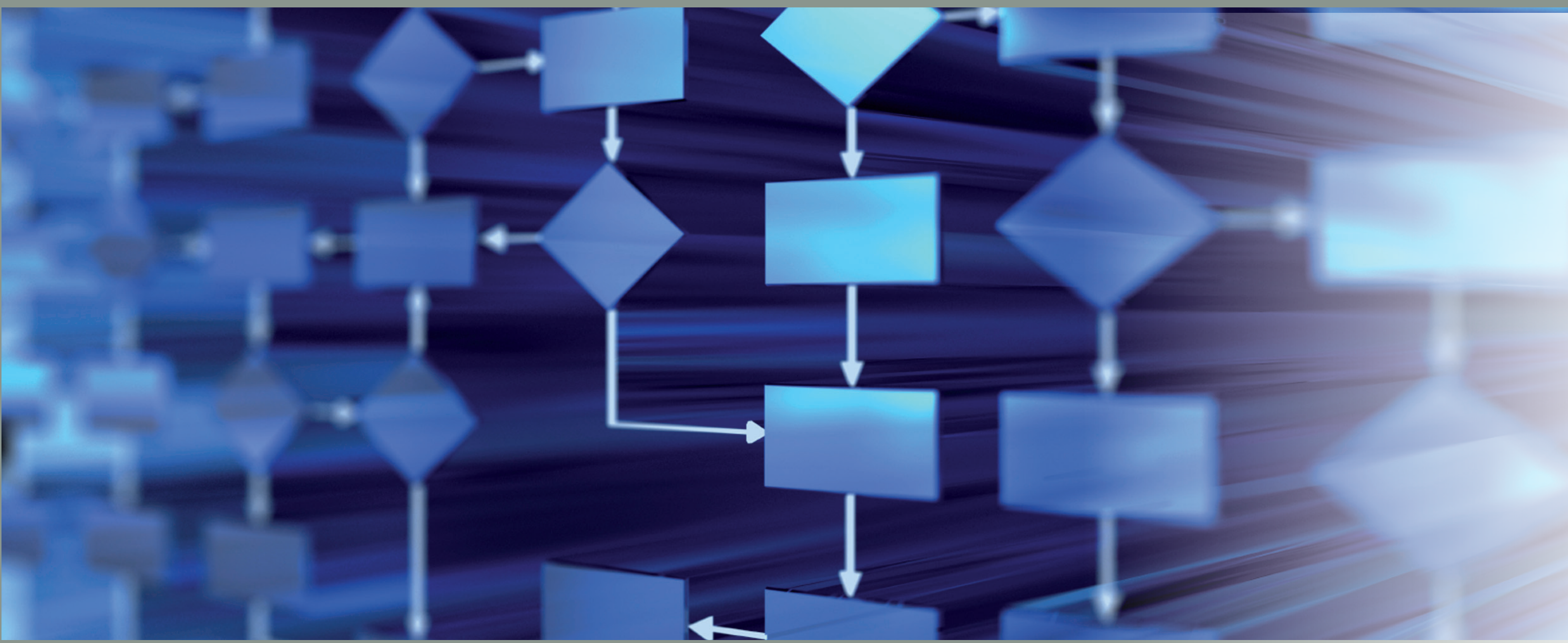
Economic Research Working Paper No. 58

Tied In: The Global Network of Local Innovation

Ernest Miguelez
Julio Raffo
Christian Chacua
Massimiliano Coda-Zabetta
Deyun Yin
Francesco Lissoni
Gianluca Tarasconi



November 2019

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Tied in: the Global Network of Local Innovation

Ernest Miguelez[2]
Julio Raffo[1]
Christian Chacua[2]
Massimiliano Coda-Zabetta[2]
Deyun Yin[1,3]
Francesco Lissoni[2,4]
Gianluca Tarasconi[5]

Abstract

In this paper we exploit a unique and rich dataset of patent applications and scientific publications in order to answer several questions concerned with two current phenomena on the way knowledge is produced and shared worldwide: its geographical spread at the international level and its spatial concentration in few worldwide geographical hotspots. We find that the production of patents and scientific publications has spread geographically to several countries, and has not kept within the traditional knowledge producing economies (Western Europe, Japan and the U.S.). We observe that part of this partial geographical spread of knowledge activities is due to the setting up of Global Innovation Networks, first toward more traditional innovative countries, and then towards emerging economies too. Yet, despite the increasing worldwide spread of knowledge production, we do not see the same spreading process within countries, and even we see some increased concentration in some of them. This may have, of course, important distributional consequences within countries. Moreover, these selected areas also concentrate a large and increasing connectivity, within their own country to other hotspots, and across countries through Global Innovation Networks.

