (Column 4) and social (Column 5) proximities in collaborations is further reinforced by geographic proximity. By

[11] This may not be the case for all inventors; Nathan (forthcoming) finds evidence of co-ethnic links to patenting for some minority ethnic 'stars'. Our result is also sensitive to the precise ethnicity proxy used – see robustness checks in Section 5.3.

[12] Table B.2 in Appendix B gives summary statistics (first panel) and correlations matrices of the proximities variables (second panel). In this case, because the social and cognitive proximity variables are based on past inventor behaviour, pairwise correlations between these proximities and dependent variables are higher than desirable (between 0.5 and 0.54), but do not indicate fatal collinearity problems. With the patent sampling base increased from 5 to 25%, model fit statistics are substantially higher than for the full panel, with the R^2 rising to about 0.8 with controls.

Table 3
Multiple patenting inventors, co-invention dummy, 1992–2007.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Base with fixed effects	Social & cognitive	Interaction terms					1990s	2000s
Geographic proximity	0.00462***	0.00383***	0.00154	0.00148	0.00262***	0.00390***	−0.00107	−0.00311*	0.000557
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Cultural/ethnic proximity	−0.000343	−0.000420	−0.000741	−0.000442	−0.000377	−0.000419	−0.000711	−0.00200**	0.0000191
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Organisational proximity	0.0237***	0.0215***	0.0213***	0.0160***	0.0210***	0.0215***	0.0171***	0.0120***	0.0239***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Social proximity		0.0625***	0.0625***	0.0618***	0.0351***	0.0627***	0.0356***	0.0261***	0.0441***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)
Cognitive proximity		−0.00498***	−0.00498***	−0.00468**	−0.00522***	−0.00430**	−0.00284	−0.00172	−0.00406
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Geographic prox.*cultural/ethnic prox.			0.00437***				0.00435***	0.0101***	−0.0000359
			(0.001)				(0.001)	(0.002)	(0.002)
Geographic prox.*organisational prox.				0.0128***			0.00886***	0.00588**	0.00894***
				(0.002)			(0.002)	(0.003)	(0.003)
Geographic prox.*social prox.					0.0666***		0.0663***	0.0725***	0.0812***
					(0.005)		(0.005)	(0.007)	(0.008)
Geographic prox.*cognitive prox.						−0.00576	−0.0187***	−0.0321***	−0.00287
						(0.006)	(0.006)	(0.008)	(0.009)
Observations	40,638	40,638	40,638	40,638	40,638	40,638	40,638	19,710	20,928
F	27.691	28.266	28.269	28.295	28.432	28.261	28.451	34.030	20.916
R^2	0.821	0.824	0.824	0.824	0.825	0.824	0.825	0.865	0.762

All models use time dummies and spells fixed effects. Robust standard errors in parentheses.
IND, **INST** and **ENV** control vectors as in Table 1.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

contrast, geographic proximity does not reinforce the recurrence of collaborations in the same technology field (Column 6). A negative interaction only appears in Column (7) where all interaction terms are included simultaneously and the direct effect becomes insignificant. This result is in line with Rigby (2013), who finds a trend in the US for cities to build invention competence around a range of related technologies – if they are based in distant locations, suggesting that geographic and cognitive proximity tend to be substitutes rather than complements. However, these results are mainly a consequence of trends in the 1990s. In the 2000s neither

Table 4
Multiple patenting inventors, co-invention counts, 1992–2007.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Base with fixed effects	Social & cognitive	Interaction terms					1990s	2000s
Geographic proximity	−0.0108	−0.0134*	−0.0512***	−0.000288	−0.0139*	−0.0183**	−0.0418***	−0.0697***	−0.0193*
	(0.008)	(0.008)	(0.011)	(0.009)	(0.008)	(0.008)	(0.011)	(0.021)	(0.011)
Cultural/ethnic proximity	0.00524	0.00526	−0.0000429	0.00538	0.00528	0.00521	0.00000667	0.00611	−0.00518
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.011)	(0.006)
Organisational proximity	0.0701***	0.0598***	0.0578***	0.0902***	0.0596***	0.0600***	0.0922***	0.154***	0.0118
	(0.011)	(0.011)	(0.011)	(0.014)	(0.011)	(0.011)	(0.014)	(0.023)	(0.014)
Social proximity		0.249***	0.0881***	0.0866***	0.0881***	0.0389**	0.0366*	0.0760**	−0.0210
		(0.024)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.034)	(0.019)
Cognitive proximity		0.0882***	0.249***	0.253***	0.238***	0.232***	0.236***	0.380***	0.0269
		(0.018)	(0.024)	(0.024)	(0.031)	(0.024)	(0.031)	(0.052)	(0.034)
Geographic prox.*cultural/ethnic prox.			0.0721***				0.0727***	0.0983***	0.0467***
			(0.014)				(0.014)	(0.025)	(0.013)
Geographic prox.*organisational prox.				−0.0712***			−0.0797***	−0.178***	0.0759***
				(0.020)			(0.021)	(0.035)	(0.021)
Geographic prox.*social prox.					0.0276		−0.000994	−0.221***	0.253***
					(0.048)		(0.049)	(0.081)	(0.055)
Geographic prox.*cognitive proximity						0.422***	0.424***	0.867***	−0.0960*
						(0.056)	(0.057)	(0.097)	(0.058)
Observations	40,638	40,638	40,638	40,638	40,638	40,638	40,638	19,710	20,928
F	4.711	4.750	4.757	4.753	4.749	4.767	4.777	3.931	6.564
R^2	0.438	0.440	0.441	0.440	0.440	0.441	0.442	0.425	0.501

All models use time dummies and spells fixed effects. Robust standard errors in parentheses.
IND, **INST** and **ENV** control vectors as in Table 1.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

the coefficient for the interaction between geographic proximity and cultural/ethnic proximity, nor between the former variable and cognitive proximity remain significant. Only the strong connections between social and geographic proximity and between the former and organisational proximity stay significant in the 2000s (Column 9). Technological development plays an ancillary role in supporting the process of search along organisational and social networks trajectories.

