FEBRUARY 2021
BxWP2021-02

BORDEAUX ECONOMICS WORKING PAPERS
CAHIERS D'ECONOMIE DE BORDEAUX

The missing link: international migration in global clusters of innovation

Massimiliano Coda-Zabetta
Spatial Dynamics Lab – University College Dublin
GREThA UMR CNRS 5113 – Université de Bordeaux

Christian Chacua
GREThA UMR CNRS 5113 – Université de Bordeaux

Francesco Lissoni
GREThA UMR CNRS 5113 – Université de Bordeaux
ICRIOS – Università Bocconi, Milano
francesco.lissoni@u-bordeaux

Ernest Miguelez
GREThA UMR CNRS 5113 – Université de Bordeaux
AQR-IREA, University of Barcelona
ernest.miguelez@u-bordeaux.fr

Julio Raffo
Innovation Economics Section, WIPO

Deyun Yin
Harbin Institute of Technology, Shenzhen
Innovation Economics Section, WIPO

GREThA UMR CNRS 5113
Université de Bordeaux
Avenue Léon Duguit
33608 Pessac – France
Tel : +33 (0)5.56.84.25.75
http://gretha.u-bordeaux.fr/

LAREFI
Université de Bordeaux
Avenue Léon Duguit
33608 Pessac – France
Tel : +33 (0)5.56.84.25.37
http://larefi.u-bordeaux.fr/

Abstract

In this chapter we look at the global network of innovative agglomerations, with a focus on their degree of internationalization and on the actors behind it – particularly high-skilled migrants. Using worldwide patent and publication geo-localized data, we identify all Global Hotspots of Innovation (GIHs) and Niche Clusters (NCs) worldwide, and study their success as a function of their international connections. In particular, we compare organizational ones, such as international collaborations orchestrated by multinational firms' collaborations, to personal ones, which may derive from migration to/from the GIHs and NCs. We find a strong role of the latter, always comparable and sometimes larger than the former.

Keywords: patents, publications, agglomeration, internationalization, migration

JEL: O30, F20, F60

Disclaimer: The views expressed in this chapter are those of the authors and do not necessarily reflect the views of the World Intellectual Property Organization (WIPO) or its Member States.

ACKNOWLEDGEMENTS: We are indebted to the participants to the 5th Geography of Innovation Conference (Stavanger, January 2020) and the online workshop for the edited book on "Cross-border Innovation in a Changing World. Players, Places and Policies" (June 2020). Any mistake remains ours. Financial support from the French National Research Agency (TKC project – reference: ANR-17-CE26-0016) and from the World Intellectual Property Organization (WIPO) is gratefully acknowledged.

This is an extended version of the book chapter forthcoming in: Castellani D., A. Perri, V. Scalera, A. Zanfei (eds), Cross-border Innovation in a Changing World. Players, Places and Policies, Oxford University Press, due for publication in 2021.

To cite this paper: CODA-ZABETTA Massimiliano, CHACUA Christian, LISSONI Francesco, MIGUELEZ Ernest, RAFFO Julio, YIN Deyun (2021), The missing link: international migration in global clusters of innovation, In: Castellani. D., Perri, A., Scalera, V., Zanfei, A. (eds.) Cross-border Innovation in a Changing World. Oxford University Press

1. Introduction[1]

The decade running from the Great Recession of 2008-09 to the Covid-19 pandemic has been marked by public discourse on the end of globalization, or at least of its "second golden age" (James and James, 2009; Rugman, 2012). Innovation-wise, this age has not yet come to the end. It started in the 1990s and has spanned over the three decades, during which the number of countries involved in the production of knowledge has increased incessantly.

The globalization of innovation, however, has not gone hand in hand with a geographical dispersion of innovative activities within countries (Crescenzi *et al.*, 2020). Instead, in both incumbent and entrant countries a limited number of locations have grown disproportionately and come to dominate the respective national systems of innovation. In some of these locations, innovation also has internationalized, thanks mostly to the unbundling and relocation of R&D activities by multinational companies (MNCs) (Awate, Larsen and Mudambi, 2015).

Last but not least, the globalization of innovation has coincided with the rapid growth of the international migration of highly skilled people, especially those with education and/or jobs in Science, Technology, Engineering or Mathematics (STEM) (Kerr et al., 2016). Compared to other aspects of globalization, STEM migration has been less studied, although some evidence indicates that its contribution to both the agglomeration and the internationalization of innovative activities is not negligible (Crown, Faggian and Corcoran, 2020; Marino *et al.*, 2020). Migrants' preference for thick labor markets pushes those with STEM qualifications towards the most important innovation clusters, thus reinforcing their gravitational pull. But to the extent that migrants in different locations are bounded by social ties that resist physical distance, they can also strengthen the links between such locations, thus easing the flow of goods and services, capital, and knowledge. Yet, these links are missing from most analyses. With this chapter, we aim to put them at center-stage.

In particular, we propose to measure the extent to which different agglomerations of innovative activity worldwide are both internationalized and invested by STEM migration, and how much their performance gain from it.

We proceed as follows. First, we briefly recall the key findings of the literature on innovative agglomerations and their global connections, and in particular those that speak in favor of a special role for personal connections involving STEM migrants (section 2). Second, we present our data sources and methodology and, based on this, we both produce a map of innovative clusters worldwide and draw a first picture of their involvement in international migration,

[1] All the data used in this study, including the raw micro data on worldwide migrant inventors, can be accessed from the authors upon e-mail request.

international collaborations, or the presence of foreign MNCs (section 3). Third, we run a series of regressions to explore to what extent the success of specific clusters owes to such involvement (section 4). Section 5 concludes.

2. Migration and personal connections across clusters

The benefits of agglomeration for knowledge-based interactions and exchanges are widely acknowledged in the literature (Castellani, 2018). Both cities and regional clusters generate the "local buzz" leading to further concentration of innovation activities (Storper and Venables, 2004). At the same time, such activities once concentrated in the MNCs' home countries and closely guarded inside them, have been increasingly trusted to foreign-located labs and/or undertaken as part of international collaborations (von Zedtwitz and Gassmann, 2002; Hall, 2011). When knowledge is not available locally, MNCs can now easily tap into distant global centers of excellence to access the resources they need (Berman, Marino and Mudambi, 2019). When it comes to scientific activities, universities and public research organizations worldwide also undertake collaborative initiatives. As a result, many national innovation centers are now connected via the organizational channels put in place by the MNCs and other global players. Such channels, or "global pipelines", do not simply allow external knowledge to flow in and out the innovative agglomerations, but play a special role in breeding their success, by limiting knowledge inbreeding and the danger of technological 'lock-ins' (Bathelt, Malmberg and Maskell, 2004; Maskell, Bathelt and Malmberg, 2006),

Scholarly work on global pipelines has documented in depth their importance for local innovative success (Cantwell and Mudambi, 2011; Lorenzen and Mudambi, 2013; Turkina, Van Assche and Kali, 2016; Castellani, 2018; Turkina and Van Assche, 2018; Castellani *et al.*, 2019). The present book largely contributes to this topic too (Belderbos et al.; Cantwell and Marra; Phene and Santangelo; Nieto and Rodriguez).

Yet, global pipelines are not the only type of links between innovation centers worldwide. In fact, personal ties matter, too, whether strong, such as kinship or personal friendship, or weak, such as relationships loosely based on common educational background or cultural traits (Lorenzen and Mudambi, 2013). But the extent at which personal relationships contribute to knowledge globalization and the success of specific clusters has yet to be assessed, both in absolute terms and relative to global pipelines.

Conceptually, one should first and foremost define the mechanisms that allow residents in innovative agglomerations, and in particular STEM workers, to have relevant personal relations abroad. We suggest that international migration is an obvious candidate, for reasons that have to do with migrants' spatial mobility and the creation of international social networks that come with it, as well as with their importance for innovation activities.

International migrants represent today over 3% of the worldwide population, this share having grown incessantly since the 1970s (Ferrie and Hatton, 2014). If one could count also the number of persons now living in their home country, but with a migratory experience (whether as students or workers), the figure would be much higher.

Migrants' family and friendship ties both span across national boundaries and may be particularly dense in specific locations abroad and/or in the country of origin, due to chain migration and path dependency (Boyd, 1989; Castles, 2002). An emerging literature explores the way these ties sustain international flows of goods and services (Rauch and Trindade, 2002; Combes, Lafourcade and Mayer, 2005; Iranzo and Peri, 2009). Strong evidence suggests that migrants increase the intensive margins of trade, which implies that their cross-border social networks diffuse otherwise unavailable or costly information assets on the price, quality, and location of the traded items.

Migration rates for the tertiary educated are both higher than for the rest of the population and increase with further education (Artuc *et al.*, 2015). This is especially true when it comes to STEM qualifications (Freeman, 2010; Kerr et al., 2016). Far from moving exclusively along a South-North or East-West axis, migrant scientists and engineers circulate widely also between advanced economies (Auriol, 2010; Franzoni, Scellato and Paula, 2012). MNCs are among their top employers and they further contribute to migration when they dispatch their staff around the world to follow foreign operations (Kerr, 2018; Choudhury, 2020). Top universities worldwide also attract international students. Whether they eventually return to their home countries, or stay and enter the local labor market or entrepreneurial scene (or move forward to other destinations), migrants contribute to extend and densify the global web of personal connections.

Some recent research suggests that STEM workers' family and friendship ties are most likely to supplement or intermingle with educational or professional ones. Family-wise, some studies suggest that inventors' parents are on average wealthier and more educated than the national average, often hold a technical degree, and raised their families within innovation hubs, where they acquire social capital (Aghion *et al.*, 2018; Bell *et al.*, 2019). Profession-wise, networks of scientific authors and inventors have been extensively studied and found to exhibit distinctive "small world" properties (Uzzi et al., 2007). Geographically, these properties are reflected in the double role that such networks play in keeping knowledge flows highly localized (to the extent that most career moves occurs locally), while at the same time allowing physically distant nodes to access local knowledge resources (Almeida and Kogut, 1999; Agrawal, Cockburn and McHale, 2006; Breschi and Lissoni, 2009; Miguelez and Moreno, 2013). More specifically, migrants' connections are found to play a role both inside the migrants' destination countries and between countries of destination and origin (Agrawal, Kapur and McHale, 2008; Kerr, 2008; Agrawal *et al.*, 2011; Breschi, Lissoni and Miguelez, 2017; Bahar, Choudhury and Rapoport, 2020; Fackler, Giesing and Laurentsyeva, 2020; Miguelez and Temgoua, 2020).

Summing up, an increasingly abundant literature suggest STEM international migrants have the potential both to contribute to the agglomeration of innovative activities, to forge ties between specific locations, and to sustain their success. In what follows we put this intuition to test.

3. Agglomeration, migration, and internationalization

We use patent data for measuring both agglomeration and internationalization and, more originally, migration. While not uncommon, this data choice is often limited to the most important patent archives, namely those of the United States Patent and Trademark Office (USPTO) or the European Patent Office (EPO). These have the advantage of both being well organized and easily accessible, and to cover the most important technology markets. However, the rise of other important patent offices in emerging countries (e.g., South Korea and China), makes this choice increasingly biased, especially when it comes to measuring the intensity of patenting activity at the local level, as we wish to do. Hence, we consider all the patent offices covered by PatStat (Worldwide Patent Statistical Database; De Rassenfosse et al., 2014; Kang and Tarasconi, 2016), and in particular all the international patent families from 1976 to 2017. Roughly speaking, an international patent family is a set of patent applications filed by the same applicant on the same invention, but in different countries. This definition includes the first filings in the applicant's home country and their international extensions (Martinez, 2010).

By considering families, instead of individual patents, we avoid counting the same invention twice or more. By focusing on international ones, we also make sure to ignore the most trivial inventions, for which no foreign markets exist, and possibly no market at all.

We follow the literature and assume that inventive activities take place close to the inventors' addresses, as reported on patent documents, which we transform into geographical coordinates (on patent geo-localization, see: Bergquist et al., 2017; de Rassenfosse et al., 2019; Morrison et al., 2017). When the same patent document reports inventors from different areas, we count the patent as many times as the number of areas, rather than fractioning it. The peculiar characteristics of knowledge assets, namely their non-rivalry and indivisibility, supports this choice (Tubiana, Miguelez and Moreno, 2020).

When identifying knowledge agglomerations worldwide, we also make use of scientific publications data, from Web of Science (WoS), for the years 1998-2018 – details in Miguelez et al. (2019).

3.1 Global Innovation Hotspots and Niche Clusters

While both the economic and geographical literature stress the importance of agglomerations for innovative activities, their identification remains an open challenge, for at least two reasons.

First, the terminology varies, ranging from "tech clusters" (when wishing to stress their peculiarities vis-à-vis industrial clusters; Kerr and Robert-Nicoud, 2020) to "innovation hubs" (where the emphasis is placed on cities and knowledge exchanges between them; Nijkamp and Kourtit, 2013) and also "hotspots", which is more neutral and used interchangeably with the former two. Other approaches, instead, identify "global cities" based on their population and economic activity, and then move on to quantify the innovation they produce (Belderbos et al., in this book; and Castellani, 2018). In this chapter, while using often – for brevity – the generic term of "cluster", we draw a distinction between "Global Innovation Hotspots" (GIHs) and "Niche Clusters" (NCs) (full definitions below).

Second, every effort to map innovation agglomerations, especially at the international level, should not rely on fixed spatial boundaries, such as administrative or political units (Carlino and Kerr, 2015). This practice suffers of both a "modifiable area unit problem" (the unit size may vary across countries, thus making quantitative comparisons impossible) and a "border effect" problem (the unit boundaries may either cut across a cross-border agglomeration, or – in case of large units include two distinct agglomerations). For these reasons, based on the coordinates we assigned to each patent and publication, we apply a DBSCAN clustering algorithm to identify a multitude of agglomerations worldwide.[2] This methodology maximizes internal density (reciprocal proximity of the activities in a cluster) and external distance (distance between clusters). We fully describe it, also with examples, in Miguelez et al. (2019).

We first run our algorithm on the entire patent and publication dataset (1976-2015), irrespective of technological or disciplinary fields, and obtain our list of GIHs. These are large knowledge production centers, which correspond to the geographical areas with the highest density of innovation/science per square kilometer. By construction, they do not overlap and are internationally comparable.

We then re-consider all the patents and publications not yet assigned to any GIH, split them by technological fields and treat each field separately with the same algorithm. In this way we

[2] DBSCAN stands for "Density-Based Spatial Clustering of Applications with Noise". See: Ester et al. (1996).

identify a number of NCs, with high innovation/scientific density in selected fields. By construction, NCs do not overlap with GIHs, and they are internationally comparable only within their specific field(s).

With our preferred calibration of DBSCAN parameters, we obtain 174 GIHs and 313 NCs, distributed across 34 countries (out of 195 in our sample).[3] While GIHs usually coincide with large urban areas, not all such areas host a large GIH. GIHs in Beijing, London, Los Angeles, New York, Seoul, and Tokyo concentrate a large amount of patents and publications, while those in Buenos Aires, Delhi, Istanbul, Mexico City, Moscow, Sao Paulo, or Tehran do not. Other big cities, such as Cairo, Bangkok, Kolkata, and Chongqing, do not host any GIH, but one or more NCs. Finally, many others – such as Jakarta, Karachi or Manila – do not host any knowledge agglomeration at all.

Figure 1 shows the evolution of a few, selected GIHs or NCs, starting from the largest one in each country (Figure 1.a), followed by the time evolution of top-10 clusters (Figure 1.b). The figures show that large GIHs such as Tokyo, Osaka, New York or San Francisco permanently occupy the top positions. They also show the role of leading GIHs in emerging economies, such as Seoul, Beijing, Shenzhen-Hong Kong and, to a lesser extent, Bengaluru, catching up over time with the leading areas.

[Insert Figure 1 about here]

3.2 Migration

No official statistics exist on the international migration of STEM workers, especially at the local level. Following Kerr (2008) and Breschi et al. (2017) we produce our own statistics on migrant inventors. To do so, we exploit two pieces of information we find on patents, namely the name and surnames of inventors, and their addresses.

3.2.1 Immigrant inventors

We analyse names and surnames by means of the IBM's Global Name Recognition system (IBM-GNR). IBM-GNR couples each name or surname to all countries in which it appears, along with information on their frequencies within each country, expressed in percentiles. The raw information exploited by IBM-GNR comes from the US Immigration Authorities archives, which

[3] More details on the methodology to identify GIHs and NCs, as well as further analytical results, are presented in the Appendix.

recorded names and nationality of all entrants in the country throughout 1990s. Such records allow establishing the diffusion of a name or surname both within each country and across all countries worldwide. For diffusion within a country, consider for example the name Rajiv, which IBM-GNR reports to be associated to India, Great Britain, Sri Lanka, the Netherlands and a few other countries. While very frequent within India (it belongs to the 90th percentile of names, ranked according to their frequency), this name is not that frequent in Great Britain nor in Sri Lanka (where it belongs to the 50th percentile only), and not frequent at all in the Netherlands (10th percentile). Similar examples can be made for surnames, too.

This allows us to evaluate the frequency of each inventor's name and surname in the country where he/she resides (according to the address in the patent), and use it to assign him/her either a native or a migrant status. In particular, we classify as native each inventor whose name and/or surname belongs to the 90th frequency percentile in his/her country of residence. Following with the previous example, we consider any Indian resident inventor named Rajiv – no matter the surname – as an Indian native. However, had an inventor named Rajiv both a common English surname (such as Smith) and a British address, he would be considered (also) a British native. Every resident who is not classified as native is classified as migrant. Hence, any Rajiv Smith with a Sri Lankan or Dutch address, or more generally living outside India or an English-speaking country, will be treated as a migrant.