Keywords: patents, scientific publications, geocoding, global innovation networks, clusters, geography of innovation

JEL codes: O30, F20, F60

[1] Innovation Economics Section, Economics and Statistics Division, WIPO
[2] *Groupe de recherche en économie théorique et appliquée, Unité mixte de recherche, Centre national de recherche scientifique, GREThA UMR CNRS 5113—Université de Bordeaux.*
[3] Corresponding author: Deyun Yin (yindeyunut@gmail.com).
[4] Invernizzi Center for Research on Innovation, Organization, Strategy and entrepreneurship (ICRIOS) – Bocconi University.
[5] ipQuants AG

Acknowledgements: We are indebted to Lesya Baudoin, Cristina Chaminade, Alica Daly, Gaétan de Rassenfosse, Maryann Feldman, Carsten Fink, Intan Hamdan-Livramento, Agénor Lahatte, Kazuyuki Motohashi, Orion Penner, Frédérique Sachwald, Florian Seliger and Usui Yoshiaki; as well as the participants to the World Intellectual Property Report (WIPR) 2019 workshop “From Global Innovation to Local Growth” (Geneva, 2019) and the II° International Workshop on Cluster Research, “Rethinking Clusters” (Padova, 2019). Financial support from the French National Research Agency (TKC project – reference: ANR-17-CE26-0016) and from WIPO is gratefully acknowledged. Any mistake remains ours.

1. Introduction

Throughout the 20th century, the production of scientific and technical knowledge was largely confined to a few countries, and especially within the so-called Triad (Japan, Western Europe, and the US), for two main reasons. First, it was these countries that hosted the largest universities, public research organizations and R&D-intensive companies, with the latter having a preference for keeping their research operations close to their headquarters (Castellacci and Archibugi, 2008; Chaminade et al., 2016; Patel and Pavitt, 1991). Second, the Triad also hosted most of the foreign-based R&D operations of multinational companies (MNCs), while the branches or controlled firms in the developing economies were mostly trusted with unsophisticated tasks such as the adaptation of products and processes to local market conditions (Krishna et al., 2012). This resulted in very limited foreign patenting activity, especially outside the Triad, in stark contrast with the increasingly global outsourcing of manufacturing (Gerybadze and Merk, 2014).

Starting in the 1990s, two changes occurred. First, some emerging economies, most notably in East Asia but also elsewhere, began gaining importance as knowledge originators (Branstetter et al., 2014). Second, foreign direct investments (FDIs) became increasingly dictated by knowledge-seeking strategies, with international R&D operations aiming at getting access to foreign knowledge (Amendolagine et al., 2019; Reddy, 1997). Return and circular high-skilled migrants reinforced this phenomenon (Saxenian, 2006), with international social networks helping to direct investments to the migrants' countries of origin (Foley and Kerr, 2013; Useche et al., 2019).

Large emerging economies such as those of China and India host nowadays not only the R&D operations of foreign MNCs, but also several important domestic actors. It is not yet clear, however, whether they can be ranked among the new global science and technology powers, or are still confined to subordinate positions in the international division of research and inventive labor (Awate et al., 2012; Branstetter et al., 2014).

Answering this question, however, requires scratching behind the surface of international statistics and country-level data. Knowledge-seeking FDIs do not target countries as a whole, but specific locations therein. Knowledge production, in fact, is dominated by economies of agglomeration, which results in its concentration in selected cities and regions (Feldman, 1994). At the same time, important locations are neither self-sufficient nor they can rely exclusively on knowledge inputs coming either from their countries' national science and innovation system or from large MNCs. Rather, they are expected to take part in a network of international collaborations, investments or personal mobility, which allow them to exchange knowledge with other locations worldwide (Bathelt et al., 2004; Lorenzon and Mudambi, 2013; Moreno and Miguelez, 2012).

In view of these considerations, it is important to examine the emergence of a global science and innovation system from a local perspective.

First, one needs to map the knowledge production centers that, from within each country, contribute to the increasing dispersal of innovative activities worldwide and the increase or reversal of international knowledge exchanges.

Second, in particular, when examining emerging countries, one also wishes to investigate whether the strengthening of their knowledge production centers goes along with an intensification of international collaboration and investments, or instead it occurs at the expense of their participation to innovative activities in the rest of world.

In what follows we produce both country- and local-level evidence. First, we investigate the extent at which the globalization of knowledge production has progressed, due both to the emergence of new countries outside the Triad and an increase of cross-country exchanges (sections 2 and 3). Second, we detect the cities and/or large metropolitan areas which stand behind the production of knowledge at the country level and examine the extent of their mutual, global connections (section 4). Section 5 concludes.

2. The internationalization of knowledge production

In the following paragraphs we exploit patent and publication data at the country level to assess the extent at which the production of knowledge and innovation has internationalized, and to identify the key actors behind the observed trends. When necessary, we compare the patent- and publication-based indicators with analogous indicators based on R&D or trade, among others.

As explained in detail in the annex, our patent data cover all the worldwide patent families, from 1976 to 2017. Roughly speaking, a (simple) patent family is the set of patent applications on the same invention filed in different countries' patent offices. The base unit of our analysis is the first filing for a set of patent applications filed in one or more countries and claiming the same invention. Each set containing one first and, potentially, several subsequent filings is defined as a patent family. By considering families, instead of individual patents, we avoid counting the same invention more than once. We assume that the inventive activity behind such patents take place in the countries of residence of the inventors, according to the latter's addresses as reported on the patent documents (occasional deviations from this assumption will be stated explicitly) (Webb et al., 2005).