Hence, as in the case of occasional inventors, organisational proximity is an essential characteristic of co-patenting teams, while social proximity emerges as another crucially important factor. Multiple patent inventors based in the UK patent much more with other inventors in close organisational proximity and within established social networks, with space mediating the relevance of all types of proximities considered.

Teams formed by multiple patent inventors collaborate more with other inventors whom they may be able to meet on a more frequent basis in the same organisation and with whom they have collaborated in the past and may be easily accessible via a number of telecommunications channels. This implies that ideas are not necessarily in the air – as implied in Marshallian and agglomeration approaches – but that they flow inside very well-structured organisational and social pipelines (Fitjar and Rodríguez-Pose, 2015). Nevertheless, the functioning of these communication channels improves in close geographical proximity. Similar conclusions are reached by Cassi and Plunket (2010), who study genomics patents in France and suggest that spatial proximity is highly complementary to social proximity, even if individual partners are in different types of organisations. Singh (2005) confirms that the geography and firm boundaries interact with interpersonal networks in shaping knowledge flows. However, our results contrast with those of Agrawal et al. (2008), who conclude that geographic and social proximity operate as substitutes in the USA.

5.2.2. Multiple patent inventor pair activity

Table 4 covers inventor pair activity. As for the full set of inventors, the results for frequency of collaborations by multiple patent inventors largely resemble the simple collaboration analysis discussed above. Social and, above all, organisational proximity remain highly significant in all specifications (Columns 1 and 2). However, in the case of co-invention counts, cognitive proximity shows a positive and significant sign, suggesting multiple collaborations happening more frequently within the same technology class.

While the interaction between cultural/ethnic and geographical proximities remains unchanged (Column 3), other interaction terms (Columns 4–7) depict a slightly different story than previous results. Organisational and geographical proximity display significant coefficients with the expected sign, but their interaction is not significant. This suggests that they do not reinforce one another, hinting at potential substitution, rather than complementarities (Column 5). Finally, the interaction between cognitive and spatial proximities is positive and significant, meaning that frequent patenting by serial inventors is potentially associated with specialised spatial clusters. Time splits (Columns 8 and 9) indicate that these differences with the collaboration dummy are most prevalent in the 1990s. For the 2000s the signs and significance of the interaction terms are in line with the previous table.

5.3. Robustness checks

Section 4 highlights a number of potential challenges to our empirical strategy. To deal with these, we subject our model to a series of robustness checks (reported in Appendix C). Results not shown here are given in an online appendix (Appendix D).

We first test for measurement and specification issues. Table C.1 covers three key tests (the others are available in the online appendix). Column 1 refits the main result for the collaboration dummy. Column 2 removes all patents where inventors and applicants are probabilistically matched (see Section 4.2). The betas change slightly but the overall pattern of results remains the same. This suggests that the patent assignment process does not affect our results.

Column 3 introduces a scaled organisational proximity measure. Proximity is assigned greater weight for single applicant patents than for co-assigned patents. The coefficient is positive and significant, confirming our intuition that there is a stronger link to co-patenting in the former cases than when other applicants are involved, even if both members of the pair work for the same organisation.

Column 4 includes a dummy for *applicant* spatial proximity, taking the value 1 if applicants share the same TTWA. We fit this in order to try and disentangle individual and applicant location issues (as explored in Lychagin et al. (2010) among others). This variable is fuzzy by construction, since it can only be directly observed for actual pairs on co-assigned patents, a selected sub-sample: for possible pairs, applicant information is taken from *any other* patent and may not correspond to an inventor's 'real' applicant in any given year. As such it may be a case of 'bad control' (Angrist and Pischke, 2009) and needs to be taken with some caution. Here, the coefficient is positive significant, and the beta for geographic proximity drops slightly. Taken at face value, this implies that while working for the same organisation is the principle link to co-patenting, working in co-located applicants also has a smaller, independent connection.

In other tests (shown in the online appendix, Table OA1), we refit the collaboration dummy model with alternative measures for geographic proximity (a linear distance threshold and a TTWA dummy for inventors in a pair) and cultural/ethnic proximity (same ONS ethnic group, same geographical origin, same major language). The overall pattern of the results does not change from the main specification. We also run all of these tests for the count model, with little or no change from our main estimates. Further checks explore the role of inventor team size, and suggest that while size is a relevant characteristic of co-patenting teams, it does not 'knock out' proximities (Table OA2).

Table C.2 repeats these checks for multiple patent inventors, and adds some further tests. As with the pooled sample, rescaling organisational proximity suggests that organisational proximity is strongest when patents have single applicants (Column 2). Column 3 fits an alternative social proximity measure, rescaled as an ordinal variable; Column 4 refits cognitive proximity using larger IPC3 technology fields instead of IPC6. Neither changes our main results. Column 5 fits applicant spatial proximity, with similar results to the pooled sample. Column 6 removes co-assigned patents, also with little change. As before, we also refit geographic and cultural–ethnic proximities (Table OA3). Specifying the latter as 'same major language' generates a marginally significant positive coefficient: this is weak evidence that linguistic ties may matter for repeat collaborators in a way that broader cultural–ethnic groupings do not. Other re-specifications do not change our main estimates. Results for collaboration counts are also available online (Table OA4).

A second set of concerns centres on estimation issues. We begin by shifting from robust standard errors to errors clustered on inventor pairs: results are identical (Table OA5). Next, we tighten the fixed effect specification to include IPC6 technology field-by-year fixed effects, allowing shocks to vary over time and technology space simultaneously (Table C.3). Compared to Column 1, the main specification in Column 2, geographic proximity, is now close to zero and non-significant. Estimates for other proximities barely change. These results suggest that geographic proximity may be sensitive to fixed effects specifications: however, model fit substantially decreases, implying a potentially sub-optimal specification.