For inventors whose name and/or surname are relative infrequent (that is which never reach the 90th percentile in any country to which they are associated), we consider them as natives of the country with the highest frequency, as long as this coincides with the country of residence. For example, the Italian-resident inventor Rajiv Coda-Zabetta would be classified as an Italian native, because Coda-Zabetta is a rare surname in any country worldwide, but it is in Italy where it is most common.

While extremely transparent, our algorithm presents two main limitations, which create both false positives and false negatives (respectively: natives treated as migrants and vice versa).

First, it may classify as migrants the members of small, but longstanding ethnic or linguistic minorities. Following our example, this is very likely the case of Rajiv in Sri Lanka. Similarly, it confounds first- and second-generation migrants, as it may be case with many Rajiv (with Indian surnames) in Great Britain.

Second, it misses out migrants who move between countries with the same dominant language, for example British inventors in the US or German ones in Switzerland. Any Mark Cavendish from the Isle of Man filing a patent while in Boston will be classified as a US-native inventor, and – similarly – any Erik Zabel from Berlin and active in Basel will be considered Swiss.

Lesser limitations of our algorithm are as follows. First, a very limited number of inventors may be classified as natives of two or more countries. Second, as our dataset does not include a unique ID for inventors, our count of native and migrant inventors is actually a count of patent-inventor pairs, which could bias our proxies for migrants' presence, should they be systematically more or less prolific than natives (but the literature suggests this not the case; Hunt, 2013).

Still, we consider our data reliable enough to both capture some general migration trends and evaluate the role of migrant inventors in most locations.

Figure 2 shows that the worldwide weight of migrant inventors has increased incessantly since the late 1980s (dashed line). The US is both the most important destination country, and the country with the highest inventor immigration rate in the figure. But the trend is similar in the UK and, at a lower level, for Germany. Some English-speaking countries such as Canada and Australia (see Appendix) have trends and levels comparable to those of the US. The same applies to small, R&D intensive European countries such as Switzerland and the Netherlands.

Japan, with negligible immigration rates, stand at the opposite end of the spectrum. In between, we find developing countries such as India and China. They start with very high immigration levels, which later on decline incessantly. Most likely, the trend inversion is due to the prevalent nature of STEM immigration in such countries, which mostly consists of foreign researchers working at local MNCs' branches. These weighed considerably on the inventor population when the indigenous innovative activity was limited, and much less thereafter.[4]

[Insert Figure 2 about here]

Generally speaking, the migration rates of inventors tend to be higher in GIHs and NCs than outside them (figure 3). While the share of migrant inventors remained stable and around 5% in both types of areas until mid-1990s, it started to growth thereafter, which translates in large differences at the end of the period (15% vs 8% in, respectively, clusters and non-cluster areas).

[Insert Figure 3 about here]

[4] The Appendix section explores differences in immigration rates across technologies. Two fields stand out: Biopharma and Information & Communication Technologies (ICTs), both of which are closely associated to higher education and academic research. It also shows immigration rates across clusters within a selection of countries.

More importantly, there is considerable variation across locations. Figure 4 reports the immigration trends for the largest clusters in the top-patenting countries (plus India). These trends remind national ones, with Bengaluru, Seoul and Shenzhen following closely, respectively, the Indian, Korean and Chinese trends; and lines for San Jose – San Francisco, London and Frankfurt reminding of those for the US, UK and Germany. Levels, however, are different and especially for the latter cities, whose immigration rates are well above their national averages. As for Tokyo, it exhibits steadily low migration rate levels.

[Insert Figure 4 about here]

3.2.2 Emigrant inventors

While identifying migrants in specific host countries is relatively straightforward, things get trickier when trying to assign them to specific countries of origin. Still, this exercise would be particularly useful if we were to assess how countries, and clusters therein, benefit not only from immigration, but also from emigration, as discussed in section 2. In this sense, we will often refer to the stock of emigrants from a given country or locations as its "diaspora" (while imprecise, this term is both evocative and largely accepted in the migration literature; see Beine et al., 2011a, 2011b; Dufoix, 2008).

We proceed by reconsidering IBM-GNR cross-country distribution of names and surnames. This allows us to identify the stock of emigrant inventors from any possible country of origin, which we then we assign to specific clusters inside such country. Following up from one of the examples above, consider the inventor Rajiv Choudhary. According to IBM-GNR, the name Rajiv comes mostly from India (84% of the cases) followed by Great Britain (10%) and Sri Lanka, Trinidad and Tobago, Canada, Australia, and the Netherlands (all with 1%). The surname Choudhary can be found in India (77% of cases), Pakistan (16%), Great Britain (6%), and Trinidad and Tobago (1%). We will then consider any Rajiv Choudhary who appears on patents with an address outside these countries as a migrant, and assign his patents fractionally to all of them, based on the combined frequencies of his name and surname.[5]

Being the IBM-GNR sourced from the US Immigration Authorities, the US is never listed as country of origin. Therefore, we cannot include it in our analysis (we will come back to this in section 4 below).

[5] In the case of India, the fractional count will be: $fraction = (84 + 77)/200 = 0.805$. Similarly, it will be 0.083 for Great Britain, 0.078 for Pakistan, 0.01 for Trinidad and Tobago, and 0.005 for the remaining countries.

Panel 5.a in figure 5 reports the evolution of inventor diasporas for the countries with the most emigrant inventors. India stands out, followed by China and, quite distanced, the UK and several European countries. Panel 5.b shows instead the inventor emigration rate, which is defined as the stock of emigrant inventors over the sum of emigrant and native resident inventors of each country. Countries such as Russia and India have more than half of their inventors residing abroad. China had large emigration rates at the beginning of the period too, but they converged to the lower levels of other countries in the last years, mainly due to the dramatic increase of the resident inventors.

We do not report specific figures for clusters. By construction, they follow closely those of the countries in which they are located. Notice that, together with the absence of data for the US, this is the most important limitation of our measure.

[Insert Figure 5 about here]

3.3 Internationalization besides migration

3.3.1 International teams

We can measure internationalization through patent teams geographical dispersion, as observed when the inventors of the same patent have addresses in different countries (for a similar approach, see Belderbos et al., this book; and Castellani, 2018). For MNCs' patents, such dispersion may result from the distribution of R&D activities across laboratories in different countries; or by collaborations between the patent applicant and one foreign institutions. For sake of simplicity, and due to the impossibility to distinguish between these two cases, we will consider them jointly and refer generically to collaborations.

Figure 6 summarizes the main trends for international collaborations for the whole world (dashed line) and for a selection of countries (top-6 patentees plus India), starting in 1991. Few stylized facts emerge.

First, international co-invention increases over time, but not monotonically. In particular, the share of international co-invented patents grew up to 11% in 2010, and declined afterwards. In unreported results, we see that this is due to a notable slowdown for international co-inventions from 2010 onwards, vis-à-vis an incessant growth of the total number of patents.

Second, international collaboration trends differ widely across countries. East Asian countries – Japan, South Korea and China – exhibit the lowest collaboration levels. For Chinese patents, this is not the case before 2000, but it becomes so by 2011-2015, as a consequence of a spectacular downward trend. The US, Europe, and India stand at the opposite end, with rather high levels.[6]

Clearly, for patenting the main trends are dictated by corporate R&D: when it comes to developing countries, its home-based growth may simply go along with a decreasing dependence on MNCs' intra-group collaborations, which shows up in a decline of international co-inventorship.

[Insert Figure 6 about here]

3.3.2 Foreign applicants

Being intellectual property, patents may have an additional and specific internationalization dimension, which does not relate to the invention process, but to ownership. Especially when coinciding with MNCs, applicants may have R&D labs in several foreign clusters, while retaining their headquarters elsewhere (Picci, 2010; Thomson, de Rassenfosse and Webster, 2013). Such MNCs are the main builders and feeders of the global pipelines discussed in section 2. They also contribute to international STEM migration, which makes imperative for us to control for their presence in both GIHs and NCs. However, telling foreign firms apart from local ones on the basis of patent data is no easy task. In first instance, we consider as foreign all the patent applicants with an address outside the country of the GIH or NC under consideration. By doing so, however, we run the risk of treating as owned by foreign firms all patents that a country's MNCs assign, for strategic or fiscal reason, to a foreign unit, possibly a patent box. To correct for this error, we treat as locally-owned all patents by applicants with address in a foreign country, but less of 10% of its patents produced by inventors in such country.

Figure 7 shows that India stands out as the country with more foreign penetration, as measured by the share of locally invented patents held by foreign companies. By the end of our period of analysis, almost 60% of patents with inventors in India have at least one foreign applicant. In contrast, China exhibits similarly large shares only at the start of the observation period: as soon as its inventive activity starts growing, its foreign share goes down to around 15%. This level is the same as that of developed countries such as the US or Germany – which however reach it by following an opposite trend. The UK is also an interesting case, as it has dramatically increased its foreign penetration over time, which is presently above 30%. At the other extreme,

[6] A more detailed analysis of this phenomenon at the cluster level is presented in the Appendix.

both Japan and South Korea exhibit very small foreign shares, with no changes over time. Clusters within each country follow patterns similar to the national ones.[7]

[Insert Figure 7 about here]

4. Clusters' performance: the role of migration and internationalization

This section concerns the relationship between international migration and innovative performance at the cluster level, by technological field and after controlling for other forms of internationalization. We produce several panel data regressions, with cluster-technology pairs as observations. This implies that any two observations may have in common the same GIH or NC, but refer to different technologies, or vice versa. In order to eliminate outliers, in particular very small cluster-technology pairs, we consider only the GIHs and NCs ranked among the first 100 clusters in each technology, for the time window 2011-2015.

For regressions including immigration variables, we drop around five hundred observations, which correspond to countries whose stocks of immigrants is hard to define, due to their linguistic proximity to their migrants' most relevant home countries. In most cases, these are small and highly innovative European countries such as Belgium (which draws most migrant inventors from France), Switzerland (which depends heavily on German migrants) and Ireland (which attracts migrants from all English-speaking countries worldwide). For regressions including emigration variables we do not consider the US clusters, for the reasons we put forward in section 3.2.2.

We measure clusters' success in two ways. First, we look at their productivity, for which we produce two proxies: the log number of patents per capita and the log number of citation-weighted patents per capita (3 years forward citations).[8] Second, we look at the clusters' position in the global network of cities and regions. This is a graph with clusters as nodes and co-inventorship activities (at least one patent in the period co-signed by inventors in two nodes) as ties. We rely on both the IB literature and social network analysis and consider more central clusters as better connected (Turkina, Van Assche and Kali, 2016; Turkina and Van Assche, 2018). We compute both degree centrality (the direct number of linkages that a cluster has with all other clusters) and eigenvector centrality (which gives a higher weight to clusters that are

[7] A more detailed analysis is presented in the Appendix.

[8] Due to the potential existence of zeros, we do not strictly log transform our variables, but apply the inverse hyperbolic sine transformation, which is defined as follows: $IHS = \log(y + (y^2 + 1)^{1/2}$.

central in the network, and therefore have access to key knowledge from anywhere in the network) (Turkina and Van Assche, 2018).

Our sample covers the period 1986-2015. We organize the data by 5-year, non-overlapping time windows. In order to deal with reverse causality concerns we time lag all the explanatory variables by one time window (that is, we use data from 1981 to 2010 to build our explanatory variables). Late comers to the international patent system, such as China and India, enter the panel only after 2000.[9]

As for the independent variables, our focal one is *Migrants*, which we measure as the percentage of patent-inventor pairs due to immigrant inventors, in each cluster-technology observation (see section 3.2.1).

Alternatively, we focus our attention on clusters' *Diasporas*, which we measure as the stock of emigrant inventors (fractional count as described in section 3.2.2). As this variable is originally computed at the country-technology level, we use the weight of the cluster in its country's total patenting to assign emigrants to specific clusters-technology.

Our main controls are *International teams* and *Foreign applicants*, which we measure, respectively, as described in sections 3.3.1 and 3.3.2.[10]

Finally, we deal with unobserved sources of endogeneity by inserting *technology-cluster* fixed effects, as well as *time-cluster* and *time-technology* interactions, which allows us to control for a wide range of confounding factors.[11]

Table 1 reports our baseline results on the relationship between immigration and cluster's performance, each column corresponding to a different dependent variable. The estimated coefficient for *Migrants* is positive and significant in all cases, except for eigenvector centrality, for which it is null. Regressions in columns (1) and (2) are log-linear and all explanatory variables are expressed in ratios, so we can read the coefficients as elasticities. This implies that an additional percentage point in the share of immigrants' inventor-patent pairs increases of 0.07% the number of per-capita patents in the average cluster-technology pair (column 1); and that the effect goes to 0.12% when patent counts are citation-weighted (column 2). As for the regression in columns (3), which is linear, the estimated coefficient suggests than a percentage

[9] More precisely, we consider as late comers all countries that adopted Triad-like patent legislations only after signing TRIPs, agreements coming with the establishment of the World Trade Organizations, such as China, India, Russia, and Brasil. Among such countries, those hosting clusters that enter our regression sample are China and India, the former having signed TRIPs in 2001 and the latter having signed it in 1995, but implemented only in 2000.

[10] Summary statistics and the correlation matrix are presented in the Appendix.

[11] We are aware of the lack of controls in our regressions. Yet, our FEs scheme account for the majority of variables that could affect productivity and centrality, including agglomeration economies, human capital accumulation, technology specific shocks, and other institutional and environmental factors.

point increase in the share of immigrants' inventor-patent pairs would increase the cluster's degree centrality score of 0.006 normalized points, that is 10% of its standard deviation.

These effects are comparable, from a quantitative viewpoint, with those pertaining to other forms of internationalization, namely international collaborations. In particular, the estimated coefficient for *International teams* is positive and significant in all regressions, except one (with patents per capita as dependent variable), with similar coefficients in columns (2) and (3).

As for foreign ownership, this is unrelated to cluster centrality (the coefficient for *Foreign applicants* is null in both columns 3 and 4), and negatively related to productivity (the coefficient for *Foreign applicants* is negative and significant in both columns 3 and 4). This confirms the positive association between the strength of domestic firms' presence and the cluster's success, which we suggested in section 3.3.

[Insert Table 1 about here]

These results, however, hide important heterogeneities across clusters, depending on their size and geographical locations. Table 2 explores them. Panel 2a in the table reports the results of split estimations of the coefficient for *Migrants*, according to the clusters' size. In particular, we distinguish between the top-25 ranked clusters in each technology, and the bottom-75, and find that it is only the coefficient for the former that is positive, significant and much larger than in the baseline regressions.

Based on our descriptive analysis, we also know that migrant inventors are mostly found in English-speaking countries and, to a lesser extent, in Continental Europe. For this reason, in panel 2b of table 2 we report split coefficients for cluster-technology pairs in English-speaking countries (US, UK, Australia and Canada), South-East Asian ones (India, China, Japan, South Korea, and Singapore), and the rest of the world (RoW) (mostly Continental Europe, plus a few clusters in Israel and Russia). With the (weak) exception of the regression for citation-weighted productivity, the estimated coefficient for South-East Asian countries is never significant, while evidence for other areas is patchier (positive and significant for one measure of productivity and one measure for centrality for English-speaking countries; positive and significant for both productivity measures, but not centrality, for the RoW).

When we split the coefficients according to both clusters' size and geography (panel 2c), we find that the positive relationship between inventors' migration and productivity is almost due entirely to the top 25 English-speaking clusters, and the top RoW ones, with very high elasticity estimates for both measures of productivity and for degree centrality. As for South-East Asia, we find some positive and relatively less significant effects for non-top clusters in two out of the four regressions.

Caution is due when it comes to causal interpretation of these results. Based on our efforts to accurately specify the model, we consider them as providing evidence of the positive effect of STEM migration on clusters' performance, but only for top clusters located in migrants' main destination countries. But some reverse interpretation may hold as well, which would point at all the other clusters' difficulties to attract foreign inventors.

[Insert Table 2 about here]

Table 3 examines the impact of STEM emigration (*Diasporas*) on clusters' productivity and centrality. Similar to the previous table, its first panel (3a) reports the baselines estimates, while the following ones (3b and 3c) report split coefficients for, respectively, clusters of different size and in different geographical areas. We first notice that the estimated coefficient for the variable of interest is always positive and significant in all panels, which implies a positive effect of emigration on all clusters in our sample. When examining split coefficients, the effects are stronger for the larger clusters. They are instead as strong, if not weaker for clusters in English-speaking countries relative to both those in South-East Asia and those in RoW (English-speaking countries, however, do not include the US, for the reasons explained in section 3.2.2). Overall, the estimated coefficients are slightly larger than for immigration, although direct comparisons should be treated with caution, due to the different ways in which the *Migrants* and *Diasporas* are calculated, as well as the differences in the samples used to obtain them.

These results suggest that both immigration and emigration can sustain productivity in clusters, but with effects much more asymmetric (across clusters) for immigration than for emigration. The tentative nature of our *Diasporas* variable and the loss of the US from our sample, however, suggest caution when interpreting our results and clearly call for further research on this issue.

[Insert Table 3 about here]

5. Conclusions

Globalization of innovative activities, which has grown incessantly over the past quarter century, has gone hand in hand with the emergence of local innovative clusters. They may be highly diversified GIHs, located inside or next to large metropolitan areas of advanced countries or emerging economies, or more specialized NCs, most often found in smaller cities and/or in less advanced economies. Based on a rich set of patent and publication data, this chapter identifies both types of agglomerations and explores their connections.