We pay special attention to internationally oriented (or foreign-oriented). Foreign-oriented patent families concern those inventions for which the applicant has sought for patent protection beyond its home patent office. This definition includes also patent applications by applicants filing only abroad, filing only through the PCT system or filing only at the EPO. Reciprocally, domestic-only patent families refer to those patent applications filed only at the applicant's home office – regardless of how many filings in the home office there are within the same family – without any subsequent foreign filing though the Paris or PCT routes. Likewise, patent applications with applicants of more than one origin are, by our definition, foreign-oriented patent families. In addition, about 30 percent of the patent families relate only to utility model protection, which are mostly domestic only. Unless otherwise stated, we use international patent families as the unit of analysis for all patent statistics reported. This relates mostly to the incomplete coverage of the domestic-only patents (and utility models) of many national collections in PATSTAT. While the top national and international offices are usually well covered – namely USPTO, JPO, KIPO, CNIPA, EPO and WIPO – some other offices have limited coverage in PATSTAT. For instance, the coverage in PATSTAT of national collection data from some top 20 patent offices – such as India, Indonesia, Iran (Islamic Republic of), Mexico and Turkey – is limited. In addition, due to more limited information, we are less likely to geocode domestic only patent families and, even if so, we often do it at a lessen quality or precision. This is another reason why we rely mostly on international patent families. Occasionally, and if explicitly stated, we will produce some statistics based on the remaining patent families, which consist of domestic firms' single patent applications to their national patent office.

As for publication data, we rely on Clarivate's Web of Science, from 1998 to 2017. In this case, we assume the research conducting to the publication to take place in the countries of the institutions and organizations to which the authors declare their affiliation. We have used all the collection (all fields) from 1998 onwards, but only worked with articles in scientific journals, and excluded proceedings, books, etc.

We first examine the international concentration of both patents and publications. In order to do so, we adapt the Herfindahl-Hirschman (HH) index to our needs, as follows:

$$HH = \sum_{i=1}^{N} s_i^2 \quad (1)$$

where N is the number of countries in our database and s_i is each country's share of the activity of interest (research, invention or others). The HH Index ranges from 1/N to 1, where values equal or above 0.25 can be reinterpreted as if there are only four or less countries responsible for the indicator described and, hence, indicative of high concentration. Notice that by taking the reciprocal of the HH index (that is: 1-HH) we obtain a measure of international dispersion.

Figure 1 reports the yearly HH index for 195 countries, for several indicators: total exports, inward foreign direct investments (FDI), R&D (as measured by either the expenditures or staff employed, in full time equivalent), internationally oriented patents, and scientific publications (plus total population, as benchmark).[1]

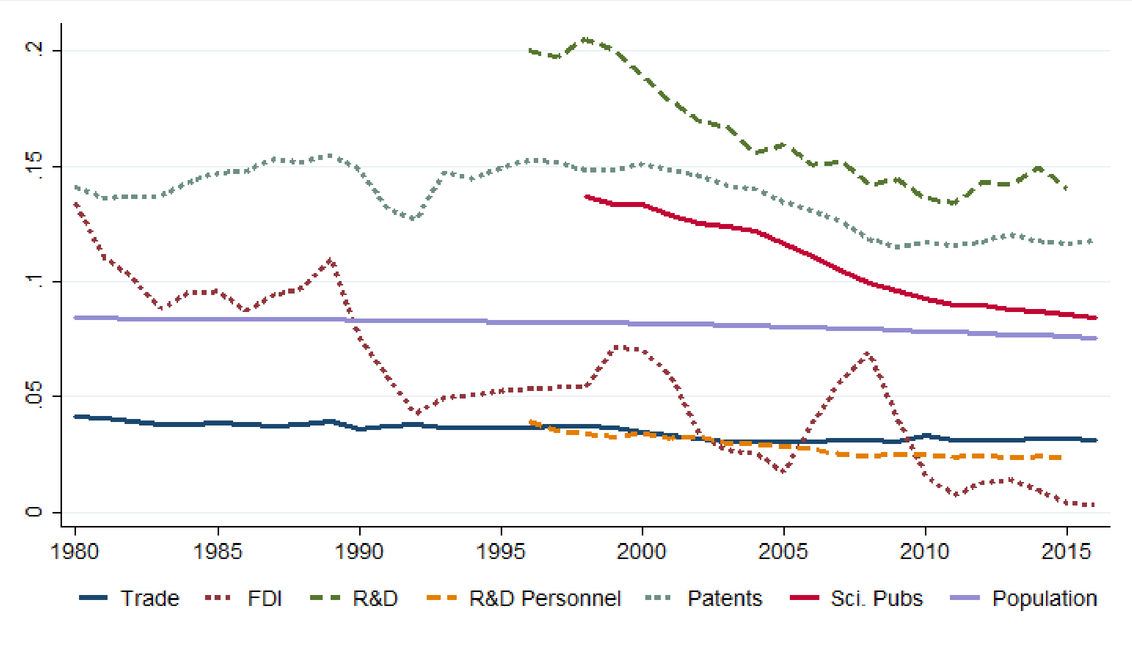
Two stylized facts stand out. First, knowledge-related and innovation activities, such as R&D expenditures, patents, and publications, are way more concentrated than population, trade and FDI. But, second, the concentration of such activities declines over time as much as, if not more, than that of trade and FDI.

Two further stylized facts emerge from **Figure 1**, which are worth stressing. First, R&D expenditures are more concentrated than R&D personnel. This implies that while we can find R&D employees everywhere across the world, their budgets differ widely. Their outputs differ too, with both patents and publications also being more concentrated than R&D personnel. This implies that researchers' and inventors' productivity, as measured by the number of patents and publications per person, differs widely by country.