Table C.4 compares linear and non-linear estimators for the counts model. As discussed in Section 4, we contrast the linear estimator with a negative binomial specification, using an observables model without full fixed effects. As all unobserved factors are omitted, the test simply illustrates different functional forms. Column 1 gives the linear model, Column 2 the raw coefficients from the negative binomial, and Column 3 the marginal effects. The latter shows that with a non-linear estimator, geographic proximity is positive but only marginally significant, compared to zero in the linear model. Coefficients for other proximities keep the same sign and significance. This implies that the choice of a linear estimator does not bias our results.

Finally, there is the concern that our results derive from a particular sub-sample of patents and inventors. Since sampling patents and inventors is the basis of our empirical strategy, it is important to test this. To do so, we rebuild the main panel five times using different samples of patents. We then re-run the collaboration dummy model on each panel in sequence; while coefficient sizes vary slightly, the sign and significance of the **PROX** variables remain unchanged (Table OA6), indicating that our results are robust to sample choice.

6. Conclusions

Innovation has become an increasingly collaborative activity in recent years. Inventors cooperate more than ever in the invention process, but our knowledge of the factors that determine the formation of inventor teams is still incomplete.

This paper has looked at the characteristics of co-patenting teams in order to explore the incidence of different proximities. Using a rich microdata set, we have been able to focus on individual inventors across the full range of technology fields and in different time periods, examining geographic, cultural/ethnic, organisational, social, and cognitive factors, individually and in combination. In doing so, we address empirically a number of issues which have generally been considered from a more theoretical perspective in the literature.

Our results – which control for fixed effects and are robust to multiple cross-checks – contain a number of important findings which in part challenge the existing literature and policy consensus on the importance of spatial clustering for knowledge exchange and innovation.

The analysis of the features of UK patenting teams shows that a significant part of innovation activities are characterised by organisational proximity. That is, they take place within the boundaries of private firms (that account for the largest share of all applicants), universities and other organisations. Sharing the same applicant captures the simultaneous importance of hyper-geographical proximity (inventors in the same organisation are likely to be co-localised in the same premises) and organisational proximity. Second, the results show that ethnic and cultural diversity are also important features of inventing teams and that social networks represent the building blocks for collaboration. Third, geographic proximity enters the picture only indirectly: it interacts with other proximities increasing their association with collaborative work. Inventors seem to rely on geographical proximity to form their teams when it is coupled with other forms of ‘advantage’.

Overall, the results highlight important differences between the proximity relationships linking UK inventors over a long time span. However, it has to be borne in mind that our results suffer from some limitations. First, without exogenous variations in the observed proximities, selection effects cannot be controlled for, and results have to be interpreted as associations. Second patent data can only capture collaborations that lead to a patented output. We do not observe unproductive collaborations (or not-yet-productive ones). Third, we cannot capture collaborations that lead to non-patentable output – for example in the form of process innovation or innovation in services (an important part of UK innovation activity as highlighted in Crescenzi et al., 2015). Finally, the reliance on patent data might lead to an over-estimation of the importance of organisational proximity. The tendency of inventors to share the same applicant inevitably reflects fundamental incentives about disclosure versus secrecy in R&D projects.

Having acknowledged these limitations, our results make innovative contributions to the existing literature on innovation and its drivers on several fronts. First, the analysis of innovation and its geography cannot be restricted to the study of spatial relationships between innovation agents: organisational structures and firms' boundaries play a key role – often underestimated in the existing literature – in shaping innovative collaborations and knowledge diffusion. Second, when inventors search for collaborators outside the boundaries of their organisations, ethnic ties or cognitive proximity are not necessarily the preferred search channel. Contrary to what part of the existing literature suggests (e.g. Docquier and Rapoport, 2012 on cultural/ethnic proximity in the US or Kogler et al., 2013; Rigby, 2013 on cognitive proximity), ‘unconstrained’ social networks that allow for diversity to emerge and avoid cognitive lock-in are prevalent in UK patenting teams. Third, the innovative approach of this paper – combining for the first time a large set of different proximities in the same inventor-level analysis – has uncovered hitherto unobserved synergies and substitution effects between various proximities. The focus of the existing literature on specific spatial units of analysis (e.g. cities) or on more limited sub-sets of proximities has often neglected the underlying complexity of the phenomena under analysis.

The three-fold innovative contribution of our paper has important implications for both academic research and practical policy-making.

In terms of future research our analysis, rather than closing doors, opens up a number of other avenues. Our results for cognitive proximity would need follow-up work testing out alternative specifications of technological closeness. We have also chosen deliberately simple social proximity measures. Further work could use more complex social proximity metrics, or focus on hub inventors' ego-networks rather than inventor pairs. Finally, given the scope of the paper, we have been unable to delve into other interesting aspects, such as age or gender differences, which intuition suggests may be important influences on what inventors do. While the paper provides useful and novel answers to the question of what are the relational and spatial features of collaborations among inventors, many other questions remain to be answered. We believe we have opened an important gate, however further analysis is needed in order to improve our understanding about the reasons why inventors talk to each other.

The results of our analysis (as well as further reflections and work in this field of research) are particularly important in order to inform innovation policies in the United Kingdom and in Europe more generally. Policy-makers have been attracted for a long time by the concept of innovation clusters and have allocated substantial public resources to their support and promotion. The rationale behind these policies has been provided by the assumption that geographical clustering would per se support knowledge exchange and innovation. Further analysis on the complementarities between geographical proximity and other forms of proximities is crucial in this regard. An emerging body of evidence seems to increasingly point in the direction of an ancillary role being played by spatial clustering: if other proximity conditions are not simultaneously in place, spatial clustering may – as our research points out – be of limited utility to innovation. At the same time the search for knowledge and competences outside the organisational boundaries of firms, research centres and universities might follow highly complex (and evolving) patterns hard to anticipate by policy-makers and difficult to generalise and target by dedicated policy tools. Policies supporting the formation of networks and collaborations within pre-selected technological fields or groups of countries/regions might work against the 'unconstrained' networks on which innovative agents instead need to rely on. In this context, policy efforts would be better targeted towards the minimisation of information asymmetries and the support of innovation agents' capabilities to scan external opportunities and identify the most suitable partnerships/collaborations.