Such connections consist of both organizational and personal ties. The former originate from international organizational arrangements largely orchestrated by MNCs, such as collaborations between different country divisions or with foreign firms or research institutions, which result in "global pipelines" channeling knowledge across locations in different countries. Personal ties result instead from the international mobility of highly skilled persons, in particular those with STEM qualifications. While organizational ties contribute to such mobility (as when MNCs dispatch their staff to foreign operations, and back) this is mostly the result of a general migration trend that has both accompanied and favored the international exchanges of goods and services, capital, and knowledge.

Global pipelines have attracted some substantial literature, much less so personal ties. So far, the latter have represented a "missing link" in studies on global innovation networks: this chapter has proposed a methodology, based on patent data, to make its measurement possible, and proposed some exploratory results.

First, we have identified GIHs and NCs on the sole basis of the patent and publication data and not on pre-defined administrative boundaries. We have also documented their increasing weight on worldwide patenting activity, and their heterogeneity in terms of both size, exposure to high-skilled migration, and dependence on global pipelines.

Second, we have selected the top cluster-technology pairs and explored – by means of panel data regressions - how the clusters' success (as measured by productivity and connectedness) is due to their organizational or personal ties worldwide. We have measured organizational ties with the international collaborations and foreign ownership of local inventive operations; but also personal ties, based on the weight of foreign inventors in such operations (immigrant inventors) and the share of inventors of local origin operating abroad (emigrant inventors, or "diaspora").

While confirming, with some qualifications, the importance of global pipelines, our results also point at the important role of migration-based ties. When immigration is considered, migrant inventors play a role in the success of the largest clusters in English-speaking countries and European ones (the former being the most important destination of STEM migrants worldwide and especially from Asia, the latter being the beneficiaries of the European free-movement area). However, non-top clusters and clusters in Asia neither attract migrant inventors, nor they appear to benefit much of their presence. Whether the two things go together, possibly because the migrants' presence is felt only above a certain threshold, is to be assessed by future research.

As for emigration, our measure of each cluster's diaspora is more tentative and we must exert caution when interpreting our results. Still, they point at a more diffuse role than immigration, with productivity and connectedness of clusters of all sizes and locations being positively affected by emigration.

Besides improving our diaspora measure, we expect that future research will extend the evidence to other indicators besides patents and produce more robust causal evidence on the migration-performance nexus of global clusters. Political shocks opening or closing national borders are frequent and some of them – both recent and less recent - have been already exploited to prove the migrants' role in diffusing knowledge (surveyed by Lissoni, 2018) or otherwise sustaining innovation in host and home countries (Kerr et al., 2016). However, most studies take place at the national level or, when they focus on cities and regions, do not explore their connections. Moving the analysis to both the local and global level is the next step.

Figure 1. Number of patents per year in selected GHIs, 1976-2015

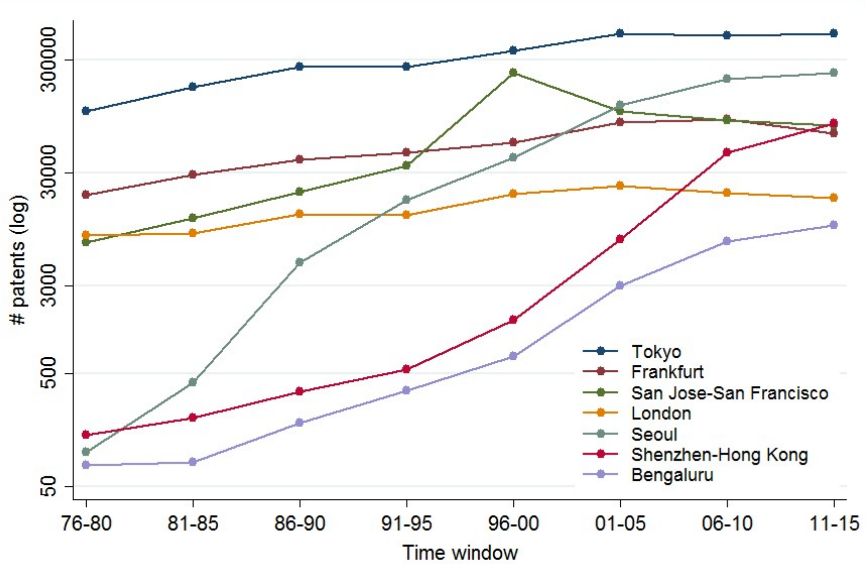


Figure 1.a. Top* clusters per country

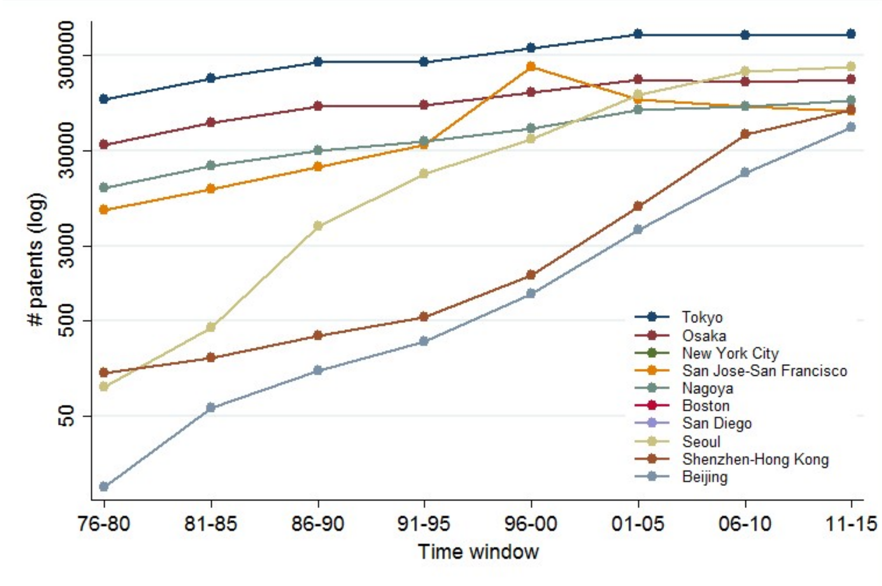


Figure 1.b. Top -10* clusters worldwide

* As per 2011-15 ranking

Figure 2. Share of immigrant inventors, total and by destination country; 1976-2015

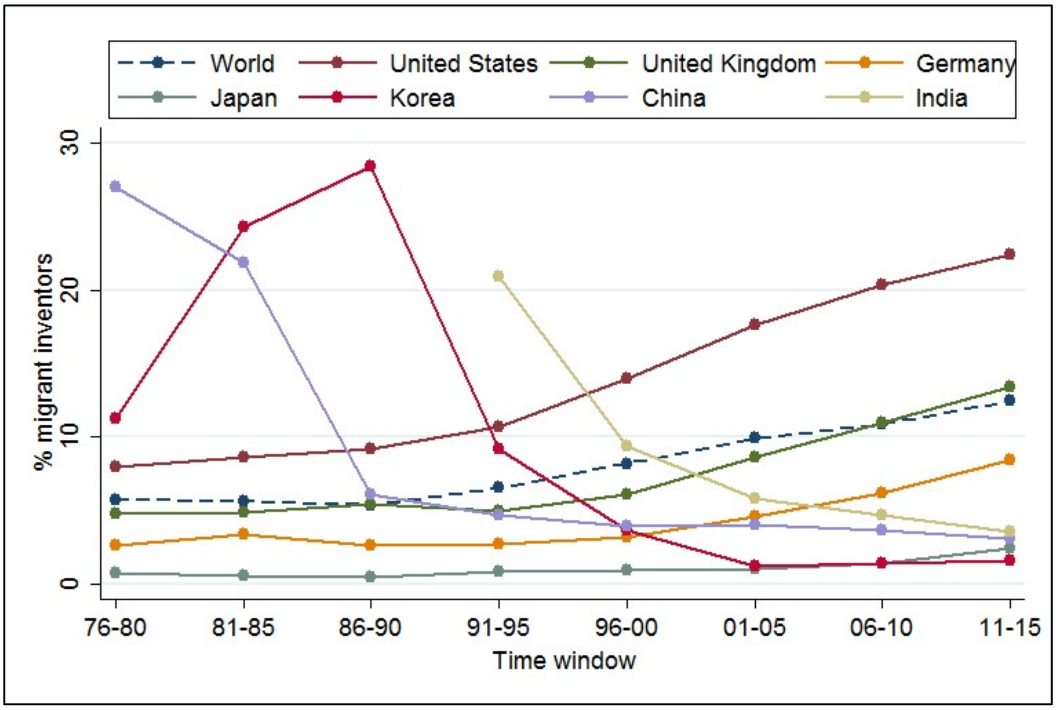


Figure 3. Share of immigrant inventors, inside and outside clusters worldwide; 1976-2015

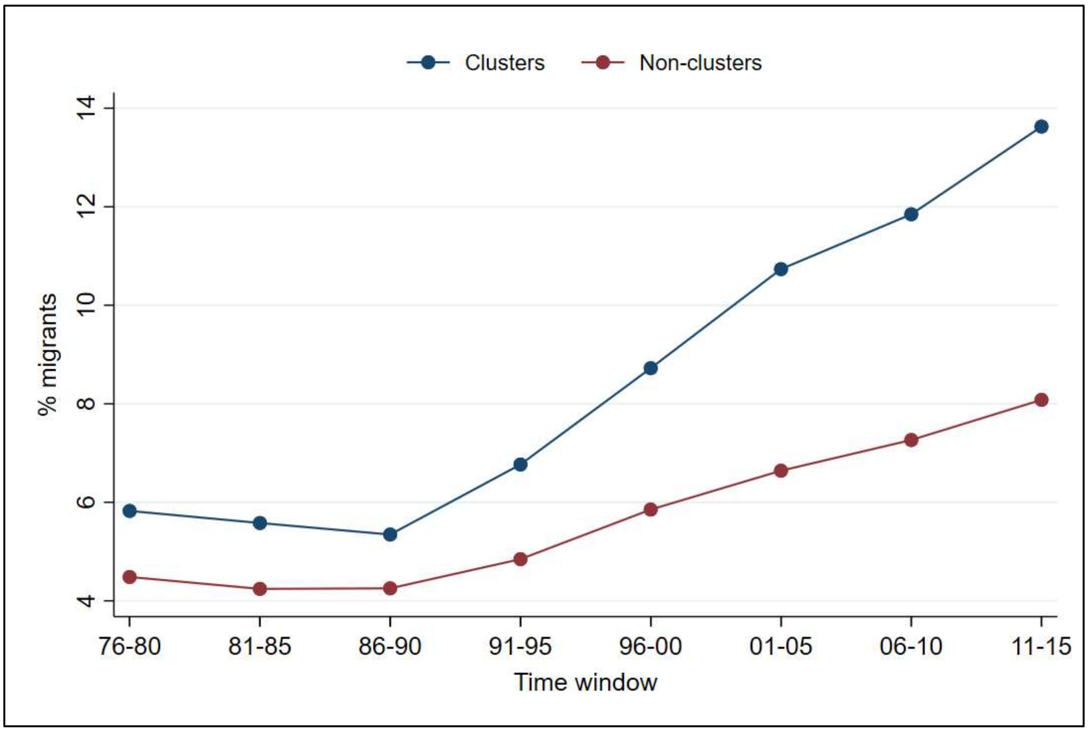
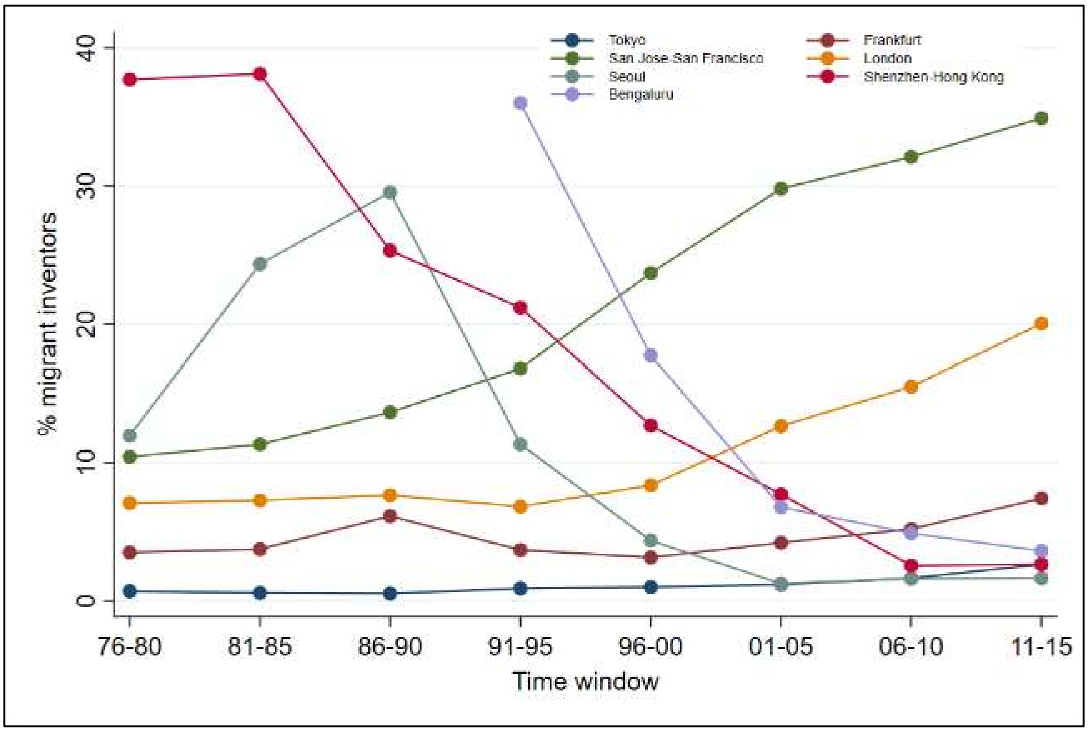


Figure 4. Share of immigrant inventors in top clusters*, selected countries; 1976-2015



* Top clusters are defined as either the GIH or the NC with the highest number of patents in its country, in 2011-15).

Figure 5. Emigrant inventors, by country of origin (selected countries);1976-2015

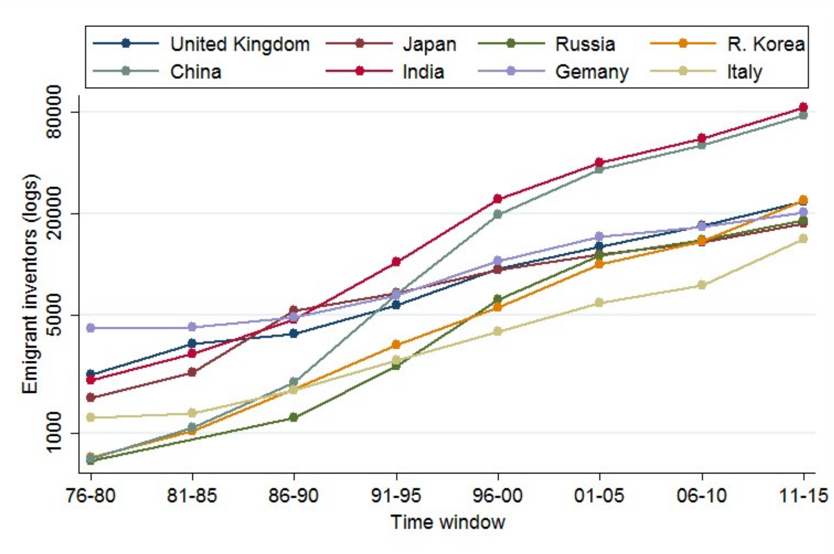


Figure 5.a: Stock of emigrant inventors

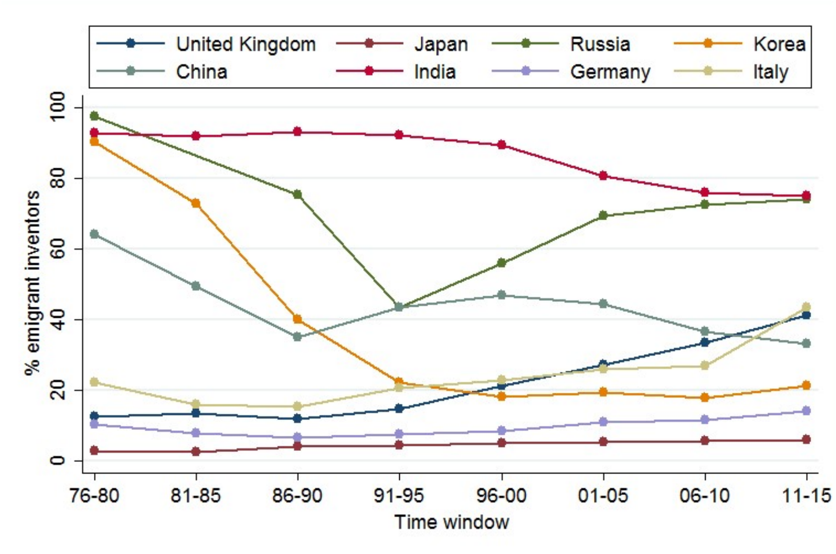


Figure 5.b: Inventor emigration rate

Figure 6. Share of patents by international inventor teams, total and selected countries; 1976-2015