Second, we observe a slight change in trend for R&D expenditures and patents after 2008, at the onset of the Great Recession. Whether this is a temporary, business-cycle-related occurrence, or the beginning of an inversion of long-term globalization trends is still an open question.

[1] We considered the countries with missing data as having 0% of the global share. This has the potential of overestimating concentration if one or more large economies are missing. After inspection of our data, we did not find any large or medium size country being unreported.

Figure 1: HH index for patents, publications and other variables



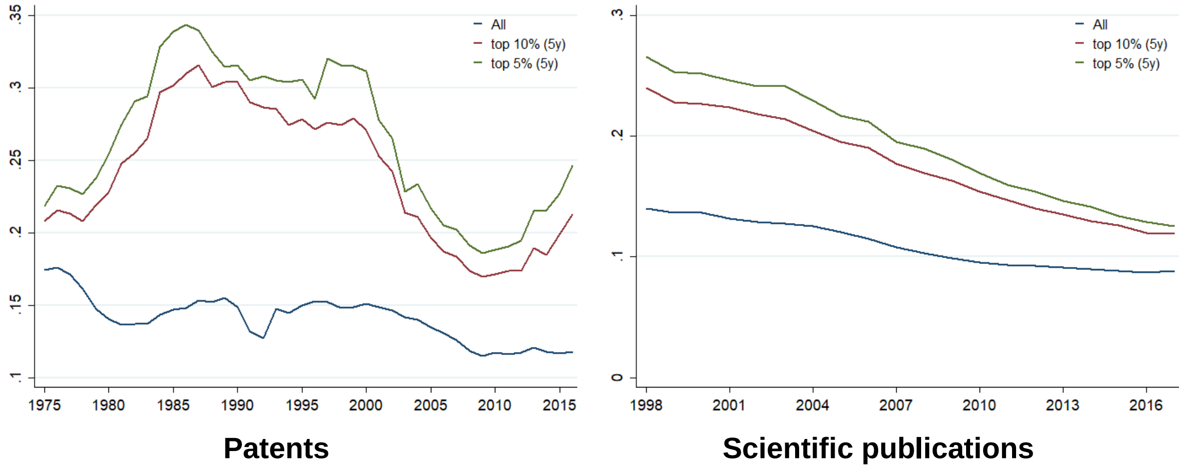
Source: Authors based on PATSTAT, PCT and Web of Science data. R&D and R&D personal are retrieved from the UNESCO Institute for Statistics, while trade and FDI data from the World Development Indicators, the World Bank.

So far, we have limited ourselves to counting patents and publications, and ignored that their scientific or technological importance (impact on subsequent research) can vary widely. We can measure such importance with the number of citations both patents and publications receive, respectively from other patents and from other publications (forward citations; Jaffe and de Rassenfosse, 2017). In the case of patents, some correlation also exists between forward citations and economic value (Jaffe and de Rassenfosse, 2017).

We ask whether the decreasing concentration described so far concerns also the most important and/or valuable items (patents and publications), namely those that received the most citations over 5 years after the date of, respectively, first filing at a patent office or appearance in a journal. To this end, **Figure 2** compares the HH indexes calculated for either the top-10% or top 5% most cited items with the general HH index (as calculated on all items, irrespective of the citations received). We first observe that the concentration of both highly-cited patents and highly-cited publications is higher than that of the general concentration, which implies that the higher the importance or value of knowledge, the more concentrated its production. At the same time, though, the concentration of highly-cited items declines, over time, faster than the general concentration, towards which it converges. This implies that emerging countries do not only produce more and more knowledge, but also knowledge that is highly valuable and important.

The main exceptions to this trend concern patent production in the 1980s (a period we cannot cover for publication) and after the Great Recession of 2008, when concentration increases. Part of the explanation of the latter period increase is the extraordinary appearance in the last decade of Eastern Asian economies, particularly Republic of Korea and China, as main hubs of patent production, which we will discuss in the next section. However, other factors may also be at play. We look at this in the following paragraphs.

Figure 2: HH index, by forward citations received



Source: Authors based on PATSTAT, PCT and Web of Science data.