Acknowledgements

The authors are enormously grateful to Guido Conaldi, Riccardo De Vita and Sara Gorgoni for their invaluable assistance with social network analysis. Thanks also to Steve Billings, Philippe Bracke, Stefano Breschi, Lorenzo Cassi, Giulia Faggio, Maryann Feldman, Steve Gibbons, Simona Iammarino, Adam Jaffe, Bill Kerr, Hans Koster, Camilla Lenzi, Francesco Lissoni, Henry Overman, Anne Plunket and Rosa Sanchís-Guarner for help and advice. Participants at the 2012 Urban Economics Association conference, 2012 EUROLIO Geography of Innovation workshop in Saint-Etienne, the 2012 ESF-APE-INV seminar in Leuven, and internal seminars at LSE and SERC contributed useful comments. The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007–2013 under Grant Agreement SSH-CT-2010-266959, from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement No. 269868 and from BIS, DCLG and the ESRC through the Spatial Economics Research Centre under award number ES/L0003945/1. All errors and omissions are our own.

Appendix A.

A.1. The ONOMAP name classification system

We use the ONOMAP name classification system to generate ethnicity information for individual inventors. Naming relates to cultural, ethnic, linguistic features of individuals, families and communities, and is highly persistent over time even after substantial population mobility (Mateos et al., 2011). ONOMAP classifies individuals according to most likely cultural, ethnic and linguistic characteristics, identified from forenames, surnames and forename-surname combinations. The reference population is 500,000 forenames and one million surnames, drawing on electoral registers and telephone directories for the UK and 27 other countries. ONOMAP classifies inventor names via an algorithm that uses surname, forename and surname-forename combinations, exploiting name-network clustering between surname and forename pairings (Mateos, 2007). In most cases both name elements share the same 'type'; in other cases the most likely type is assigned, based on frequencies in the reference population. The final classification comprises 185 'cultural–ethnic–linguistic' (CEL) types, and larger 67 subgroups. ONOMAP also provides information on CEL components such as geographical origin and major language, as well as the nine 'macro-ethnic' groups developed by the UK Office of National Statistics.[13]

ONOMAP has been extensively tested on individual datasets where ethnicity is known, typically matching over 95% of names and giving very low measurement error (Lakha et al., 2011; Petersen et al., 2011). ONOMAP usefully provides information at several levels of detail and across several dimensions of identity. It is also able to deal with Anglicisation of names, and names with multiple origins, giving it additional granularity and validity. Conversely, ONOMAP is unable to observe immigrants, and should be interpreted as assigning *most likely* cultural identity. However, unlike the MELISSA commercial database used by Kerr (2008), which only identifies high-level ethnicities, ONOMAP provides much more detail and dimensionality. ONOMAP also matches 99% of inventor names (compared with Kerr's 92–98% success rates).[14]

A.2. List of variables

Variable name	Definition	Source
Dependent variables		
DCoinvent	Dummy variable for inventor pairs, coded as 1 if pair patent together in a given year, 0 if not.	KITES-PATSTAT
#Coinvent	Continuous variable for inventor pairs, recording the count of collaborations in a given year.	KITES-PATSTAT
Independent variables		
Geographic proximity	Inverse linear distance in km between Travel to Work Area (TTWA) centroids occupied by each inventor in a pair, based on inventor address. Normalised to take maximum value 1.	KITES-PATSTAT, UK Office of National Statistics
Organisational proximity	Dummy taking the value 1 if pair belong to the same applicant, 0 if not.	KITES-PATSTAT
Cultural/ethnic proximity	Dummy for inventor pairs, set as 1 if both are in the same ONOMAP 'cultural–ethnic–linguistic' (CEL) subgroup, 0 if not, blank if unknown. ONOMAP coding is based on inventor name information	KITES-PATSTAT, ONOMAP
Cognitive proximity	Dummy for multiple patent inventor pairs, set as 1 if both have previously patented in the same 6-digit IPC technology field, 0 if not.	KITES-PATSTAT
Social proximity	Inverse social distance between inventors in a pair. For a given year, social distance is defined as the number of steps between pair members in the previous five years, from 0 (collaboration) to minus infinity (no connection).	KITES-PATSTAT, University of Greenwich

[13] The full set of ONS 1991 groups is White, Black Caribbean, Black African, Indian, Pakistani, Bangladeshi, Chinese and Other. The full set of twelve geographical origin zones is Africa, Americas, British Isles, Central Asia, Central Europe, East Asia, Eastern Europe, Middle East, Northern Europe, South Asia, Southern Europe and Rest of the World. See Nathan (2014) for the full classification of 67 CEL subgroups.

[14] We remove all conflict cases from the sample.

Variable name	Definition	Source
Control variables		
IND	Vector of individual characteristics controls for each inventor in a pair: 1. Dummy taking the value 1 if inventor is active in the pre-sample period 1978–1991, 0 if not; 2. Inventor's average patenting in the pre-sample period 1978–1991, zeroed if inventor is inactive pre-sample; 3. Dummies for inventor's type of patenting activity in pre-sample period: (i) always solo, (ii) always co-inventing, (iii) mix solo and co-inventing, (iv) inactive. Inactive is set as the reference category.	KITES-PATSTAT
INST	Vector of institutional characteristics controls for each inventor in a pair: 1. Dummies for type of inventor's applicant type, coded as (i) business/private research lab, (ii) university/public research lab; (iii) foundation/NGO/consortium; (iv) individual. Unknown is the reference category.	KITES-PATSTAT
ENV	Vector of macro/area characteristics controls for each inventor in a pair: 1. Year dummies; 2. Grouping variable for 6-digit IPC technology fields, zeroed for potential pairs; 3. TTWA dummies.	KITES-PATSTAT

Italics denotes used in robustness checks only. Controls are used in observables model only.