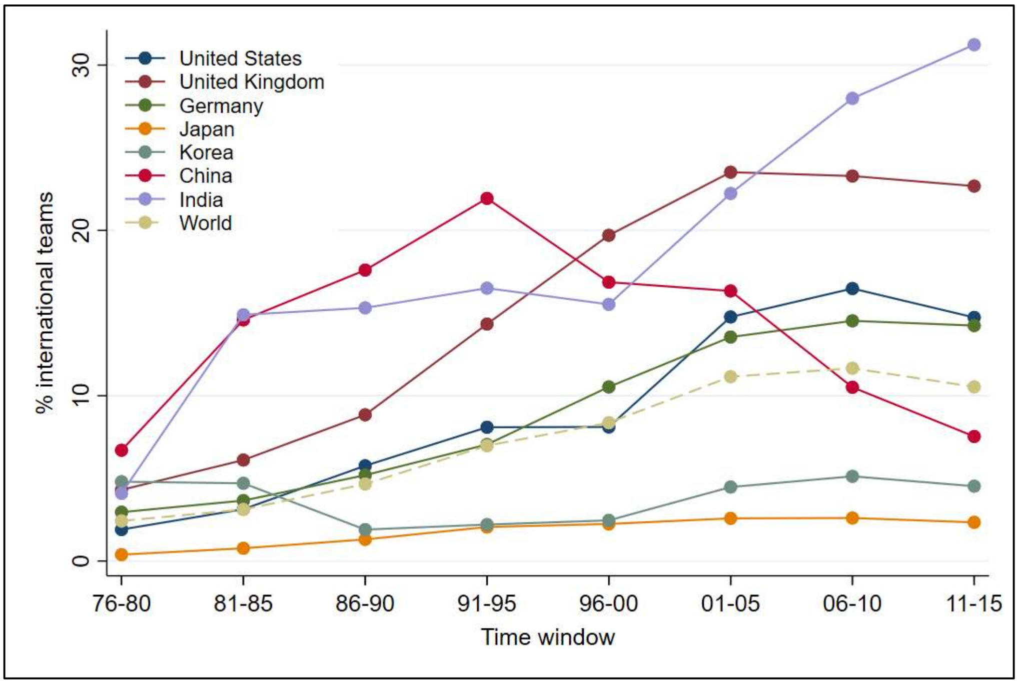
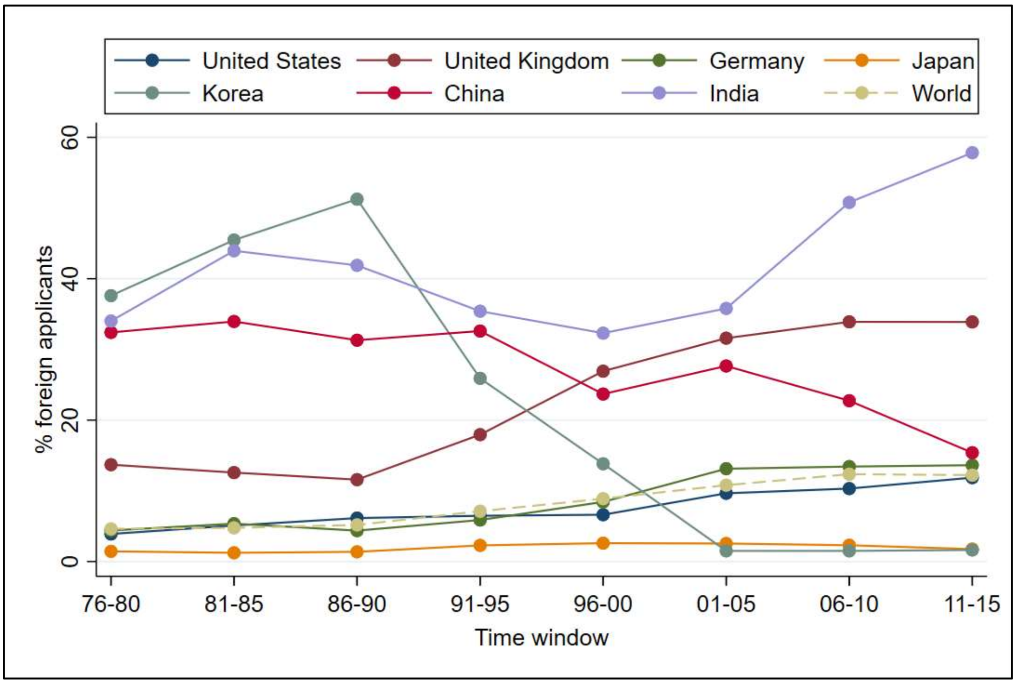


Figure 7. Share of patents by foreign applicants*, total and selected countries; 1976-2015



* All patents with at least a foreign applicant

Table 1. Effect of inventor immigration on clusters' success, 1986-2015

	(1) Patents p.c.	(2) Citation- weighted p.p.c.	(3) Degree centrality	(4) Eigenvector centrality
International teams	0.0459	0.111**	0.00848***	0.0163***
	(0.0308)	(0.0518)	(0.00293)	(0.00442)
Foreign applicants	-0.109***	-0.146***	0.00286	0.00562
	(0.0307)	(0.0519)	(0.00251)	(0.00344)
Migrants	0.0706**	0.119**	0.00612**	-0.00332
	(0.0326)	(0.0572)	(0.00269)	(0.00310)
Constant	0.448***	0.812***	0.0630***	0.0705***
	(0.00506)	(0.00875)	(0.000461)	(0.000648)
Observations	6,486	6,486	6,486	6,486
Adjusted R2	0.944	0.930	0.957	0.920
Cluster*Tech FE	Yes	Yes	Yes	Yes
Cluster*Time FE	Yes	Yes	Yes	Yes
Tech*Time FE	Yes	Yes	Yes	Yes

i. Observations consist of cluster-technology pairs observed over 5-year, non-overlapping time windows, from 1986 to 2015 (dependent variables) and from 1981 to 2010 (regressors).

ii. Patents p.c. stands for the number of patents per thousand inhabitants in the time window (inverse hyperbolic sine transformation). Citation-weighted p.p.c. stands for the number of patents per capita, weighted by their forward citations in a 3-year time window after patent priority year. Degree centrality and eigenvector centrality refer to a cluster position in the global innovation (0-to-1 normalized values).

iii. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 2. Effect of inventor immigration on clusters' success, 1986-2015; by cluster's rank and localization

	(1) Patents p.c.	(2) Citation- weighted p.p.c.	(3) Degree centrality	(4) Eigenvector centrality
		Panel 2.a		
Migrants * Top-25	0.330***	0.394***	0.0216***	-0.00449
	(0.105)	(0.136)	(0.00741)	(0.00701)
Migrants * non-top-25	-0.00499	0.0394	0.00161	-0.00298
	(0.0313)	(0.0639)	(0.00305)	(0.00331)
Constant	0.445***	0.810***	0.0628***	0.0705***
	(0.00563)	(0.00915)	(0.000476)	(0.000653)
Observations	6,486	6,486	6,486	6,486
Adjusted R2	0.944	0.930	0.957	0.920
		Panel 2.b		
Migrants * English	0.167**	0.0162	0.0210**	-0.0176
	(0.0758)	(0.173)	(0.00824)	(0.0120)
Migrants * South-E. Asia	0.0189	0.0978*	0.00210	-0.00205
	(0.0342)	(0.0589)	(0.00274)	(0.00228)
Migrants * RoW	0.243*	0.467**	0.00532	0.0153
	(0.139)	(0.183)	(0.00841)	(0.0113)
Constant	0.439***	0.811***	0.0622***	0.0709***
	(0.00653)	(0.0126)	(0.000634)	(0.000954)
Observations	6,486	6,486	6,486	6,486
Adjusted R2	0.944	0.930	0.957	0.920
		Panel 2.c		
Migrants * English (top25)	1.121***	1.187***	0.129***	-0.0231
	(0.209)	(0.233)	(0.0232)	(0.0292)
Migrants * English (non-top25)	-0.0515	-0.250	-0.00339	-0.0163
	(0.0690)	(0.190)	(0.00772)	(0.0124)
Migrants * S-E Asia (top25)	-0.0679	-0.00500	-0.00916	-0.00658
	(0.0961)	(0.134)	(0.00572)	(0.00502)
Mig. * S-E Asia (non-top25)	0.0457	0.130**	0.00558*	-0.000653
	(0.0299)	(0.0619)	(0.00304)	(0.00237)
Migrants * RoW (top25)	1.603***	1.660***	0.0411*	0.0402
	(0.351)	(0.457)	(0.0224)	(0.0286)
Migrants * RoW (non-top25)	-0.134	0.137	-0.00446	0.00830
	(0.128)	(0.177)	(0.00910)	(0.0116)
Constant	0.430***	0.801***	0.0612***	0.0710***
	(0.00682)	(0.0122)	(0.000698)	(0.00101)
Observations	6,486	6,486	6,486	6,486
Adjusted R2	0.945	0.931	0.957	0.920

Cluster*Tech FE	Yes	Yes	Yes	Yes
Cluster*Time FE	Yes	Yes	Yes	Yes
Tech*Time FE	Yes	Yes	Yes	Yes

i. Observations consist of cluster-technology pairs observed over 5-year, non-overlapping time windows, from 1986 to 2015 (dependent variables) and from 1981 to 2010 (regressors, which enter the regression with 1-period lags).
ii. Patents p.c. stands for the number of patents per thousand inhabitants in the time window (inverse hyperbolic sine transformation). Citation-weighted p.p.c. stands for the number of patents per capita, weighted by their forward citations in a 3-year time window after patent priority year. Degree centrality and eigenvector centrality refer to a cluster position in the global innovation (0-to-1 normalized values).
iii. All regressions also include the explanatory variables *International teams* and *Foreign applicants*.
iv. For the full list of top-25 clusters see the Appendix section. English includes Australia, Canada, Ireland, the UK and the US. South-East Asia includes China, India, Japan, South Korea and Singapore.
v. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table 3. Effect of inventor emigration (diasporas) on clusters' success, 1986-2015; by cluster's rank and localization

	(1) Patents p.c.	(2) Citation weighted p.p.c	(3) Degree centrality	(4) Eigenvector centrality
Panel 3.a				
Diasporas (ihs)	0.104***	0.0979***	0.00484***	0.00507***
	(0.0106)	(0.0158)	(0.000701)	(0.000840)
Constant	0.104**	0.324***	0.0285***	0.0208***
	(0.0409)	(0.0615)	(0.00268)	(0.00325)
Observations	4,596	4,596	4,596	4,596
Adjusted R2	0.958	0.931	0.946	0.861
Panel 3.b				
Diasporas * Top-25	0.208***	0.185***	0.0116***	0.00589***
	(0.0153)	(0.0217)	(0.00126)	(0.00133)
Diasporas * non-top-25	0.0785***	0.0762***	0.00316***	0.00487***
	(0.00983)	(0.0156)	(0.000677)	(0.000845)
Constant	0.0441	0.274***	0.0245***	0.0204***
	(0.0381)	(0.0601)	(0.00271)	(0.00334)
Observations	4,596	4,596	4,596	4,596
Adjusted R2	0.961	0.932	0.948	0.861
Panel 3.c				
Diasporas * English	0.0485***	0.105***	0.00941***	0.00428**
	(0.0142)	(0.0323)	(0.00156)	(0.00199)
Diasporas * S-E Asia	0.106***	0.0860***	0.00111	0.00191**
	(0.0145)	(0.0196)	(0.000753)	(0.000857)
Diasporas * RoW	0.120***	0.112***	0.00857***	0.00984***
	(0.0149)	(0.0205)	(0.00110)	(0.00122)
Constant	0.0953**	0.306***	0.0232***	0.0153***
	(0.0415)	(0.0625)	(0.00291)	(0.00349)
Observations	4,596	4,596	4,596	4,596
Adjusted R2	0.959	0.931	0.947	0.862

Cluster*Tech FE	Yes	Yes	Yes	Yes
Cluster*Time FE	Yes	Yes	Yes	Yes
Tech*Time FE	Yes	Yes	Yes	Yes

i. Observations consist of cluster-technology pairs observed over 5-year, non-overlapping time windows, from 1986 to 2015 (dependent variables) and from 1981 to 2010 (regressors, which enter the regression with 1-period lags).

ii. Patents p.c. stands for the number of patents per thousand inhabitants in the time window (inverse hyperbolic sine transformation). Citation-weighted p.p.c. stands for the number of patents per capita, weighted by their forward citations in a 3-year time window after patent priority year. Degree centrality and eigenvector centrality refer to a cluster position in the global innovation (0-to-1 normalized values).

iii. All regressions also include the explanatory variables *International teams* and *Foreign applicants*.

iv. For the full list of top-25 clusters see the Appendix section. English includes Australia, Canada, Ireland, the UK and the US. South-East Asia includes China, India, Japan, South Korea and Singapore.

v. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

.

References

Aghion, P. *et al.* (2018) 'On the returns to invention within firms: Evidence from Finland', in. *AEA Papers and Proceedings*: 208–12.

Agrawal, A. *et al.* (2011) 'Brain drain or brain bank? The impact of skilled emigration on poor-country innovation', *Journal of Urban Economics*, 69(1): 43–55.

Agrawal, A., Cockburn, I. and McHale, J. (2006) 'Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships', *Journal of Economic Geography*, 6(5): 571–591.

Agrawal, A., Kapur, D. and McHale, J. (2008) 'How do spatial and social proximity influence knowledge flows? Evidence from patent data', *Journal of Urban Economics*, 64(2): 258–269.

Almeida, P. and Kogut, B. (1999) 'Localization of knowledge and the mobility of engineers in regional networks', *Management science*, 45(7): 905–917.

Artuc, E. *et al.* (2015) 'A Global Assessment of Human Capital Mobility: The Role of Non-OECD Destinations', *World Development*, 65: 6–26.

Auriol, L. (2010) 'Careers of doctorate holders: employment and mobility patterns.', *OECD STI Working Paper. Paris: OECD Publishing.*

Awate, S., Larsen, M. M. and Mudambi, R. (2015) 'Accessing vs sourcing knowledge: A comparative study of R&D internationalization between emerging and advanced economy firms', *Journal of International Business Studies*, 46(1): 63–86.

Bahar, D., Choudhury, P. and Rapoport, H. (2020) 'Migrant inventors and the technological advantage of nations', *Research Policy*, p. 103947.

Bathelt, H., Malmberg, A. and Maskell, P. (2004) 'Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation', *Progress in Human Geography*, 28(1): 31–56.

Beine, M., Docquier, F. and Özden, Çaglar (2011) *Diaspora effects in international migration: key questions and methodological issues*. The World Bank. Policy Research Working Papers

Beine, M., Docquier, F. and Özden, Çağlar (2011) 'Diasporas', *Journal of Development Economics*, 95(1): 30–41.

Belderbos, R., Benoit, F., Edet, S., Lee, G., Riccaboni, M. (2021) Global Cities' Innovation Network. In: Castellani. D., Perri, A., Scalera, V., Zanfei, A. (eds.) *Cross-border Innovation in a Changing World. Oxford University Press*

Bell, A. *et al.* (2019) 'Who becomes an inventor in America? The importance of exposure to innovation', *The Quarterly Journal of Economics*, 134(2): 647–713.

Bergquist, K., Fink, C. and Raffo, J. (2017) 'Identifying and ranking the world's largest clusters of inventive activity', *The Global Innovation Index 2017: Innovation Feeding the World*: 161–76.

Berman, A., Marino, A. and Mudambi, R. (2019) 'The global connectivity of regional innovation systems in Italy: a core–periphery perspective', *Regional Studies* 54.5 (2020): 677-691

Boyd, M. (1989) 'Family and personal networks in international migration: recent developments and new agendas', *International Migration Review*, 23(3): 638–670.

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

Breschi, S., Lissoni, F. and Miguelez, E. (2017) 'Foreign-origin inventors in the USA: testing for diaspora and brain gain effects', *Journal of Economic Geography*, 17.5 (2017): 1009-1038.

Cantwell, J. A. and Mudambi, R. (2011) 'Physical attraction and the geography of knowledge sourcing in multinational enterprises', *Global Strategy Journal*, 1(3–4): 206–232.

Cantwell, J. and Marra, M. (2021) 'The Evolution of International Networks for Innovation'. In: Castellani. D., Perri, A., Scalera, V., Zanfei, A. (eds.) *Cross-border Innovation in a Changing World. Oxford University Press*

Carlino, G. and Kerr, W. R. (2015) 'Agglomeration and innovation', in *Handbook of regional and urban economics*. Elsevier: 349–404.

Castellani, D. (2018) 'The changing geography of innovation and the multinational enterprise'. In: Cook, G., Johns, J., McDonald, F., Beaverstock, J. and Pandit, N. (eds.) The Routledge Companion to International Business and Economic Geography. Routledge, Abingdon

Castellani, D. *et al.* (2019) 'Connectivity and the Location of MNEs Across the Value Chain. Evidence from US Metropolitan Areas', *Academy of Management Proceedings*, 2019(1), p. 12098.

Castles, S. (2002) 'Migration and community formation under conditions of globalization', *International migration review*, 36(4): 1143–1168.

Choudhury, P. R. (2020) ‘Intra-firm Geographic Mobility: Value Creation Mechanisms and Future Research Directions’, in *Employee Inter-and Intra-Firm Mobility*. Emerald Publishing Limited.

Combes, P.-P., Lafourcade, M. and Mayer, T. (2005) ‘The trade-creating effects of business and social networks: evidence from France’, *Journal of International Economics*, 66(1): 1–29.

Crescenzi, R. *et al.* (2020) ‘The geography of innovation and development: global spread and local hotspots’. *LSE Geography & Environment Discussion Paper*, (4).

Crown, D., Faggian, A. and Corcoran, J. (2020) ‘Foreign-Born graduates and innovation: Evidence from an Australian skilled visa program’, *Research Policy*, p. 103945.