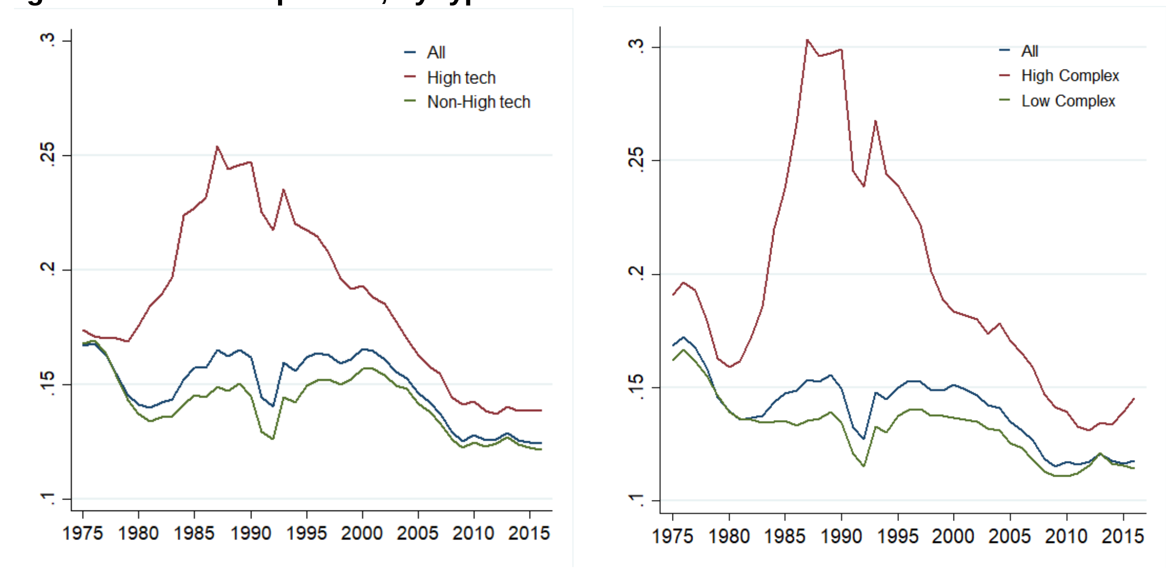
We hypothesize that the patterns found in figures 1 and 2 may be due to a succession of cycles marked by the appearance and rapid development of certain breakthrough technologies, followed by diffusion. At the beginning of each cycle (in the 1970s and the 2010s), highly-cited patents concentrate in leading countries, while during the diffusion phases inventive activities spread out (on this hypothesis, see Crescenzi et al., 2019).

Figure 3 looks at patents and splits them either into high versus non-high technologies (left panel) or into low- and high-complex technologies (right panel), based on their IPC classification (where IPC stands for International Patent Classification) (to identify high-complexity patents, we follow Fleming and Sorenson, 2004, and Sorenson et al., 2006). Intuitively, the high versus non-high distinction (as defined by Eurostat[2]) attains to the maturity reached by the technology with, say, nanotechnology standing higher than textile. As for discrete versus complex technologies, it attains at whether a new product or process can be protected, respectively, by a single or a handful of patents (as it is the case with new chemical formulas for drugs or materials) or by a large number of patents combined which result from the combination of many new features, each protected by a patent (as it is the case with, say, smartphones or means of transport).

Patents in high and complex technologies show a similar pattern: they are systematically more concentrated than the others, their concentration increases during the 1980s (after an initial decline in the 1975s), declines later on and picks up again in the 2010s. However, the 1980s-2000s pattern is more extreme for patents in complex technologies. This lends some support to the hypothesis that the most disruptive and breakthrough innovation activities first determined a concentration increase in high-income countries, with low-and middle-income countries following only later on (D'Agostino and Santangelo, 2012; Crescenzi et al., 2019). As for the increase of concentration in the 2010s, this is most visible for high-complex technologies.

[2] http://ec.europa.eu/eurostat/cache/metadata/Annexes/pat_esms_an2.pdf

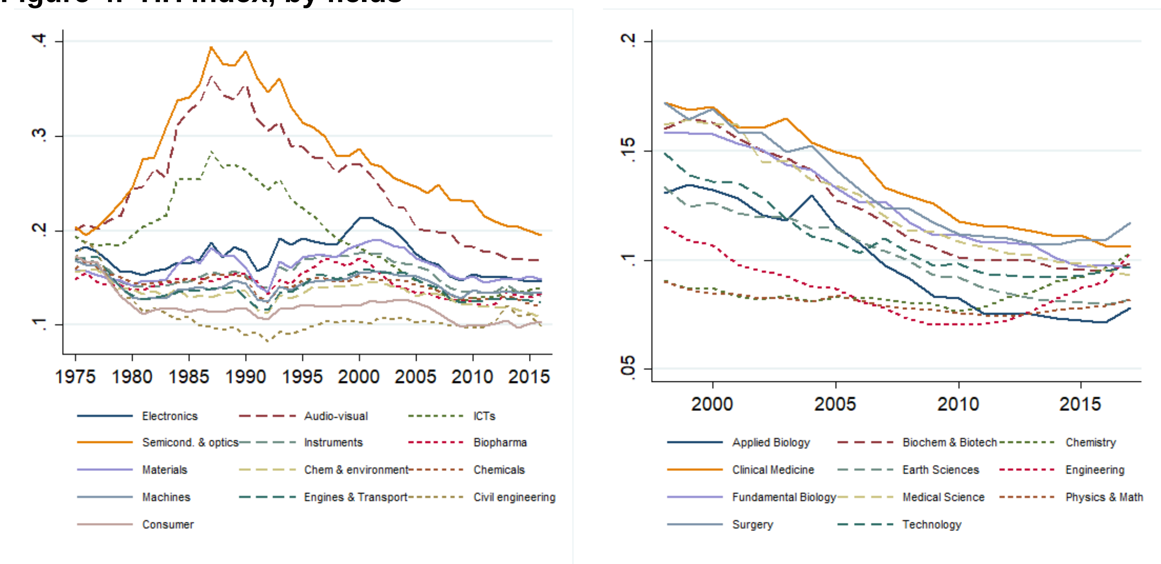
Figure 3: HH index patents, by type



High-tech vs low-tech patents **Complex vs discrete tech.**

Source: Authors based on PATSTAT, PCT and Web of Science data.