Appendix B. Summary statistics and correlation matrices

Tables B.1–B.4

Table B.1
Patents by applicant types.

Variable	Obs	Mean	Std dev	Min	Max
Applicant is individual (inventor)	116,325	0.0663	0.248	0	1
Applicant is business or private research centre	116,325	0.555	0.490	0	1
Applicant is university or public research centre	116,325	0.0396	0.193	0	1
Applicant is foundation or NGO or consortium or other	116,325	0.00289	0.0533	0	1
No applicant info mapped	116,325	0.304	0.452	0	1

Source: Own elaboration using KITES-PATSTAT.

Table B.2
Summary statistics and correlation matrix: full sample.

Variable	Obs	Mean	Std dev	Min	Max
Dcoinvent	190,313	0.0201	0.140	0	1
#coinvent	190,313	0.0244	0.285	0	32
Team size, for *ij* pairs part of a team	190,313	0.0730	0.631	0	12
Geographic proximity (linear distance)	190,313	0.0636	0.233	0	1
Geographic proximity (200 km)	190,313	0.232	0.332	0	1
Geographic proximity (TTWA)	190,313	0.0583	0.234	0	1
Organisational proximity	190,313	0.0631	0.243	0	1
Organisational proximity (scaled)	186,407	0.127	0.487	0	2
Cultural/ethnic proximity	190,313	0.294	0.455	0	1
Cultural/ethnic proximity (geographic origin)	190,313	0.459	0.498	0	1
Cultural/ethnic proximity (ONS)	190,313	0.545	0.498	0	1
Cultural/ethnic proximity (same language)	190,313	0.442	0.497	0	1
Applicants share local area dummy	176,774	0.0572	0.232	0	1

Variable	Geo prox (TTWA)	Geo prox (linear)	Geo prox (200 km)	Cultural/ ethnic proximity	Cultural/ ethnic prox (geo)	Cultural/ ethnic prox (ONS)	Cultural/ ethnic prox (lang)	Organisational proximity	Organisational prox (scaled)	Applicants share local area
Geographic proximity (TTWA)	1.000									
Geographic proximity (linear distance)	0.999***	1.000								
Geographic proximity (200 km)	0.574***	0.599***	1.000							
Cultural/ethnic proximity	0.123***	0.133***	0.294***	1.000						

Table B.2 (*Continued*)

Variable	Geo prox (TTWA)	Geo prox (linear)	Geo prox (200 km)	Cultural/ ethnic proximity	Cultural/ ethnic prox (geo)	Cultural/ ethnic prox (ONS)	Cultural/ ethnic prox (lang)	Organisational proximity	Organisational prox (scaled)	Applicants share local area
Cultural/ethnic proximity (geographic)	0.141***	0.154***	0.372***	0.700***	1.000					
Cultural/ethnic proximity (ONS)	0.165***	0.180***	0.444***	0.589***	0.840***	1.000				
Cultural/ethnic proximity (same language)	0.136***	0.148***	0.357***	0.724***	0.966***	0.813***	1.000			
Organisational proximity	0.440***	0.441***	0.270***	0.052***	0.046***	0.036***	0.045***	1.000		
Organisational proximity (scaled)	0.442***	0.443***	0.272***	0.053***	0.048***	0.038***	0.046***	0.997***	1.000	
Applicants share local area dummy	0.370***	0.372***	0.276***	0.097***	0.112***	0.113***	0.107***	0.386***	0.391***	1.000

Source: Own elaboration using KITES-PATSTAT.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Table B.3
Summary statistics: multiple patent inventors.

Variable	Obs	Mean	Std dev	Min	Max
Dcoinvent	40,638	0.00426	0.0651	0	1
#coinvent	40,638	0.0141	0.349	0	24
Team size, for *ij* pairs part of a team	40,638	0.0102	0.189	0	5
Geographic proximity (linear distance)	40,638	0.0677	0.235	0	1
Geographic proximity (200 km)	40,638	0.317	0.329	0	1
Geographic proximity (TTWA)	40,638	0.0597	0.237	0	1
Organisational proximity	40,638	0.0401	0.196	0	1
Organisational proximity (scaled)	38,401	0.0849	0.403	0	2
Cultural/ethnic proximity	40,638	0.494	0.500	0	1
Cultural/ethnic proximity (geography)	40,638	0.735	0.441	0	1
Cultural/ethnic proximity (ONS)	40,638	0.841	0.366	0	1
Cultural/ethnic proximity (same language)	40,638	0.703	0.457	0	1
Social proximity	40,638	0.005	0.068	0	1
Social proximity (scaled)	40,638	1.011	0.139	1	3
Cognitive proximity	40,638	0.00779	0.0879	0	1
Cognitive proximity (3-digit IPC)	40,638	0.00763	0.0870	0	1
Applicants share local area dummy	38,699	0.0652	0.247	0	1

Source: Own elaboration using KITES-PATSTAT.

Table B.4
Correlation matrix: multiple patent inventors.