De Rassenfosse, G., Dernis, H. and Boedt, G. (2014) ‘An introduction to the Patstat database with example queries’, *Australian economic review*, 47(3): 395–408.

Dufoix, S. (2008) ‘Diasporas'.Berkeley: University of California Press.

Ester, M. *et al.* (1996) ‘A density-based algorithm for discovering clusters in large spatial databases with noise.’, in Knowledge Discovery and Data Mining proceedings, 96(34),: 226–231.

Fackler, T. A., Giesing, Y. and Laurentsyeva, N. (2020) ‘Knowledge remittances: Does emigration foster innovation?’, *Research Policy*, 49(9): 103863.

Ferrie, J. P. and Hatton, T. J. (2014) ‘Two Centuries of International Migration’, in *In: Chiswick, B & Miller, P. (eds.) Handbook of the Economics of International Migration.* Elsevier. Amsterdam.

Franzoni, C., Scellato, G. and Paula, S. (2012) ‘Foreign-Born Scientists: Mobility Patterns for 16 Countries.’, *Nature Biotechnology*, (30): 1250–53.

Freeman, R. B. (2010) ‘Globalization of scientific and engineering talent: international mobility of students, workers, and ideas and the world economy’, *Economics of Innovation and New Technology*, 19(5): 393–406.

Hall, B. H. (2011) *The internationalization of R&D*. UNU-MERIT Working Paper Series 049. United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology.

Hunt, J. (2013) *Are Immigrants the Best and Brightest U.S. Engineers?* NBER Working Paper 18696. National Bureau of Economic Research, Inc.

Iranzo, S. and Peri, G. (2009) ‘Migration and trade: Theory with an application to the Eastern–Western European integration’, *Journal of International Economics*, 79(1): 1–19.

James, H. and James, H. (2009) *The end of globalization: lessons from the Great Depression*. Harvard University Press.

Kang, B. and Tarasconi, G. (2016) ‘PATSTAT revisited: Suggestions for better usage’, *World patent information*, 46: 56–63.

Kerr, S. P. *et al.* (2016) ‘Global Talent Flows’, *Journal of Economic Perspectives*, 30(4): 83–106.

Kerr, W. R. (2008) ‘Ethnic scientific communities and international technology diffusion’, *The Review of Economics and Statistics*, 90(3): 518–537.

Kerr, W. R. (2018) *The gift of global talent: How migration shapes business, economy & society*. Stanford University Press.

Kerr, W. and Robert-Nicoud, F. (2020) 'Tech clusters', *Journal of Economic Perspectives* 34.3 (2020): 50-76.

Lissoni, F. (2018) ‘International migration and innovation diffusion: an eclectic survey’, *Regional Studies*, 52(5): 702–714.

Lorenzen, M. and Mudambi, R. (2013) ‘Clusters, Connectivity and Catch-up: Bollywood and Bangalore in the Global Economy’, *Journal of Economic Geography*, 13(3): 501–534.

Marino, A. *et al.* (2020) ‘Ties that bind: Ethnic inventors in multinational enterprises’ knowledge integration and exploitation’, *Research Policy*, 49(9), p. 103956.

Martinez, C. (2010) *Insight into Different Types of Patent Families*. OECD Science, Technology and Industry Working Paper 2010/2. OECD Publishing.

Maskell, P., Bathelt, H. and Malmberg, A. (2006) ‘Building global knowledge pipelines: The role of temporary clusters’, *European Planning Studies*, 14(8): 997–1013.

Miguelez, E. *et al.* (2019) *Tied in: the Global Network of Local Innovation*. Groupe de Recherche en Economie Théorique et Appliquée (GREThA).

Miguelez, E. and Moreno, R. (2013) ‘Research networks and inventors’ mobility as drivers of innovation: evidence from Europe.’, *Regional Studies*, (47): 1668–1685.

Miguelez, E. and Temgoua, C. N. (2020) ‘Inventor migration and knowledge flows: A two-way communication channel?’, *Research Policy*, 49(9): 103914.

Morrison, G., Riccaboni, M. and Pammolli, F. (2017) ‘Disambiguation of patent inventors and assignees using high-resolution geolocation data’, *Scientific Data*, 4:201764.

Nieto, M.J. and Rodriguez, A. (2021) The locus of cross-border innovation: learning from external sources. In: Castellani. D., Perri, A., Scalera, V., Zanfei, A. (eds.) *Cross-border Innovation in a Changing World. Oxford University Press*

Nijkamp, P. and Kourtit, K. (2013) 'The "New Urban Europe": Global challenges and local responses in the urban century', *European Planning Studies*, 21(3): 291–315.

Phene, A. and Santangelo, G. D. (2021) 'Subsidiary technological scope and learning: A fresh perspective on the role of the subsidiary for the organization of MNC innovation'. In: Castellani. D., Perri, A., Scalera, V., Zanfei, A. (eds.) *Cross-border Innovation in a Changing World. Oxford University Press*

Picci, L. (2010) 'The internationalization of inventive activity: A gravity model using patent data', *Research Policy*, 39(8): 1070–1081.

de Rassenfosse, G., Kozak, J. and Seliger, F. (2019) 'Geocoding of worldwide patent data', *Scientific Data*, 6(1): 260.

Rauch, J. E. and Trindade, V. (2002) 'Ethnic Chinese Networks In International Trade', *The Review of Economics and Statistics*, 84(1): 116–130.

Rugman, A. (2012) *The end of globalization*. Random House.

Storper, M. and Venables, A. J. (2004) 'Buzz: face-to-face contact and the urban economy', *Journal of economic geography*, 4(4): 351–370.

Thomson, R., de Rassenfosse, G. and Webster, E. (2013) 'Quantifying the effect of R&D offshoring on industrial productivity at home.', *Mimeo, University of Melbourne.*

Tubiana, M., Miguelez, E. and Moreno, R. (2020) *In knowledge we trust: learning-by-interacting and the productivity of inventors*. 202013. University of Barcelona, Research Institute of Applied Economics.

Turkina, E. and Van Assche, A. (2018) 'Global connectedness and local innovation in industrial clusters', *Journal of International Business Studies*, 49(6): 706–728.

Turkina, E., Van Assche, A. and Kali, R. (2016) 'Structure and evolution of global cluster networks: evidence from the aerospace industry', *Journal of Economic Geography*, 16(6): 1211–1234.

Uzzi, B., Amaral, L. A. and Reed-Tsochas, F. (2007) 'Small-world networks and management science research: A review', *European Management Review*, 4(2): 77–91.

von Zedtwitz, M. and Gassmann, O. (2002) 'Market versus technology drive in R&D internationalization: four different patterns of managing research and development', *Research Policy*, 31(4): 569–588.

Appendixes

Appendix 1. Technological and scientific fields used in the analysis

For the purpose of the analysis, we grouped patents in 13 technological fields. These fields were based on the 35 fields of technology from WIPO's technology concordance table relying on the International Patent Classification (IPC) symbols.[12] The criteria to group the fields were: (1) to keep the resulting fields of a comparable size; and, (2) to group them according to the co-occurrence of WIPO's categories. The resulting fields range from 4.4% to 13.8% of total and group the 35 WIPO fields as follows:

1. **Electronics** (6.9%): Electrical machinery, apparatus, energy (1).
2. **Audio-visual** (4.4%): Audio-visual technology (2).
3. **ICTs** (13.8%): Telecommunications (3); Digital communication (4); Basic communication processes (5); Computer technology (6); and, IT methods for management (7).
4. **Semiconductors & optics** (7.3%): Semiconductors (8); and, Optics (9).
5. **Instruments** (10.8%): Measurement (10); Analysis of biological materials (11); Control (12); and, Medical technology (13).
6. **Biopharma** (7.6%): Organic fine chemistry (14); Biotechnology (15); Pharmaceuticals (16); and, Food chemistry (18).
7. **Materials** (4.9%): Materials, metallurgy (20); Surface technology, coating (21); and, Micro-structural and nano-technology (22).
8. **Chem & environment** (4.4%): Chemical engineering (23); and, Environmental technology (23).
9. **Chemicals** (8.5%): Macromolecular chemistry, polymers (17); Basic materials chemistry (19); and, Other special machines (29).
10. **Machines** (9.7%): Handling (25); Machine tools (26); and, Textile and paper machines (27).
11. **Engines & Transport** (12.3%): Engines, pumps, turbines (27); Thermal processes and apparatus (30); Mechanical elements (31); and, Transport (32).
12. **Civil engineering** (4.5%): Civil engineering (35).
13. **Consumer goods** (5.1%): Furniture, games (33); and, Other consumer goods (34).

Similar to patents, we also grouped the scientific publications in 12 scientific fields based on the subject tags to scientific publications in the Web of Science SCIE data. We based these fields in the existing categories by the *Observatoire de Sciences et Techniques* (OST) also with the criterion to group the publications in fields of a comparable size. The resulting fields range from 5.6% to 12.9% of total and group the 35 WIPO fields as follows:

1. **Applied Biology (7%)**: Plant Sciences; Veterinary Sciences; Agriculture; Zoology; Transplantation; Biology; Life Sciences & Biomedicine - Other Topics; Ecology; Entomology; Fisheries; Forestry; Agriculture, Dairy & Animal Science; Agronomy;

[12] www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf

Agriculture, Multidisciplinary; Mycology; Soil Science; Biodiversity & Conservation; Biodiversity Conservation; Horticulture; Agricultural Engineering; Materials Science, Textiles; Ornithology; Biochemistry & Molecular Biology.

2. **Biochem & Biotech (9.2%)** : Biochemistry & Molecular Biology; Cell Biology; Biotechnology & Applied Microbiology; Genetics & Heredity; Chemistry.
3. **Chemistry (12.9%)** : Chemistry; Chemistry, Multidisciplinary; Materials Science, Multidisciplinary; Chemistry, Physical; Polymer Science; Chemistry, Analytical; Chemistry, Organic; Electrochemistry; Nanoscience & Nanotechnology; Crystallography; Chemistry, Inorganic & Nuclear; Chemistry, Applied; Chemistry, Medicinal; Materials Science, Coatings & Films; Materials Science, Ceramics; Materials Science, Composites; Materials Science, Characterization & Testing; Materials Science, Paper & Wood; Oncology.
4. **Clinical Medicine (12%)** : Oncology; Radiology, Nuclear Medicine & Medical Imaging; Psychiatry; Clinical Neurology; Pediatrics; Medicine, General & Internal; Pathology; Dermatology; Toxicology; Health Care Sciences & Services; Rheumatology; Critical Care Medicine; Otorhinolaryngology; Allergy; Rehabilitation; Emergency Medicine; Tropical Medicine; Andrology; Environmental Sciences & Ecology.
5. **Earth Sciences (6%)** : Environmental Sciences & Ecology; Environmental Sciences; Geology; Marine & Freshwater Biology; Water Resources; Meteorology & Atmospheric Sciences; Geochemistry & Geophysics; Geosciences, Multidisciplinary; Oceanography; Engineering, Environmental; Paleontology; Mineralogy; Geography, Physical; Physical Geography; Engineering, Geological; Limnology; Engineering.
6. **Engineering (9.3%)** : Engineering; Energy & Fuels; Metallurgy & Metallurgical Engineering; Mechanics; Engineering, Chemical; Instruments & Instrumentation; Thermodynamics; Engineering, Mechanical; Engineering, Civil; Construction & Building Technology; Engineering, Biomedical; Engineering, Multidisciplinary; Engineering, Manufacturing; Engineering, Industrial; Transportation; Engineering, Aerospace; Mining & Mineral Processing; Engineering, Petroleum; Transportation Science & Technology; Engineering, Ocean; Engineering, Marine; Neurosciences & Neurology.
7. **Fundamental Biology (7.4%)** : Neurosciences & Neurology; Microbiology; Biophysics; Physiology; Reproductive Biology; Biochemical Research Methods; Virology; Evolutionary Biology; Developmental Biology; Mathematical & Computational Biology; Medical Laboratory Technology; Parasitology; Materials Science, Biomaterials; Anatomy & Morphology; Neuroimaging; Microscopy; Cell & Tissue Engineering; Immunology.
8. **Medical Science (7.7%)** : Immunology; Gastroenterology & Hepatology; Hematology; Respiratory System; Infectious Diseases; Medicine, Research & Experimental; Research & Experimental Medicine; Peripheral Vascular Disease; Physics.
9. **Physics & Math (9.4%)** : Physics; Physics, Applied; Optics; Physics, Condensed Matter; Physics, Multidisciplinary; Mathematics, Applied; Physics, Atomic, Molecular & Chemical; Spectroscopy; Physics, Particles & Fields; Physics, Mathematical; Statistics & Probability; Physics, Nuclear; Physics, Fluids & Plasmas; Mathematics, Interdisciplinary Applications; Public, Environmental & Occupational Health.

10. **Social & Human Sciences (6.3%)** : Public, Environmental & Occupational Health; Psychology; Nutrition & Dietetics; Sport Sciences; Nursing; Behavioral Sciences; Geriatrics & Gerontology; Business & Economics; Substance Abuse; Integrative & Complementary Medicine; History & Philosophy Of Science; Health Policy & Services; Education & Educational Research; Economics; Psychology, Experimental; Education, Scientific Disciplines; Anthropology; Audiology & Speech-Language Pathology; Gerontology; Psychology, Clinical; Psychology, Biological; Social Sciences - Other Topics; Primary Health Care; Management; Environmental Studies; Psychology, Multidisciplinary; Medical Ethics; Biomedical Social Sciences; Social Sciences, Biomedical; Legal Medicine; Medicine, Legal; Social Issues; Mathematical Methods In Social Sciences; Social Sciences, Mathematical Methods; Psychology, Developmental; Psychology, Applied; Linguistics; Ethics; Hospitality, Leisure, Sport & Tourism; Archaeology; Geography; Ergonomics; Agricultural Economics & Policy; Philosophy; Women's Studies; Social Sciences, Interdisciplinary; Urban Studies; History; Business; Sociology; Art; Government & Law; Law; Music; Psychology, Mathematical; Education, Special; Business, Finance; Communication; Family Studies; Social Work; Language & Linguistics; Ethnic Studies; Criminology & Penology; Psychology, Educational; Psychology, Psychoanalysis; History Of Social Sciences; Planning & Development; Public Administration; Religion; Arts & Humanities - Other Topics; Humanities, Multidisciplinary; Demography; Psychology, Social; International Relations; Industrial Relations & Labor; Literary Theory & Criticism; Literature; Surgery.
11. **Surgery (7.2%)** : Surgery; Urology & Nephrology; Cardiac & Cardiovascular Systems; Obstetrics & Gynecology; Ophthalmology; Orthopedics; Dentistry, Oral Surgery & Medicine; Anesthesiology; Science & Technology - Other Topics.
12. **Technology (5.6%)** : Science & Technology - Other Topics; Telecommunications; Nuclear Science & Technology; Automation & Control Systems; Operations Research & Management Science; Computer Science, Information Systems; Computer Science, Artificial Intelligence; Computer Science, Theory & Methods; Computer Science, Interdisciplinary Applications; Acoustics; Computer Science, Software Engineering; Imaging Science & Photographic Technology; Remote Sensing; Computer Science, Hardware & Architecture; Medical Informatics; Information Science & Library Science; Robotics; Green & Sustainable Science & Technology; Computer Science, Cybernetics; Logic; Architecture; .

Appendix 2. Identification of GIHs and NCs and main results

While the economic and geographical literatures abound of references to the importance of agglomerations for innovative activities, the identification of the location and boundaries of such agglomerations remains an open challenge. Even the terminology varies, ranging from "tech clusters" (when wishing to stress their peculiarities vis-à-vis industrial clusters; Kerr and Robert-Nicoud, 2019) to "innovation hubs" (where the emphasis is placed on innovative cities and their centrality in global networks of knowledge exchanges; Nijkamp and Kourtit, 2013). "Hotspot" is also a recurring term, which is more neutral and used interchangeably with "hub" and "cluster". In this chapter we use the generic form of "cluster" to refer to the spatial units we are going to identify. We will divide them, however, between "Global Innovation Hotspots" (GIHs) and "Niche Clusters" (NCs) (full definitions below). Moreover, when the technological dimension is attached to the spatial unit identified, we refer to them as "cluster-technology".

Every effort to map innovation agglomerations, especially at the international level, should not rely on fixed spatial boundaries, such as administrative or political units (Carlino and Kerr, 2015). Albeit very common, for opportunity reasons, this practice suffers of both a "modifiable area unit problem" (the unit size may vary across countries, thus making quantitative comparisons impossible) and a "border effect" problem (the unit boundaries may either cut across a cross-border agglomeration, or – in case of large units - include two distinct agglomerations). What we need instead is a continuous distance metric, one that first places innovation activities in space and then assigns them to agglomeration units that both maximize internal density (reciprocal proximity of the activities in the unit) and external distance (distance between units).