Figure 4: HH index, by fields



Patents **Publications**

Source: Authors based on PATSTAT, PCT and Web of Science data. Note: The humanities field has been excluded from panel b. Early on, WoS did not include many Humanities journals and those few were in English from the U.S. and UK. The number of journals has increased over time and, in the Humanities (where English does not rule as in other disciplines), this means having included non-English speaking journals, where publications form non-Anglophone countries are most numerous. To put it differently, the high level and dramatic converge you see witness of the changing composition of WoS for Humanities, rather than any real pattern.

We further explore this issue in **Figure 4**, where we split patents by specific technologies and by scientific field.[3] Left panel shows that the patterns observed for high technologies is mostly driven by Semiconductors and Optics, AudioVisuals, and, to a lesser extent, by Information and Communication technologies (ICTs). Conversely, the increase of concentration in the 2010s seems largely dominated by BioPharma technologies and, to a lesser extent, by Civil Engineering and Consumer Goods. Right panel repeats the analysis for scientific fields (which start only in 1998). All the fields show a steady decrease of the HH index. However, there are

[3] For details on the grouping of patents and publications into broad fields, see annex 2.

still large heterogeneity across fields, especially during the first part of the period analyzed. In addition, some fields, such as Applied Biology, Chemistry, and Engineering, show a slight upturn at the very end of the period, which is possibly in accordance with what we found for patents in related technologies.

Overall, we can conclude that the geographical dispersion of knowledge production has increased for the most part of our observation period, but also that the general trend hides important differences across technologies and level of importance of patents, as well as scientific fields (albeit less so).

While aggregate concentration figures clearly suggest that the production of scientific and technological knowledge has increasingly dispersed across countries, they beg the questions of which countries contributed the most to this trend.

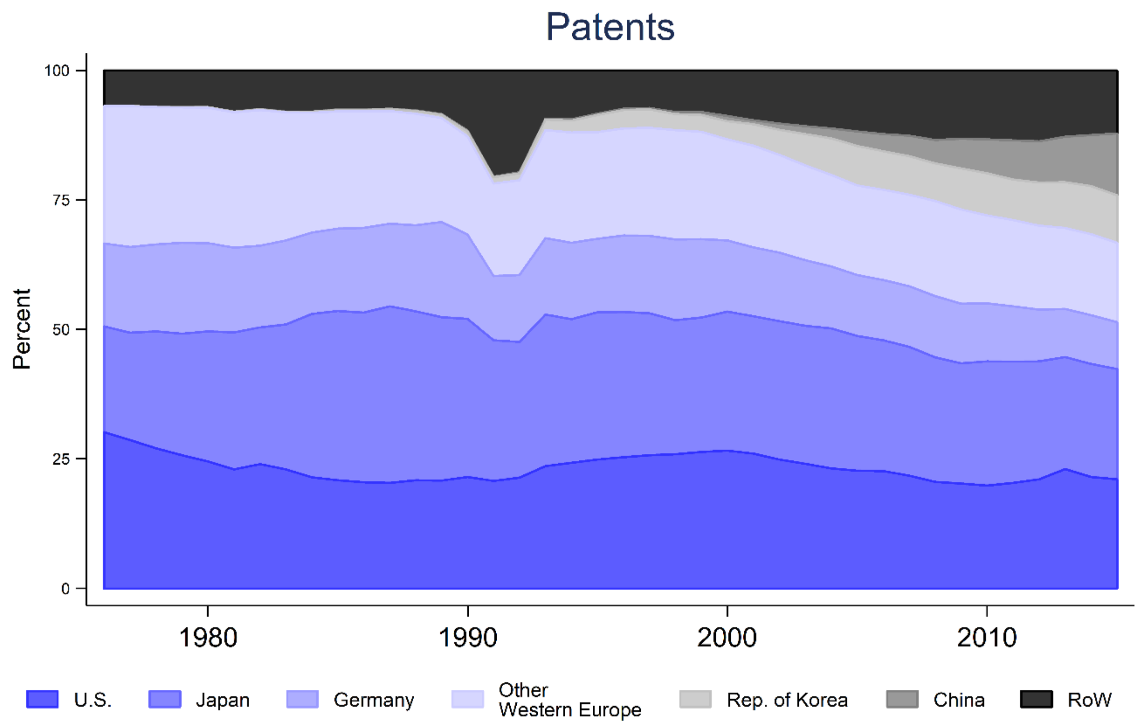
Figure 5 shows that, from 1970 to around 2000, only three countries – namely the United States of America (U.S.), Japan and Germany – accounted for two thirds of all patenting activity worldwide. Adding the remaining Western European economies – in particular the United Kingdom (U.K.), France, Switzerland and Italy – takes the figure to around 90%.