Variable	Geo prox (TTWA)	Geo prox (linear)	Geo prox (200 km)	Cultural proximity	Cultural prox (geo)	Cultural prox (ONS)	Cultural prox (lang)	Org proximity	Org prox (scaled)	Social proximity	Social proximity (scaled)	Cognitive proximity	Cogn prox (3-digit IPC)	Applicants share local area
Geographic proximity (TTWA)	1.000													
Geographic proximity (linear distance)	1.000***	1.000												
Geographic proximity (200 km)	0.523***	0.546***	1.000											
Cultural/ethnic proximity	0.017***	0.020***	0.057***	1.000										
Cultural/ethnic proximity (geographic)	0.029***	0.031***	0.032***	0.594***	1.000									
Cultural/ethnic proximity (ONS)	0.045***	0.048***	0.068***	0.431***	0.719***	1.000								
Cultural/ethnic proximity (same language)	0.030***	0.031***	0.036***	0.643***	0.924***	0.668***	1.000							
Organisational proximity	0.445***	0.448***	0.296***	0.024***	0.048***	0.027***	0.052***	1.000						
Organisational proximity (scaled)	0.454***	0.457***	0.303***	0.024***	0.050***	0.029***	0.053***	0.999***	1.000					
Social proximity	0.114***	0.113***	0.072***	0.004	0.014***	0.006	0.014***	0.174***	0.173***	1.000				
Social proximity (scaled)	0.118***	0.117***	0.075***	0.005	0.015***	0.006	0.015***	0.178***	0.178***	0.992***	1.000			
Cognitive proximity	0.020***	0.020***	0.014***	0.002	0.031***	0.018***	0.035***	0.052***	0.052***	0.081***	0.078***	1.000		
Cognitive proximity (3-digit IPC)	0.016***	0.016***	0.011**	0.003	0.030***	0.018***	0.034***	0.043***	0.043***	0.054***	0.052***	0.989***	1.000	
Applicants share local area dummy	0.321***	0.323***	0.226***	0.065***	0.053***	0.043***	0.056***	0.369***	0.380***	0.102***	0.107***	0.036***	0.028***	1.000

Source: Own elaboration using KITES-PATSTAT.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Appendix C. Selected robustness checks

Tables C.1–C.4

Table C.1
Robustness checks: omitted variables, full sample, coinvention dummy.

	(1)	(2)	(3)	(4)
Geographical proximity	−0.0320*** (0.001)	−0.0339*** (0.001)	−0.0299*** (0.001)	−0.0383*** (0.001)
Cultural/ethnic proximity	−0.0158*** (0.001)	−0.0145*** (0.001)	−0.0157*** (0.001)	−0.0163*** (0.001)
Organisational proximity	0.120*** (0.001)	0.118*** (0.001)		0.123*** (0.001)
Organisational proximity (scaled)			0.0564*** (0.001)	
Applicants share local area				0.00671*** (0.001)
Observations	190,313	189,834	186,407	172,941
F	27.508	27.160	27.215	25.484
R^2	0.662	0.660	0.661	0.666

All models use time dummies and spells fixed effects for *i,j*. Robust standard errors in parentheses.
* $p<0.1$.
** $p<0.05$.
*** $p<0.01$.

Table C.2
Robustness checks: omitted variables, multiple patenting inventors sample, coinvention dummy.

	(1)	(2)	(3)	(4)	(5)	(6)
Geographical proximity	0.00383*** (0.001)	0.00533*** (0.001)	0.00387*** (0.001)	0.00382*** (0.001)	0.00297*** (0.001)	0.00383*** (0.001)
Cultural/ethnic proximity	−0.000420 (0.001)	−0.000384 (0.001)	−0.000432 (0.001)	−0.000431 (0.001)	−0.000349 (0.001)	−0.000274 (0.001)
Organisational proximity	0.0215*** (0.001)		0.0214*** (0.001)	0.0215*** (0.001)	0.0180*** (0.001)	0.0207*** (0.001)
Organisational proximity (scaled)		0.0110*** (0.001)				
Social proximity	0.0625*** (0.003)	0.0626*** (0.003)		0.0625*** (0.003)	0.0533*** (0.003)	0.0554*** (0.002)
Social proximity (three categories)			0.0290*** (0.001)			
Cognitive proximity	−0.00498*** (0.002)	−0.00498*** (0.002)	−0.00474** (0.002)		−0.00582*** (0.002)	−0.00338* (0.002)
Cognitive proximity (3-digit IPC)				−0.00898*** (0.002)		
Applicants share local area					0.00716*** (0.001)	
Observations	40,638	38,329	40,638	40,638	36,345	40,624
F	28.266	27.419	28.207	28.281	25.831	28.174
R^2	0.824	0.821	0.824	0.824	0.826	0.824

All models use time dummies and spells fixed effects for *i,j*. Robust standard errors in parentheses.
* $p<0.1$.
** $p<0.05$.
*** $p<0.01$.

Table C.3
Robustness checks: technology field*year controls.

Coinvention dummy	(1)	(2)
Geographic proximity	−0.0321*** (0.001)	0.00234 (0.001)
Cultural/ethnic proximity	−0.0158*** (0.001)	−0.00494*** (0.001)
Organisational proximity	0.120*** (0.001)	0.306*** (0.001)
Observations	190,313	190,313
F	27.593	8.394
R^2	0.663	0.404

All models use time dummies and spells fixed effects for *i,j*. Robust standard errors in parentheses.
* $p<0.1$.
** $p<0.05$.
*** $p<0.01$.

Table C.4
Robustness checks: non-linear estimator, reduced form model.

Coinvention counts	(1)	(2)	(3)
Geographic proximity	0.000539	0.490**	0.0166*
	(0.007)	(0.191)	(0.009)
Cultural/ethnic proximity	−0.00964***	−0.352***	−0.0119***
	(0.001)	(0.066)	(0.003)
Organisational proximity	0.0996***	3.215***	0.109***
	(0.014)	(0.366)	(0.010)
Observations	117,456	117,456	117,456
F-Statistic	22.622		
R^2	0.169		
Log-likelihood	−20280.048	−6766.321	
χ^2		257127.194	

All models use time dummies. Robust standard errors in parentheses.
Reduced form model with observables vectors, not spells fixed effects.
Column 1 gives OLS results. Column 2 gives negative binomial results. Column 3 gives average marginal effects for the negative binomial.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

Appendix D. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.respol.2015.07.003.

References

Abowd, J., Creecy, R., Kramarz, F., Technical Report 2002-06 2002. Computing Person and Firm Effects Using Linked Longitudinal Employer–Employee Data. U.S. Census Bureau, Washington, DC.

Ackerberg, D., Botticini, M., 2002. Endogenous matching and the empirical determinants of contract form. J. Polit. Econ. 110, 564–591.