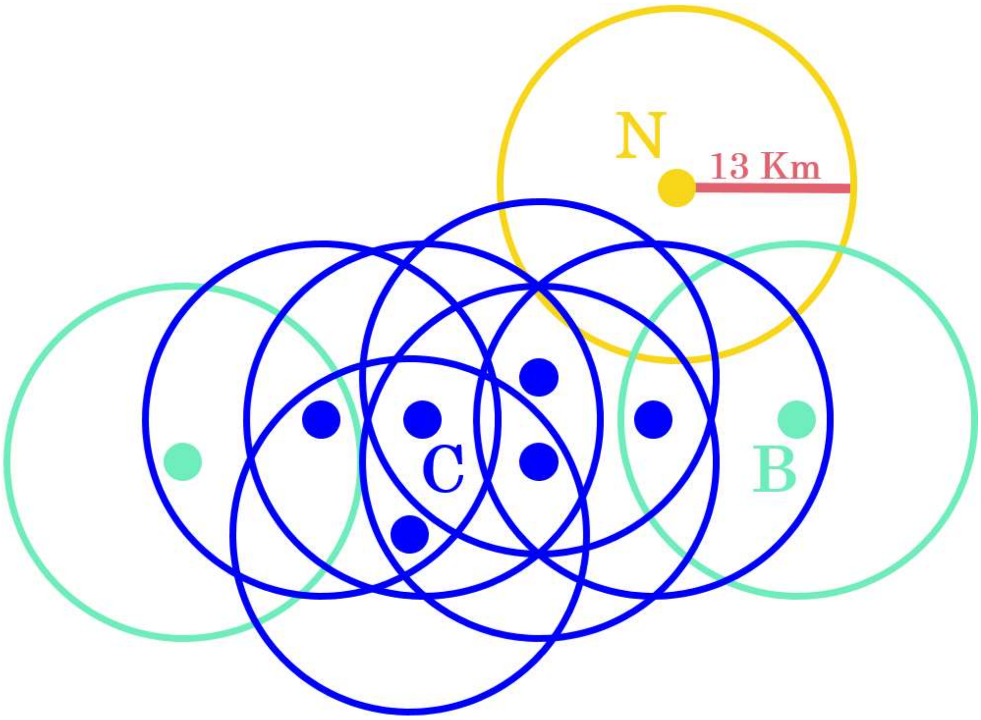
For these reasons, we first geolocalize with high precision all patents in our dataset, and then apply a DBSCAN clustering algorithm to identify a multitude of agglomerations worldwide (where DBSCAN stands for "Density-Based Spatial Clustering of Applications with Noise"; Ester et al., 1996). When available, the algorithm was applied to the inventors' addresses, following their geocoding; in the latter's absence, to the addresses of patent applicants. Geocoding of addresses relied on both previous efforts (de Rassenfosse et al., 2019; Ikeuchi et al., 2017; Li et al., 2014; Morrison et al., 2017; Yin and Motohashi, 2018) and our own work using ESRI or Geonames, as well as geocoded postal codes official national sources. Again, each patent is assigned to a coordinate point with a whole count. However, if different coordinate points belonging to the same patent are grouped in the same identified cluster, we counted all of them only once, as we did in our analysis at the country level.

In a nutshell, the algorithm consists in:

- considering all points in space containing one or more patents (patents with the same exact latitude and longitude);
- searching for other points in an arbitrarily small area around each point (in our case, a circle with a radius of 13 kilometres for patents, as in Bergquist et al., 2017; 23 kilometres for publication data), irrespective of any administrative boundary;
- and in aggregating all significantly overlapping areas up to when they contain only "core" and "border" points, the former being all the points surrounded by an arbitrarily high number of other points (critical threshold), the latter being all non-core points falling into a core point's area (so to exclude all "noise" points, which are neither core nor close to any core)

Figure A.1 displays an idealized map containing an agglomeration of points (where C is a representative core point and B a representative border point), plus a noise point (N) that does not belong to any agglomeration. Notice that the agglomeration results from having set the threshold at 4, while in reality we have set it equal to the median of the number of neighbors around each core point. Choosing as threshold the median of the number of neighbors, allows half of the points to be considered as core points by the clustering algorithm (more details available in Miguelez et al., 2019).

Figure A.1. DBSCAN core points, border points and noise points

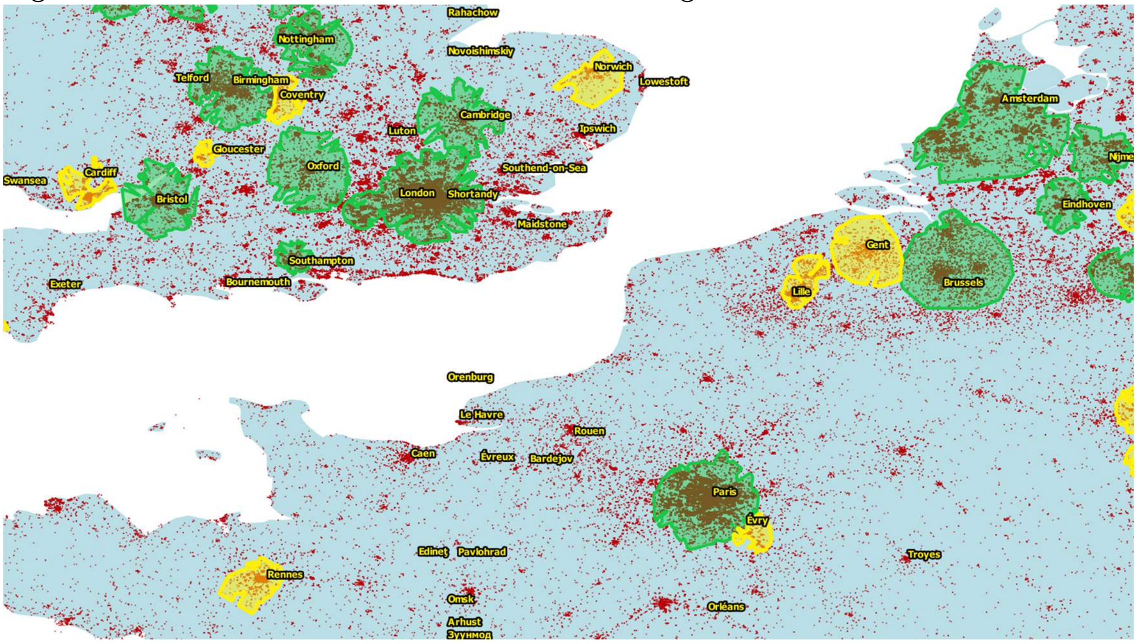


We first run our algorithms on the entire patent (1976-2015) and publication (1998-2018) datasets, irrespective technological or disciplinary fields, and identify a certain number of patents and publications agglomerations. After having identified the two types of agglomerations separately, we merge them and keep the outer borders in case that some patent and publication agglomerations do not overlap exactly, thus obtaining our list of "Global Innovation Hotspots" (GIHs). These are large knowledge production centers, capturing the geographical areas with the highest density of patents and publications per square kilometer. By construction, they do not overlap and are internationally comparable.

After identifying the GIHs, we re-consider all the patents not yet assigned to any of them, and treat them separately with the same algorithm, this time by technological and scientific field. We also set a lower critical threshold for the identification of core points – the median value of per-field distributions being lower than general ones. In this way we identify a number of "Niche Clusters" (NCs), with high innovation density in one or more specific technological and scientific fields. By construction, they do not overlap with the GIHs, but they are only internationally comparable within their specific field (or fields).

Different calibrations of the DBSCAN algorithms produce different numbers of GIHs and NCs. With our preferred calibration, we obtain 174 of the former and 313 of the latter, distributed across 34 countries (out of 195 in our sample). When ranked by either the number of patents or that of publications, all of the top 30 agglomerations in our ranking turn out to be GIHs. They are located in just 16 countries and are responsible for almost 70% of the patents and around 50% of the scientific articles published worldwide. More generally, very few patents and publications are produced outside our GIHs and NCs, and even fewer outside the countries that host them. Figure A.2 shows a zoomed-in example of the outcome of the cluster identification process (part of Western Europe and South UK). Red points identify geolocalized patenting activity, while green and yellow shades identify, respectively, GIHs' and NCs' borders.

Figure A.2. Outcome of the cluster identification algorithm



Red points identify geolocalized patenting activity, while green and yellow shades identify, respectively, GIHs' and NCs' borders

Despite being measured independently from administrative units, most GIHs and NCs fall within the largest and/or most prosperous urban areas of the world. Figure A.3 displays them on a worldwide map, and zooms-in a selection of geographical areas.

Figure A.3. Global Innovation hotspots and niche clusters

Red and green points identify, respectively, GIHs' and NCs' locations

Sources: Authors' estimation based on PATSTAT (for patent data)

Still, some less dense urban areas in high-income and innovative countries host a large number of important NCs. Table A.1 ranks both the top clusters by number of patents, for two exemplary technologies (Audio-visual and Biopharma). As expected, GIHs always come on top, but some NCs (shaded in grey) do better than many GIHs. For instance, the NC of Rennes (France) produce more patents in Audio-visual technologies than GIHs such as Amsterdam, Singapore, Ann Arbor, Stockholm, Cambridge or Copenhagen. The NCs in Bern or Mumbai do better in biopharma patents than GIHs such as Singapore, Taipei, Nurnberg, Rochester or even Bengaluru. The complete list of the top 50 GIHs or NCs for 12 distinct technological fields can be found in Miguelez et al. (2019).

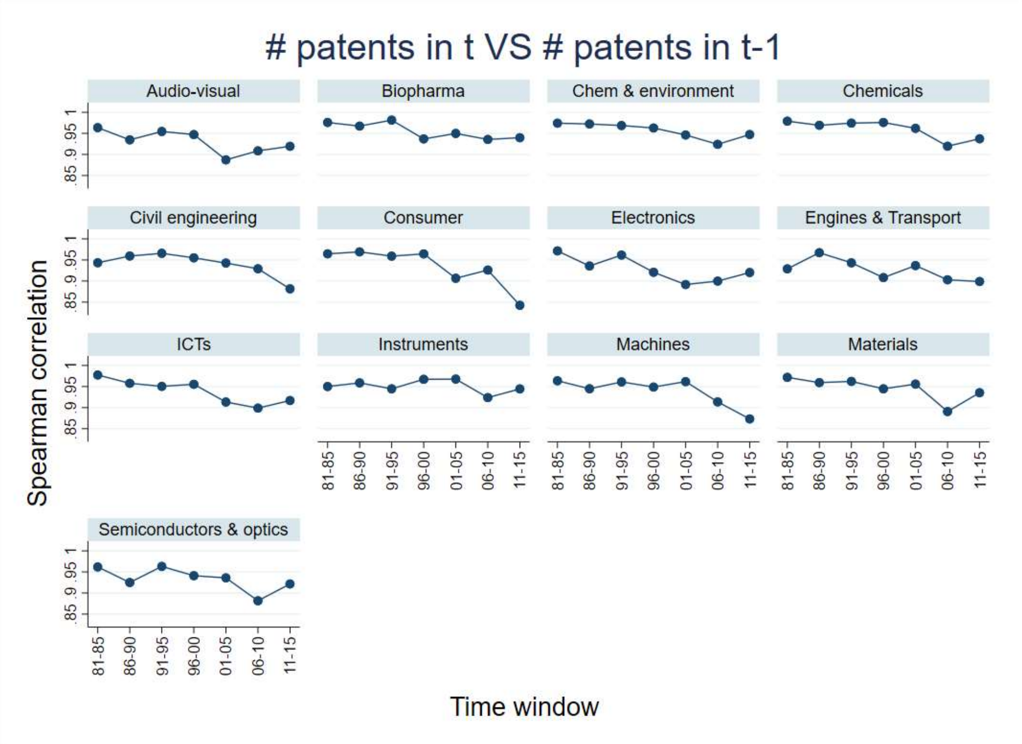
Table A.1: Ranking GIH and NC, for selected fields, patents

Audio-visual			Biopharma		
Position	Share patents	Cluster name	Position	Share patents	Cluster name
1	32.35%	Tokyo	1	9.82%	Tokyo
2	9.25%	Osaka	2	4.33%	Osaka
3	7.28%	Seoul	3	3.91%	San Jose-San Francisco
4	2.34%	Nagoya	4	3.88%	Mannheim
5	2.33%	San Jose-San Francisco	5	3.63%	New York City
6	2.11%	Shenzhen-Hong Kong	6	2.73%	Köln-Dusseldorf
7	1.42%	Taipei	7	2.60%	Paris
8	1.27%	New York City	8	2.35%	Boston
9	1.15%	Paris	9	2.25%	Seoul
....	....	....	....	....	....
44	0.29%	Rennes	74	0.25%	Bern
45	0.29%	Amsterdam	75	0.25%	Mumbai
46	0.29%	Ann Arbor	76	0.23%	Singapore
47	0.29%	Singapore	77	0.23%	Taipei
48	0.29%	Stockholm	78	0.22%	Bengaluru
49	0.27%	Cambridge	79	0.22%	Nürnberg
50	0.26%	Copenhagen	80	0.22%	Rochester
51	0.26%	Cheonan	81	0.22%	Beolgyo
52	0.25%	Basel	82	0.21%	Bridgeport
53	0.25%	Grenoble	83	0.21%	Vancouver
54	0.25%	Portland	84	0.21%	Okayama

Note: Grey-shaded lines indicate Niche Clusters. All other agglomerations are Global Innovation Hotspots
Source: Authors' estimation based on PATSTAT

The rankings of these clusters from period to period are relatively stable. This is shown in figure A.4, which reports the Spearman's rank correlation coefficients across time periods, per technology. The correlation from period to period is always very high (0.85-0.95), though it is never perfect. This is especially the case after 2000, when some reshuffling seems to take place, followed by renewed stability. This is possibly related with the entry of new actors. Thus, in unreported results we repeated Spearman's correlations per period with respect to the initial period 1976-1980, and correlations there become poorer and poorer over time, being around 0.4-0.6 for the last period.

Figure A.4. Clusters' ranking Spearman's rank correlation, over time

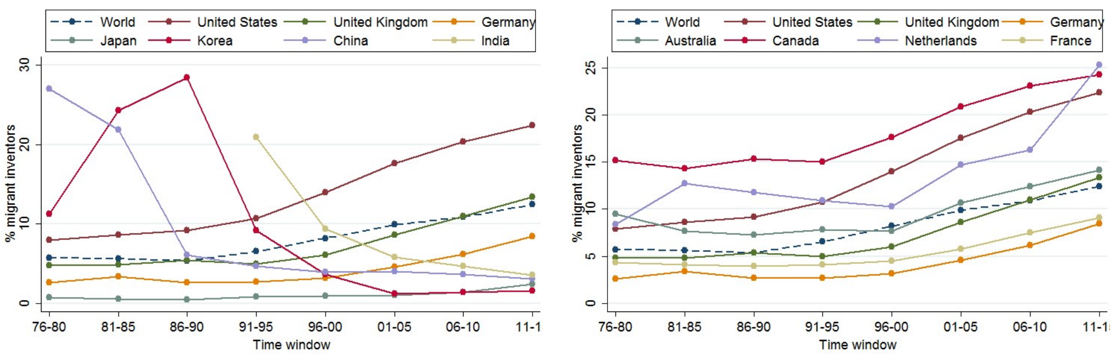


Appendix 3. The GNR-IBM migrant inventor dataset and comparisons to PCT-WIPO

Figure A.5 shows that the worldwide weight of migrant inventors has increased incessantly since the late 1980s (dotted line). The US is both the most important destination country, and the country with the highest inventor immigration rate in the figure. But the trend is similar in the UK and, at a lower level, for Germany. Some English-speaking countries such as Canada and Australia have trends and levels comparable to those of the US. The same applies to small, R&D intensive European countries such as Switzerland and the Netherlands.

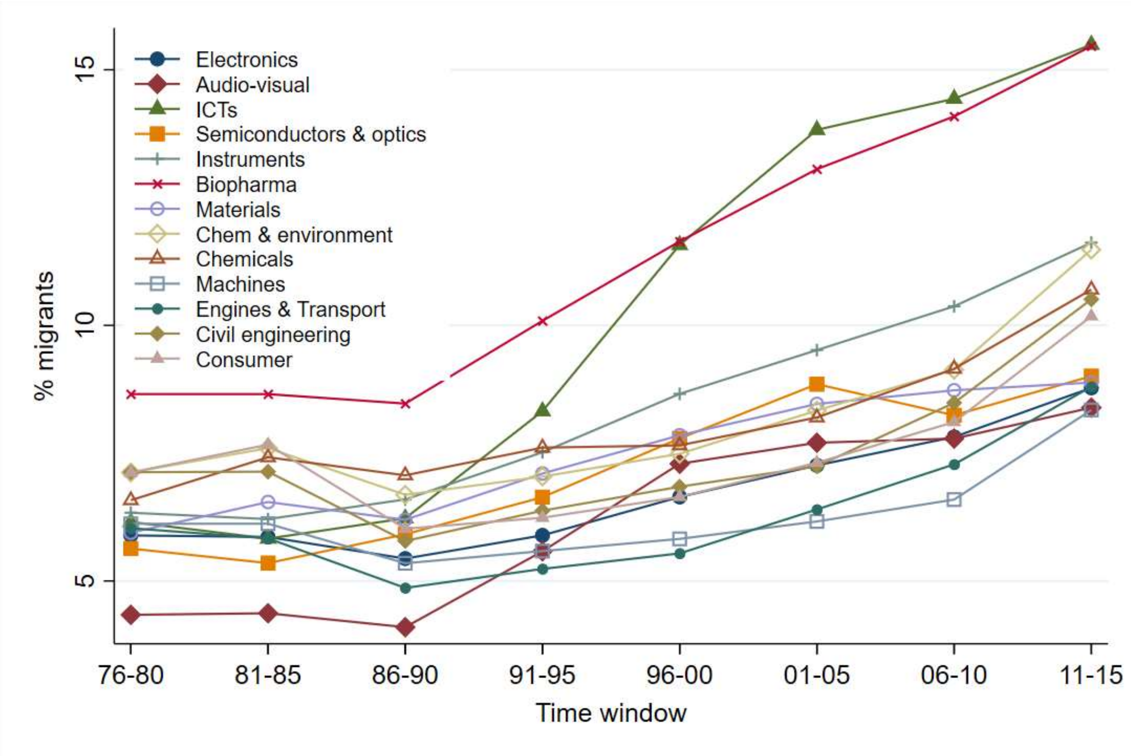
Japan, with negligible immigration rates, stand at the opposite end of the spectrum. In between, we find developing countries such as India and China. They start with high immigration levels, higher than those in the US or Europe; and then they see these levels declining incessantly. Most likely, the trend inversion is due to the prevalent nature of STEM immigration in such countries, which mostly consists of foreign researchers working at local MNCs' branches. These inventors weighed considerably when the indigenous innovative activity was limited, but lost importance thereafter.