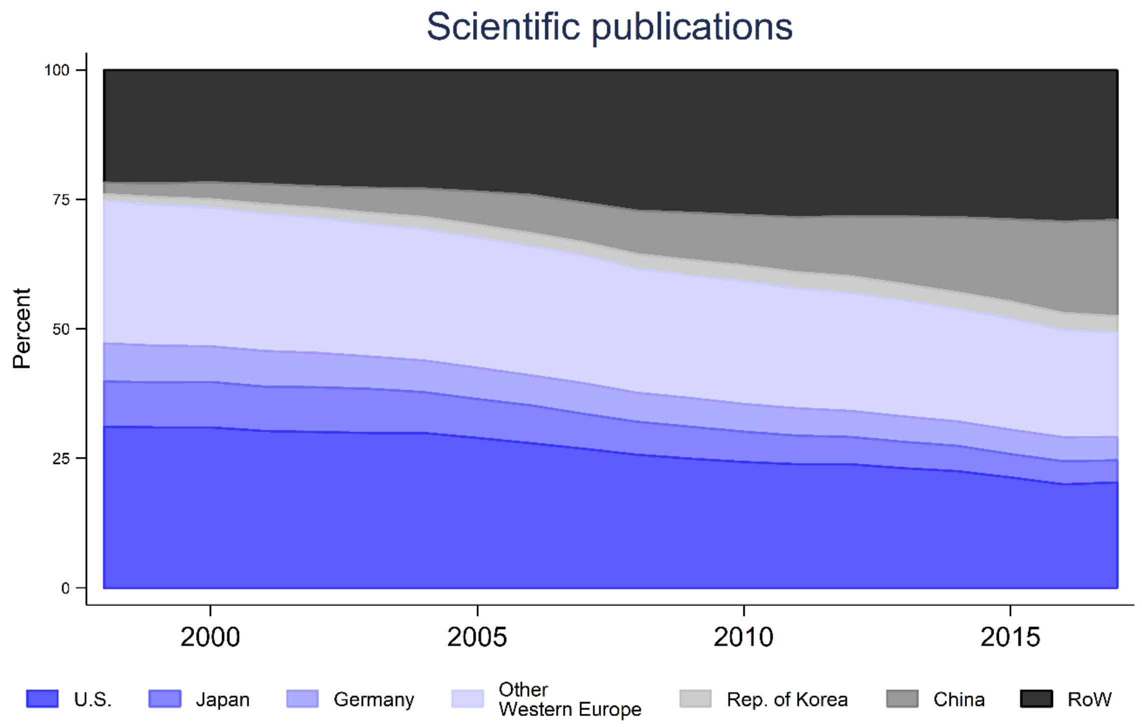
Starting in the 2000s, the rest of the world – a heterogeneous group that ranges from some high-income countries, such as Canada or the Republic of Korea, to mostly middle- and low-income economies – has outpaced in its share of knowledge production not only Western Europe but also the U.S. and Japan. China and the Republic of Korea are the two countries with the most impressive record, but they do not explain entirely the observed trend. Even after adding their shares to those of the Triad, the remaining countries in the rest of the world increase their share of patents from less than 6% at the beginning of the 1970s to around 13% during the 2000s. The share arrives to around one third if we add China and the Republic of Korea to the figures of the rest of the world.

In the last two decades, scientific publications have spread even more widely, with the rest of the world (including China and R. Korea) moving from less than a quarter of all scientific publication to around 40% at the end of the period. Again, this change is largely due to China and South Korea, but the growth is visible even without taking them into account.

Table 1 suggests that the most dynamic newcomer countries besides China and R. Korea are mostly found in Asia, whether in its West (Turkey, Israel), South-East (India, Singapore) and Centre (Iran). African countries stand out for their absence (with their main knowledge producing countries, Egypt and South Africa, contributing very little to both worldwide patenting and publishing). In Latin America, only Brazil is a meaningful player and one which exhibits some growth. Oceania – mostly pushed by Australia – has seen a small but steady increase in scientific publication shares, while patent shares have decreased since the early 2000s.

Figure 5: Evolution of patenting and publication share by top economies





Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: Other Western Europe excludes Germany.[4]

[4] Western Europe includes the 15 economies that were members of the EU prior to May 1, 2004, along with Andorra, Iceland, Liechtenstein, Malta, Monaco, Norway, San Marino and Switzerland.

Table 1: Evolution of patenting and scientific publishing, by regions and selected countries

	Patents							Publications			
Region (country)	**1970-1979**	**1980-1989**	**1990-1999**	**2000-2004**	**2005-2009**	**2010-2014**	**2015-2017**	**2000-2004**	**2005-2009**	**2010-2014**	**2015-2017**
SCSE Asia	**0.1%**	**0.1%**	**0.6%**	**1.0%**	**1.6%**	**2.1%**	**2.0%**	**3.2%**	**4.8%**	**6.7%**	**7.5%**
India	0.0%	0.0%	0.1%	0.5%	1.0%	1.4%	1.3%	2.0%	2.6%	3.2%	3.5%
Singapore	0.0%	0.0%	0.1%	0.3%	0.4%	0.4%	0.3%	0.4%	0.5%	0.5%	0.5%
CEE	**3.2%**	**3.8%**	**4.9%**	**1.1%**	**1.3%**	**1.4%**	**1.3%**	**5.8%**	**5.9%**	**5.8%**	**5.6%**
Russian Federation	0.7%	1.4%	2.7%	0.4%	0.5%	0.5%	0.4%	2.4%	1.9%	1.7%	1.8%
Poland	0.2%	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	1.1%	1.3%	1.3%	1.3%
LAC	**0.3%**	**0.3%**	**0.3%**	**0.4%**	**0.5%**	**0.6%**	**0.6%**	**3.0%**	**3.5%**	**4.0%**	**4.0%**
Brazil	0.1%	0.1%	0.1%	0.2%	0.2%	0.3%	0.3%	1.5%	2.0%	2.3%	2.3%
Western Asia	**0.3%**	**0.3%**	**0.7%**	**1.1%**	**1.4%**	**1.6%**	**1.7%**	**2.3%**	**2.8%**	**3.0%**	**3.1%**
Turkey	0.0%	0.0%	0.0%	0.1%	0.2%	0.3%	0.4%	1.0%	1.5%	1.7%	1.7%
Israel	0.2%	0.3%	0.6%	0.9%	1.2%	1.1%	1.1%	0.9%	0.8%	0.6%	0.6%
Oceania	**0.8%**	**1.1%**	**1.1%**	**1.4%**	**1.3%**	**0.9%**	**0.9%**	**2.4%**	**2.4%**	**2.6%**	**2.8%**
Australia	0.7%	1.0%	1.0%	1.2%	1.1%	0.8%	0.8%	2.0%	2.1%	2.3%	2.5%
Africa	**0.3%**	**0.2%**	**0.2%**	**0.3%**	**0.2%**	**0.2%**	**0.2%**	**1.1%**	**1.3%**	**1.6%**	**1.8%**
Egypt	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.3%	0.4%	0.5%
South Africa	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.3%	0.4%	0.4%	0.4%
Total	**4.8%**	**5.8%**	**7.8%**	**5.3%**	**6.4%**	**6.8%**	**6.7%**	**17.8%**	**20.7%**	**23.6%**	**24.9%**