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: evidence from U.S. universities, 1981–1999. Res. Policy 34, 259–285.

Agrawal, A., Cockburn, I., Galasso, A., Oettl, A., 2014. Why are some regions more innovative than others? The role of small firms in the presence of large labs. J. Urban Econ. 81, 149–165.

Agrawal, A., Cockburn, I., McHale, J., 2006. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. J. Econ. Geogr. 6, 571–591.

Agrawal, A., Kapur, D., McHale, J., 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. J. Urban Econ. 64, 258–269.

Andrews, M., Schank, T., Upward, R., 2006. Practical fixed-effects estimation methods for the three-way error-components model. Stata J. 6, 461–481.

Angrist, J., Pischke, J.-S., 2009. Mostly Harmless Econometrics. Princeton University Press, Princeton.

Azoulay, P., Ding, W., Stuart, T., 2007. The determinants of faculty patenting behaviour: demographics or opportunities? Academic science and entrepreneurship: dual engines of growth? J. Econ. Behav. Organ. 63, 599–623.

Boschma, R., 2005. Proximity and innovation: a critical assessment. Reg. Stud. 39, 61–74.

Boschma, R., Frenken, K., 2009. The Spatial Evolution of Innovation Networks: a proximity perspective. In: Boschma, R., Martin, R. (Eds.), Handbook of Evolutionary Economic Geography. Edward Elgar, Cheltenham.

Boschma, R., Minondo, A., Navarro, M., 2013. The emergence of new industries at the regional level in Spain: a proximity approach based on product relatedness. Econ. Geogr. 89, 29–51.

Brusoni, S., Giuri, P., Mariani, M., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes, W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Le Bas, C., Luzzi, A., Magazzini, L., Nesta, L., Nomaler, Ö., Palomeras, N., Patel, P., Romanelli, M., Verspagen, B., 2007. Inventors and invention processes in Europe: results from the PatVal-EU survey. Res. Policy 36, 1107–1127.

Cantwell, J., 2005. MNCs, local clustering and science–technology relationships. In: Santangelo, G. (Ed.), Technological Change and Economic Catch-Up: The Role of Science and Multinationals. Edward Elgar, Cheltenham, pp. 75–94.

Cassi, L., Plunket, A., 2010. The determinants of co-inventor tie formation: proximity and network dynamics. In: Papers in Evolutionary Economic Geography # 10.15. Utrecht University, Utrecht.

Crescenzi, R., Gagliardi, L., 2015. Moving People with Ideas – Innovation, Inter-regional Mobility and Firm Heterogeneity, Working Papers of the Taubman Centre, Harvard Kennedy School of Government, #1-2015.

Crescenzi, R., Gagliardi, L., Iammarino, S., 2015. Foreign multinationals and domestic innovation: intra-industry effects and firm heterogeneity. Res. Policy 44, 596–609.

Crescenzi, R., Gagliardi, L., Percoco, M., 2013. Social capital and the innovative performance of Italian provinces. Environ. Plan. A 45 (4), 908–929.

Crescenzi, R., Rodríguez-Pose, A., Storper, M., 2007. The territorial dynamics of innovation: a Europe–United States comparative analysis. J. Econ. Geogr. 7, 673–709.

D'Este, P., Guy, F., Iammarino, S., 2013. Shaping the formation of university–industry research collaborations: what type of proximity does really matter? J. Econ. Geogr. 13, 537–558.

D'Este, P., Iammarino, S., 2010. The spatial profile of university–business research partnerships. Pap. Reg. Sci. 89, 335–350.

Docquier, F., Rapoport, H., 2012. Globalization, brain drain, and development. J. Econ. Lit. 50, 681–730.

Evans, T.S., Lambiotte, R., Panzarasa, P., 2011. Community structure and patterns of scientific collaboration in business and management. Scientometrics 89, 381–396.

Fitjar, R.D., Rodríguez-Pose, A., 2011. When local interaction does not suffice: sources of firm innovation in urban Norway. Environ. Plan. A 43, 1248–1267.

Fitjar, R.D., Rodríguez-Pose, A., 2015. Nothing is in the Air. University of Stavanger.

Fleming, L., King, C., Juda, A.I., 2007. Small worlds and regional innovation. Organ. Sci. 18, 938–954.

Freeman, R.B., Ganguli, I., Murciano-Goroff, R., 2014. Why and Wherefore of Increased Scientific Collaboration. National Bureau of Economic Research Working Paper Series No. 19819, NBER, Cambridge, MA.

Giuri, P., Mariani, M., 2012. When distance disappears: inventors, education, and the locus of knowledge spillovers. Rev. Econ. Stat. 95, 449–463.

Glänzel, W., 2001. National characteristics in international scientific co-authorship relations. Scientometrics 51, 69–115.

Glänzel, W., Schubert, A., 2005. Analysing scientific networks through co-authorship. In: Moed, H., Glänzel, W., Schmoch, U. (Eds.), Handbook of Quantitative Science and Technology Research. Springer, Netherlands, pp. 257–276.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78, 1360–1380.

Griffith, R., Lee, S., Van Reenen, J., 2011. Is Distance Dying At Last? Falling Home Bias in Fixed Effects Models of Patent Citations, Cemmap Working Paper CWP18/11. The Institute for Fiscal Studies/Department of Economics, UCL, London.

Guimaraes, P., Portugal, P., 2010. A simple feasible alternative procedure to estimate models with high-dimensional fixed effects. Stata J. 10, 628–649.

Hall, B., Jaffe, A., Trajtenberg, M., 2001. The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. NBER, Cambridge, MA.

Jackson, M., 2006. The economics of social networks. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics: Ninth World Congress of the Econometric Society. Cambridge University Press, Cambridge.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Q. J. Econ. 108, 577–598.

Jones, B.F., 2009. The burden of knowledge and the "death of the renaissance man": is innovation getting harder? Rev. Econ. Stud. 76, 283–317.

Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-university research teams: shifting impact, geography, and stratification in science. Science 322, 1259–1262.

Kaiser, U., Kongsted, H.C., Rønde, T., IZA Discussion Paper 5654 2011. Labor Mobility, Social Network Effects, and Innovative Activity. IZA, Bonn.

Kapur, D., McHale, J., 2005. Sojourns and Software: internationally mobile human capital and high tech industry development in India, Ireland and Israel. In: Arora, A., Gambardella, A. (Eds.), From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, India, Ireland and Israel. OUP, Oxford, pp. 236–274.

Kerr, W., HBS Working Paper 09-003 2008. The Agglomeration of US Ethnic Inventors. Harvard Business School, Boston, MA.

Kerr, W., 2010. The agglomeration of US ethnic inventors. In: Glaeser, E. (Ed.), Agglomeration Economics. University of Chicago Press, Chicago, pp. 237–276.
Kerr, W., Lincoln, W., NBER Working Paper 15768 2010. The Supply Side of Innovation: H-1b Visa Reforms and US Ethnic Invention. NBER, Cambridge, MA.
Kerr, W.R., 2013. U.S. High-Skilled Immigration, Innovation, and Entrepreneurship: Empirical Approaches and Evidence. National Bureau of Economic Research Working Paper Series No. 19377, NBER, Cambridge, MA.
King, G., Zeng, L., 1999. Logistic Regression in Rare Events Data. Harvard University, Cambridge, MA.
Kogler, D.F., Rigby, D.L., Tucker, I., 2013. Mapping knowledge space and technological relatedness in US cities. Eur. Plan. Stud. 21 (9), 1374–1391.
Lakha, F., Gorman, D., Mateos, P., 2011. Name analysis to classify populations by ethnicity in public health: validation of Onomap in Scotland. Public Health 125, 688–696.
Laursen, K., Masciarelli, F., 2007. The Effect of regional Social Capital and External Knowledge Acquisition on Process and Product Innovation, DRUID Working Paper.
Lee, S., Bozeman, B., 2005. The impact of research collaboration on scientific productivity. Soc. Stud. Sci. 35, 673–702.
Lettl, C., Rost, K., von Wartburg, I., 2009. Why are some independent inventors ‘heroes’ and others ‘hobbyists’? The moderating role of technological diversity and specialization. Res. Policy 38.
Levin, D., Cross, R., 2004. The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. Manage. Sci. 50.
Leydesdorff, L., Etzkowitz, H., 1998. The triple helix as a model for innovation studies. Sci. Public Policy 25, 195–203.
Lissoni, F., Tarasconi, G., Sanditov, B., CESPRI Working Paper 181 2006. The KEINS Database on Academic Inventors: Methodology and Contents. Universita’ Bocconi, Milan.
Lobo, J., Strumsky, D., 2008. Metropolitan patenting, inventor agglomeration and social networks: a tale of two effects. J. Urban Econ. 63, 871–884.
Lychagin, S., Pinkse, J., Slade, M.E., Van Reenen, J., NBER Working Paper 16188 2010. Spillovers in Space: Does Geography Matter? NBER, Cambridge, MA.
Mateos, P., 2007. A review of name-based ethnicity classification methods and their potential in population studies. Pop. Space Place 13, 243–263.
Mateos, P., Longley, P., O’Sullivan, D., 2011. Ethnicity and population structure in personal naming networks. PLoS ONE 6, e22943.
Nathan, M., 2014. Same difference? Minority ethnic inventors, diversity and innovation in the UK. J. Econ. Geogr., http://dx.doi.org/10.1093/jeg/lbu006 (Online first, May 10).
OECD, 2009. OECD Patent Statistics Manual. OECD, Paris.
Paier, M., Scherngell, T., NEMO Working Paper #10 2008. Determinants of Collaboration in European R&D Networks: Empirical Evidence from a Binary Choice Model Perspective. ARC, Vienna.
Petersen, J., Longley, P., Gibin, M., Mateos, P., Atkinson, P., 2011. Names-based classification of accident and emergency department users. Health Place 17, 1162–1169.
Rabinow, J., 1976. Are patents needed? Idea J. Law Technol. 18, 19–35.
Rigby, D.L., 2013. Technological relatedness and knowledge space: entry and exit of US cities from patent classes. Reg. Stud., http://dx.doi.org/10.1080/00343404.2013.854878.
Ruef, M., 2002. Strong ties, weak ties and islands: structural and cultural predictors of organizational innovation. Ind. Corp. Change 11.
Saxenian, A.-L., Sabel, C., 2008. Venture capital in the ‘periphery’: the new Argonauts, global search and local institution-building. Econ. Geogr. 84, 379–394.
Schmidheiny, K., 2013. Binary Response Models. University of Basel, Basel.
Seaborn, T., 1979. Talking about the automat. The open channel. Softw. Pat. Inst. (IEEE Comput.) 12, 87–88.
Sedikes, C., Ariely, D., Olsen, N., 1999. Contextual and procedural determinants of partner selection: of asymmetric dominance and prominence. Soc. Cognit. 17, 118–139.
Seely Brown, J., Duguid, P., 2002. The Social Life of Information. Harvard Business School Press, Cambridge, MA.
Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. Manage. Sci. 51, 756–770.
Storper, M., Venables, A., 2004. Buzz: face to face contact and the urban economy. J. Econ. Geogr. 4, 351–370.
Thompson, P., Fox-Kean, M., 2005. Patent citations and the geography of knowledge spillovers: a reassessment. Am. Econ. Rev. 95, 450–460.
Torre, A., Rallet, A., 2005. Proximity and localization. Reg. Stud. 39, 47–59.
Von Hippel, E., 2005. Democratising Innovation. MIT Press, Cambridge, MA.
Yeung, H., 2009. Regional development and the competitive dynamics of global production networks: an East Asian perspective. Reg. Stud. 43, 325–351.