Figure A.5: Inventor migration, selected countries



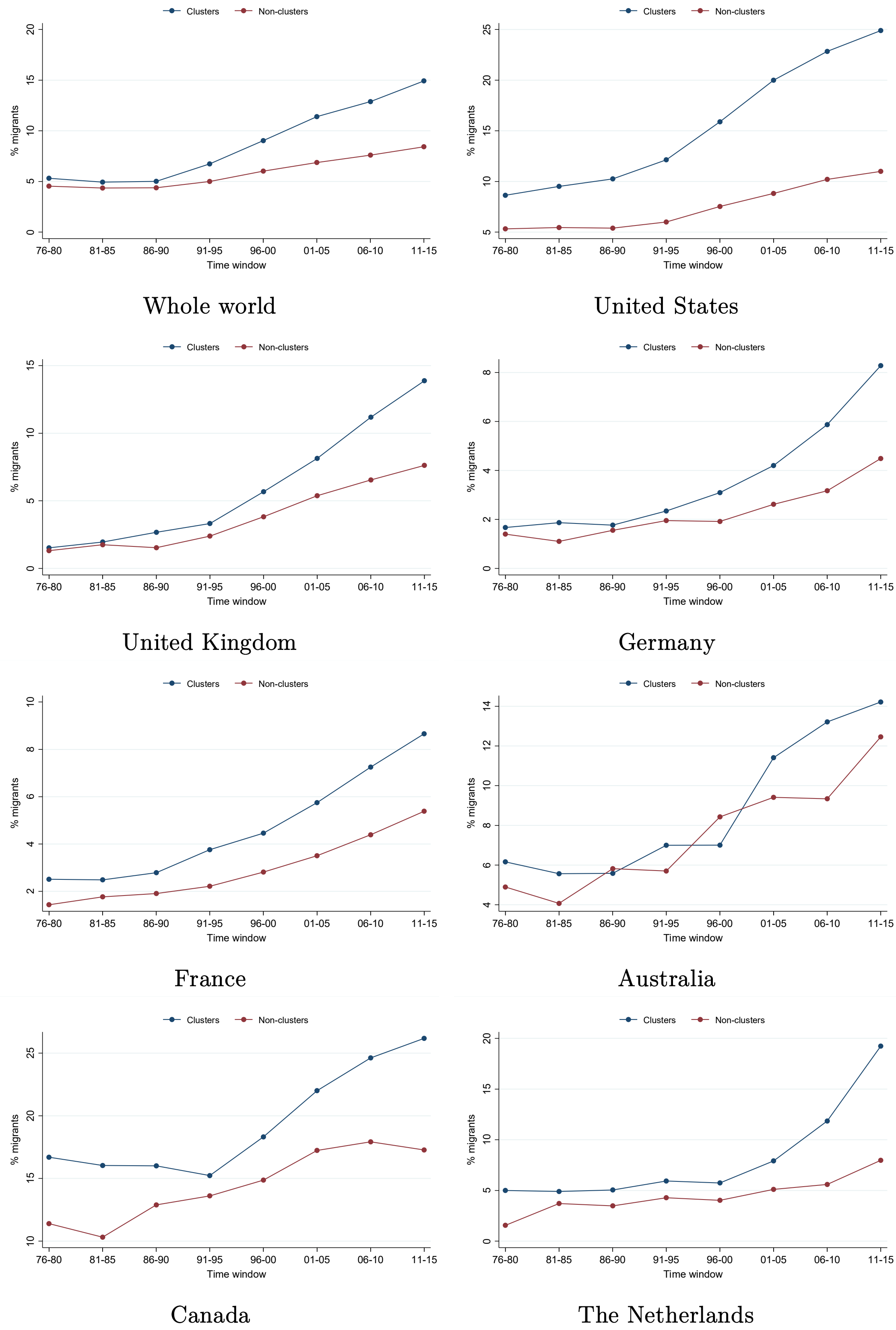
It is worth mentioning that inventor migration rates differ widely not only across countries, but also across technological fields (figure A.6). Two fields stand out for both rapid growth and high levels of migration rates, namely Biopharma and Information & Communication Technologies (ICTs), both of which are closely associated to higher education and academic research. Technologies whose advancement rely less on scientific education and research – such as Consumer goods, Engines & Transport, and Machines – stand at the opposite end.

Figure A.6. Share of migrant inventors worldwide, by technological field; 1976-2015



We next examine how GIHs and NCs are affected by international migration. Generally speaking, the migration rates of inventors tend to be higher in GIHs and NCs than outside them (figure A.7). While the share of migrant inventors remained stable and around 5% in both types of areas until mid-1990s, it started to growth after thereafter, which translates in large differences at the end of the period (15% vs 8% in, respectively, clusters and non-cluster areas).

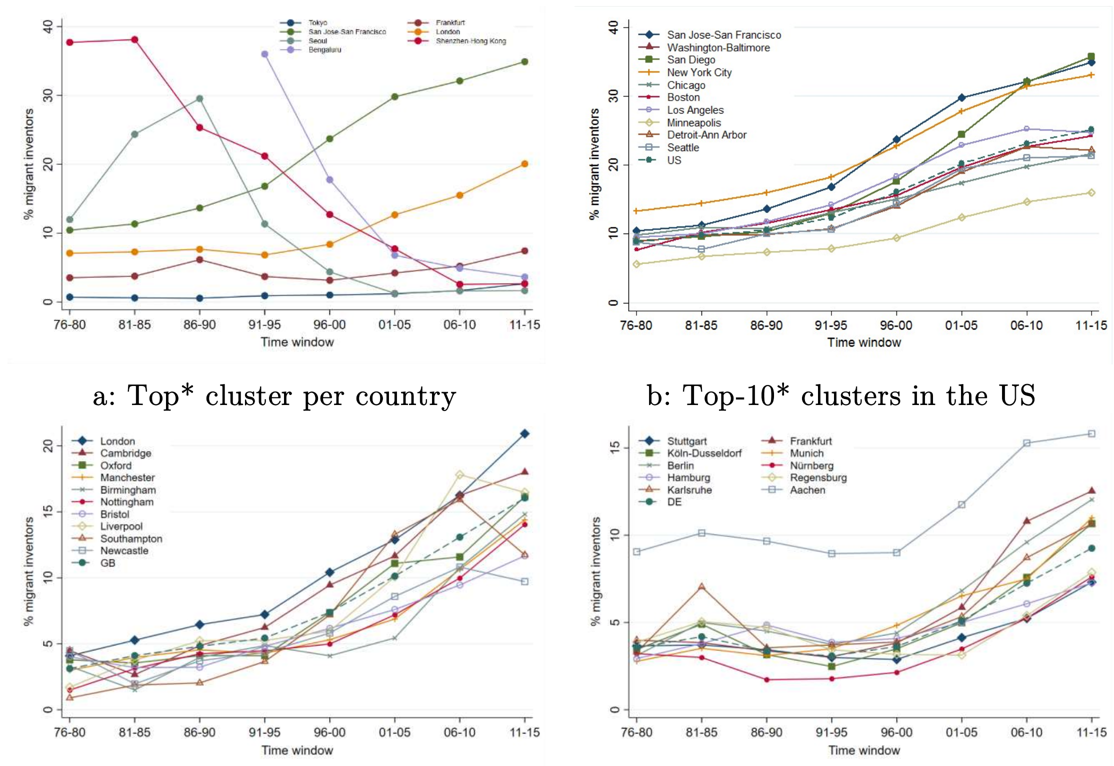
Figure A.7. Share of migrants – inside vs outside clusters



More importantly, there is considerable variation across locations. Figure A.8 reports the immigration trends for the largest clusters in the top-patenting countries (plus India).

These trends remind national ones, with Bengaluru, Seoul and Shenzhen following closely, respectively, the Indian, Korean and Chinese trends; and lines for San Jose – San Francisco, London and Frankfurt reminding of those for the US, UK and Germany. Levels, however, are different and especially for the latter cities, whose immigration rates are well above their national averages. As for Tokyo, it exhibits steadily low migration rate levels. Looking closer at the US (right-top panel) we notice that three agglomerations, the two of California ones plus New York, have double migration rates than the country as a whole. Two other West and East coast agglomerations (respectively Los Angeles and Boston) also stand out. On the contrary, the only mid-Western agglomeration in the top-10, Minneapolis, stands below. Data for the UK show that the above-average agglomerations, for migration rates, are the top one in size (London) and the two ones most clearly associated with academic research and its transfer (Cambridge and Oxford). As for Germany, the most interesting case is that of Aachen, which witnesses of the importance of intra-European migration, especially in border zones: the city, which stands next to both the Netherlands and Belgium, receives commuters from both countries (as well as other EU nationals residing there). The same is likely for Eindhoven and Brussels (unreported), which are right across the border.

Figure A.8. Share of migrants in clusters



a: Top* cluster per country

b: Top-10* clusters in the US

c: Top-10* clusters in the UK

d: Top-10* clusters in Germany

* As per 2011-15 ranking

For the sake of comparison, we also report here several figures and trends on inventor migration, built using inventors' nationality from PCT data (Fink and Miguelez, 2017) – figures A.9, A.10 and A.11.

Figure A.9: Share of PCT migrant inventors, selected countries, 1990-2010

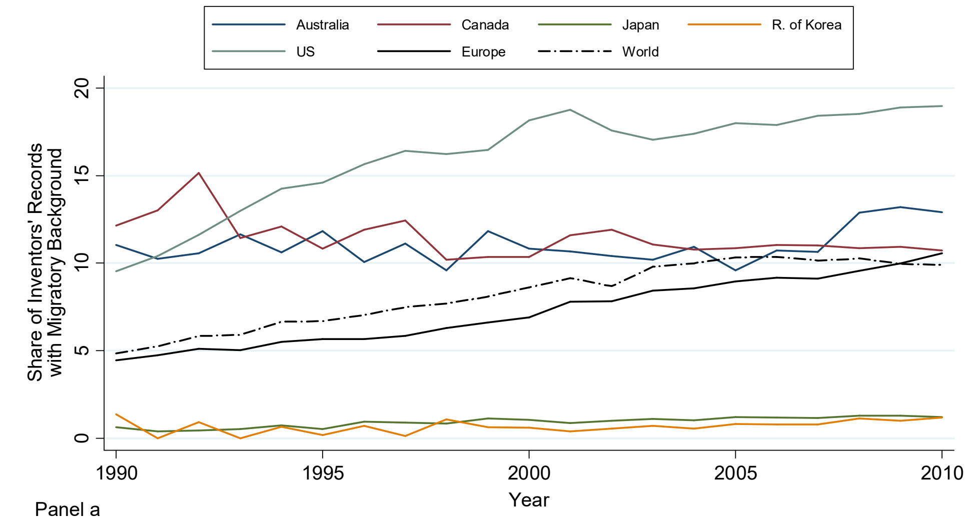


Figure A.10: Share of PCT migrant inventors, selected countries, 2001-2010

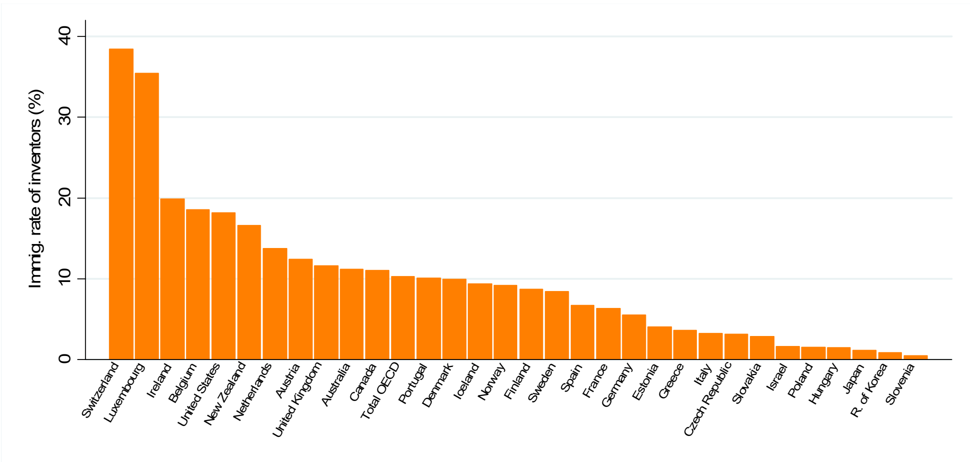
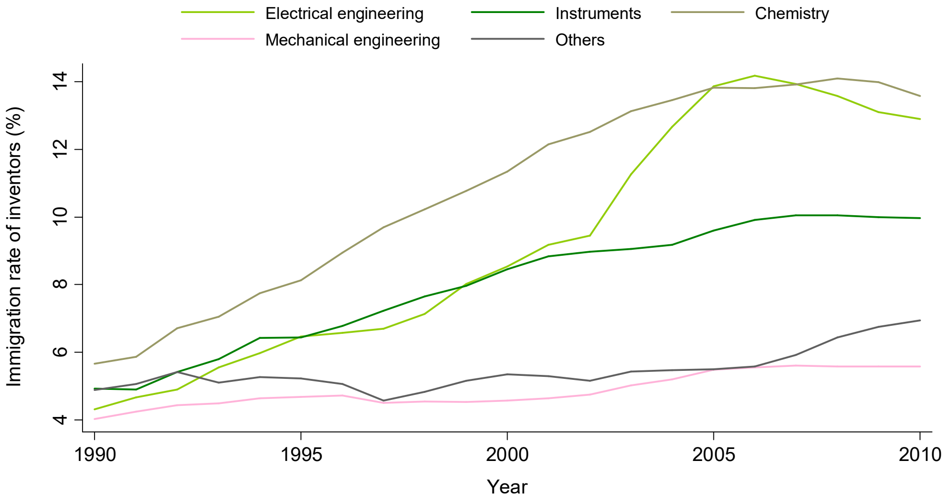


Figure A.11: Share of PCT migrant inventors, by broad technology, 1990-2010



Appendix 4. International teams, at the local level

At the local level, we can split patents between those produced within one or another form of agglomeration (at least one inventor's address coincides with a GIH or NCs) and those produced outside them (no inventor address to be found in any cluster). Based on this distinction, and for patents with two inventors or more, figure A.12 shows that inventive activities taking place in GIHs and NCs are the most internationalized (continuous lines). The difference is even starker when we focus only on highly cited patents (10% most cited patents, per year and technology; dotted lines). The figure also shows that an internationalization trend exists, but also that it considerably flattens out after 2005 (in the case of highly cited patents it even becomes negative).

Figure A.12: Share of teams (>1 inventor) based in more than 1 country – inside vs outside clusters

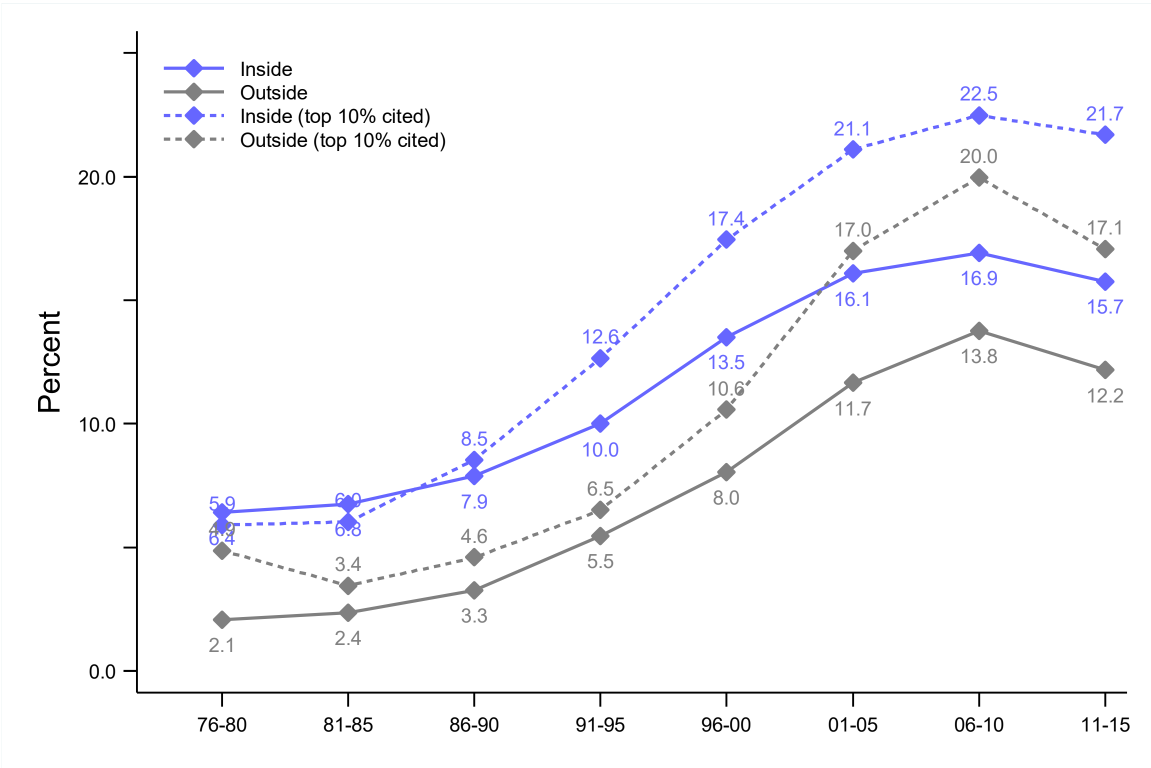
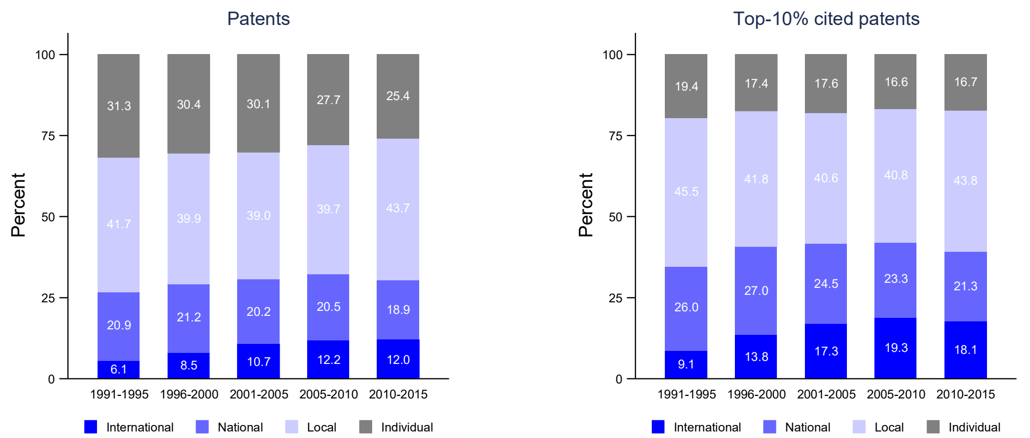


Figure A.13 examines the extent at which patents within GIHs and NCs owe to team activities as well as the composition of teams, over time. It distinguishes both between individual (single-inventor) patents and team-based (co-invented) patents and, for co-invented patents, between all-local, national, or international teams.

We first find a confirmation of the increasing importance of teams in inventive activities, witness the remarkable decline of individual patents. The more the hotspots and niche clusters collaborate, the denser the network of knowledge they create. Second, we notice that the share of local-only teams is larger than that of national and international ones. As for trends, the share of international-team patents increase up until around 2005, and flattens out after then, in coincidence with the general slowdown of internationalization of inventive activities we outlined above.

Figure A.13: GIHs' and NCs' share of co-inventorship, by team composition

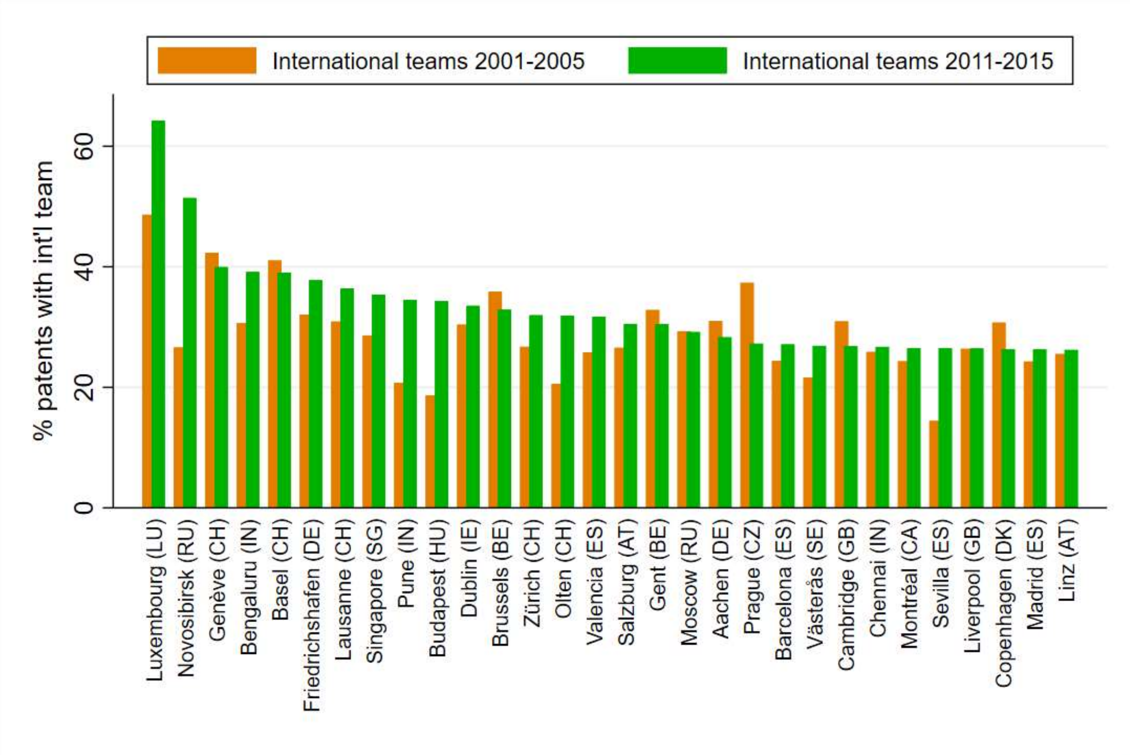


a: All patents

b: Top-10% cited patents

Figure A.14 looks at top-30 the clusters with at least 1000 patents share of international teams, in the period 2011-2015. For comparison purposes, the figure also reports the share of international teams in 2001-2005. European clusters generally rank high, especially in small countries such as Luxemburg, Switzerland, Ireland and other small countries. Some Indian and Russian clusters, however, also rank high in this list.

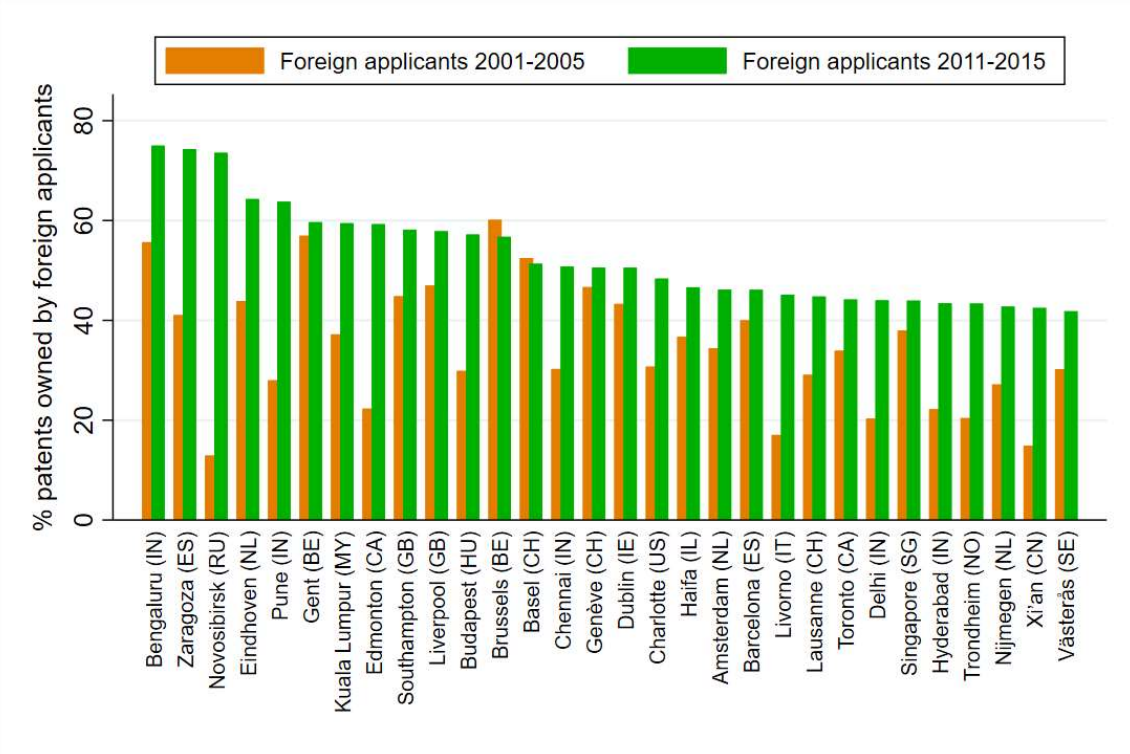
Figure A.14: Share of international teams by cluster



Appendix 5. Foreign applicants, at the local level

At the local level, figure A.15 reports information for the clusters with at least 1000 patents and the largest penetration of foreign applicants in the period 2011-2015. For comparison purposes, the figure also reports the share of foreign applicants in 2001-2005. We notice that Indian clusters (either GIH or NC) rank high in this respect, as already suggested in the main text. Other clusters from relatively small, open countries are also there, such as the Dutch, Belgian or Swiss ones. As for North American, European, Japanese, and Chinese clusters, none of the largest ones appears in the figure.

Figure A.15: Share foreign applicants by cluster



Appendix 6. Summary statistics and correlation matrix

Table A.2: Summary statistics regression sample

Variable	Obs	Mean	Std. Dev.	Min	Max
Patents p.c.	6,486	0.45	0.57	0	100.91
Citation weighted pat. p.c.	6,486	0.82	0.87	0	5.31
Degree centrality	6,486	0.06	0.06	0	6.45
Eigenvector centrality	6,486	0.07	0.07	0	0.52
Int'l teams	6,486	0.13	0.12	0	0.43
Foreign applicants	6,486	0.12	0.15	0	1
Migrants	6,486	0.09	0.10	0	1
Diaspora	4,596	3.77	1.69	0	1

Table A.3: Correlation matrix regression sample

	1	2	3	4	5	6	7	8
1. Patents p.c.	1							
2. Citation weighted	0.83	1						
3. Degree cen.	0.15	0.29	1					
4. Eigenvector cen.	-0.03	0.13	0.76	1				
5. Int'l teams	-0.06	0.00	0.14	0.03	1			
6. Foreign applicants	-0.13	-0.12	-0.02	-0.11	0.54	1		
7. Migrants	-0.12	0.03	0.21	0.21	0.23	0.37	1	
8. Diaspora	0.22	0.19	0.67	0.60	0.25	0.22	0.03	1

Appendix 7. Baseline results for international teams and foreign MNC

Table A.4. Effect of foreign MNC and teams' internationalization on clusters' success

	(1)	(2)	(3)	(4)
	Patents p.c.	Citation weighted	Degree cen.	Eigenvector cen.
International teams	0.182***	0.378***	0.0252***	0.0279***
	(0.0441)	(0.0731)	(0.00381)	(0.00430)
Foreign applicants	-0.385***	-0.580***	-0.0168***	-0.00321
	(0.0402)	(0.0636)	(0.00320)	(0.00315)
Constant	0.469***	0.824***	0.0630***	0.0682***
	(0.00709)	(0.0113)	(0.000607)	(0.000653)
Observations	6,978	6,978	6,978	6,978
Adjusted R2	0.801	0.814	0.878	0.850
Cluster FE	Yes	Yes	Yes	Yes
Tech FE	Yes	Yes	Yes	Yes
Cluster*Tech FE	No	No	No	No
Cluster* Time FE	Yes	Yes	Yes	Yes
Tech* Time FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Patents p.c. stands for the number of patents per thousand inhabitants in the time window (inverse hyperbolic sine transformation). Citation weighted stands for the number of patents per capita, weighted by their forward citations in a 3-year time window after patent priority year. Degree centrality and eigenvector centrality are normalized

Table A.5. Effect of foreign MNC and teams' internationalization on clusters' success

	(5) Patents p.c.	(6) Citation weighted	(7) Degree cen.	(8) Eigenvector cen.
International teams	0.0262	0.0935**	0.0108***	0.0159***
	(0.0266)	(0.0437)	(0.00261)	(0.00400)
Foreign applicants	-0.0736***	-0.116**	0.00257	0.00482
	(0.0279)	(0.0472)	(0.00234)	(0.00316)
Constant	0.449***	0.801***	0.0624***	0.0688***
	(0.00420)	(0.00712)	(0.000414)	(0.000615)
Observations	6,978	6,978	6,978	6,978
Adjusted R2	0.944	0.931	0.956	0.917
Cluster FE	No	No	No	No
Tech FE	No	No	No	No
Cluster*Tech FE	Yes	Yes	Yes	Yes
Cluster* Time FE	Yes	Yes	Yes	Yes
Tech* Time FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Patents p.c. stands for the number of patents per thousand inhabitants in the time window (inverse hyperbolic sine transformation). Citation weighted stands for the number of patents per capita, weighted by their forward citations in a 3-year time window after patent priority year. Degree centrality and eigenvector centrality are normalized

Table A.6. Effect of foreign MNC and teams' internationalization on clusters' success, by cluster size

	(1) Patents p.c.	(2) Citation weighted	(3) Degree cen.	(4) Eigenvector cen.
International teams * Top-25	0.444***	0.521***	0.0213**	-0.0127
	(0.102)	(0.133)	(0.0100)	(0.0113)
International teams * non-top-25	-0.0181	0.0500	-0.00258	0.00310
	(0.0269)	(0.0457)	(0.00278)	(0.00354)
Foreign applicants * Top-25	-0.141	-0.134	0.0110	-0.0108*
	(0.0893)	(0.108)	(0.00690)	(0.00590)
Foreign applicants * non-top-25	-0.0546**	-0.104**	-0.000778	0.00152
	(0.0264)	(0.0495)	(0.00250)	(0.00289)
Constant	0.441***	0.792***	0.0776***	0.0744***
	(0.00509)	(0.00784)	(0.000482)	(0.000575)
Observations	6,978	6,978	6,978	6,978
Adjusted R-squared	0.944	0.931	0.964	0.934
Cluster*Tech FE	Yes	Yes	Yes	Yes
Cluster* Time FE	Yes	Yes	Yes	Yes
Tech* Time FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Patents p.c. stands for the number of patents per thousand inhabitants in the time window (inverse hyperbolic sine transformation). Citation weighted stands for the number of patents per capita, weighted by their forward citations in a 3-year time window after patent priority year. Degree centrality and eigenvector centrality are normalized.

Appendix 8. List of clusters at the top-25 ranking of at least 1 cluster-technology

Top-25 clusters-technology are located in the following clusters: Amsterdam, Beijing, Bengaluru, Berlin, Boston MA, Chicago IL, Daejeon, Dallas TX, Detroit-Ann Arbor MI, Eindhoven, Frankfurt, Greenville NC, Hamamatsu, Houston TX, Icheon-si, Kingwood Area TX, Köln-Dusseldorf, London, Los Angeles CA, Minneapolis MN, Munich, Nagoya, New York City NY, Nürnberg, Osaka, Paris, Philadelphia PA, Portland OR, Regensburg, San Diego CA, San Jose-San Francisco CA, Seattle WA, Seoul, Shanghai, Shenzhen-Hong Kong, Shiojiri, Shizuoka, Stuttgart, Tel Aviv, Tokyo, and Washington DC-Baltimore MD. English includes Australia, Canada, Ireland, the UK and the US. South-East Asia includes China, India, Japan, R. of Korea and Singapore.

GREThA UMR CNRS 5113
Université de Bordeaux
Avenue Léon Duguit
33608 Pessac – France
Tel : +33 (0)5.56.84.25.75
http://gretha.u-bordeaux.fr/

LAREFI
Université de Bordeaux
Avenue Léon Duguit
33608 Pessac – France
Tel : +33 (0)5.56.84.25.37
http://larefi.u-bordeaux.fr/

Derniers numéros – Last issues

- 2021-01 **The Impact of Services Trade Restrictiveness on Food Trade**
  by Amara ZONGO
- 2020-21 **Zero Rating, Content Quality and Network Capacity**
  by Emmanuel LORENZON
- 2020-20 **Uninformed Bidding in Sequential Auctions**
  by Emmanuel LORENZON
- 2020-19 **Patent assertion entities and patent ownership transparency: strategic recording of patent transactions at the USPTO**
  by Valerio STERZI
- 2020-18 **Trade Uncorked: Genetic Resistance and Quality Heterogeneity in Wine Exports**
  by Olivier BARGAIN, Jean-Marie CARDEBAT, Raphaël CHIAPPINI
- 2020-17 **Responsabilité des travailleurs vis-à-vis de l'environnement et taxation optimale**
  by Sarra BEN SLIMANE, Damine BAZIN, Jérôme BALLET
- 2020-16 **Empirical Evidence of the Lending Channel of Monetary Policy under Negative Interest Rates**
  by Whelsy BOUNGOU
- 2020-15 **In knowledge we trust: learning-by-interacting and the productivity of inventors**
  by Matteo TUBIANA, Ernest MIGUELEZ, Rosina MORENO
- 2020-14 **Traditional Norms, Access to Divorce and Women's Empowerment: Evidence from Indonesia**
  by Olivier BARGAIN, Jordan LOPER, Roberta ZIPARO
- 2020-13 **Climate warming vs ecological competition for marine tropical biodiversity and fisheries**
  by Hélène GOMES, Coralie KERSULEC, Luc DOYEN, Fabian BLANCHARD, Abdoul CISSE, Nicolas SANZ

Ion LAPTEACRU and Ernest MIGUELEZ are the scientific coordinators of the Bordeaux Economics Working Papers. The layout and distribution are provided by Cyril MESMER and Julie VISSAGUET.