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: CEE = Central and Eastern Europe; LAC = Latin America and the Caribbean; SCSE Asia = Southern, Central and South-eastern Asia.[5]

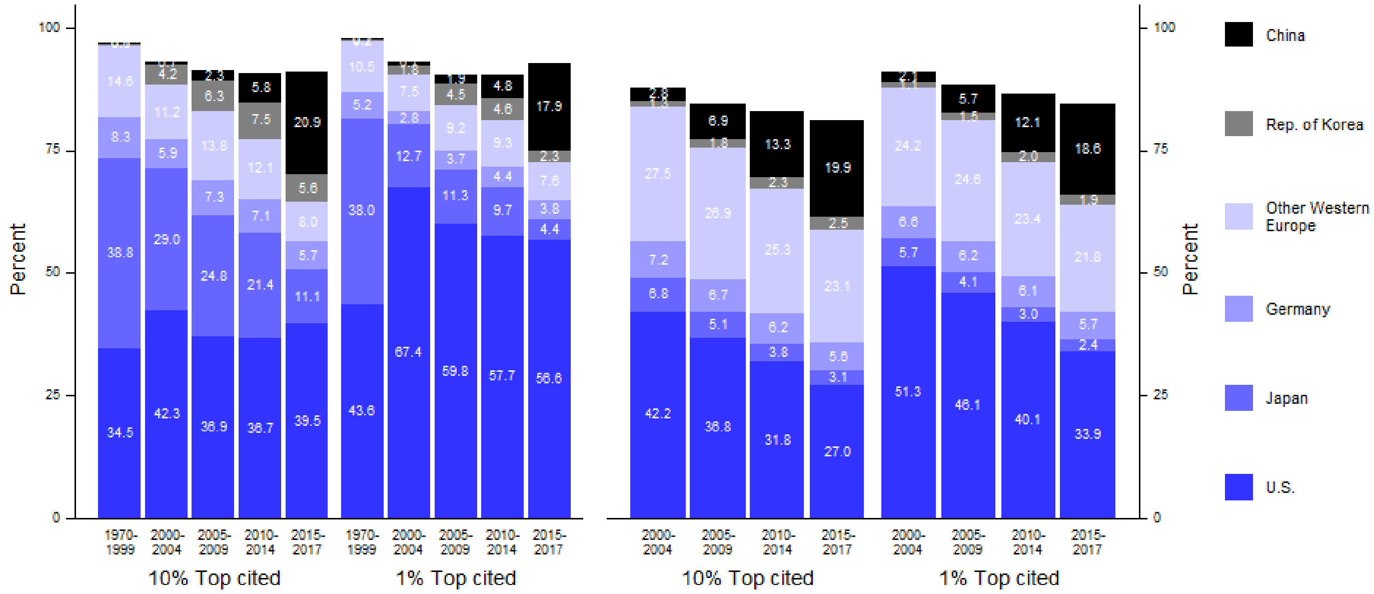
Figures for highly-cited patents and scientific publications exhibit a similar geographical pattern as that for all items, but with a greater and more resilient concentration in the U.S., and a more rapid declines of European and Japanese shares (**Figure 6**). China stands out for its particularly rapid ascent, compared to other countries in the rest of the world.

Summing up, knowledge production has increased in volume and spread more globally, but there is still a limited set of countries that produce the bulk of it, especially when it comes to highest quality outputs.

We now move on to enquire on whether, at least for the countries involved, the internationalization process has also translated into a globalization one, that is in a tighter integration of the established and emergent knowledge producers.

[5] These regions are closely based on the geographic regions from the U.N. Statistics Divisions (UNSD) methodology (https://unstats.un.org, accessed March 2019). The only differences are that CEE includes all countries in the UNSD's Northern and Southern Europe categories not included in Western Europe and that SCE includes Mongolia.

Figure 6: Evolution of top-cited patents and scientific publications by top economies and regions



Source: Authors based on PATSTAT, PCT and Web of Science data.

3. International knowledge production: how much is it global, too?

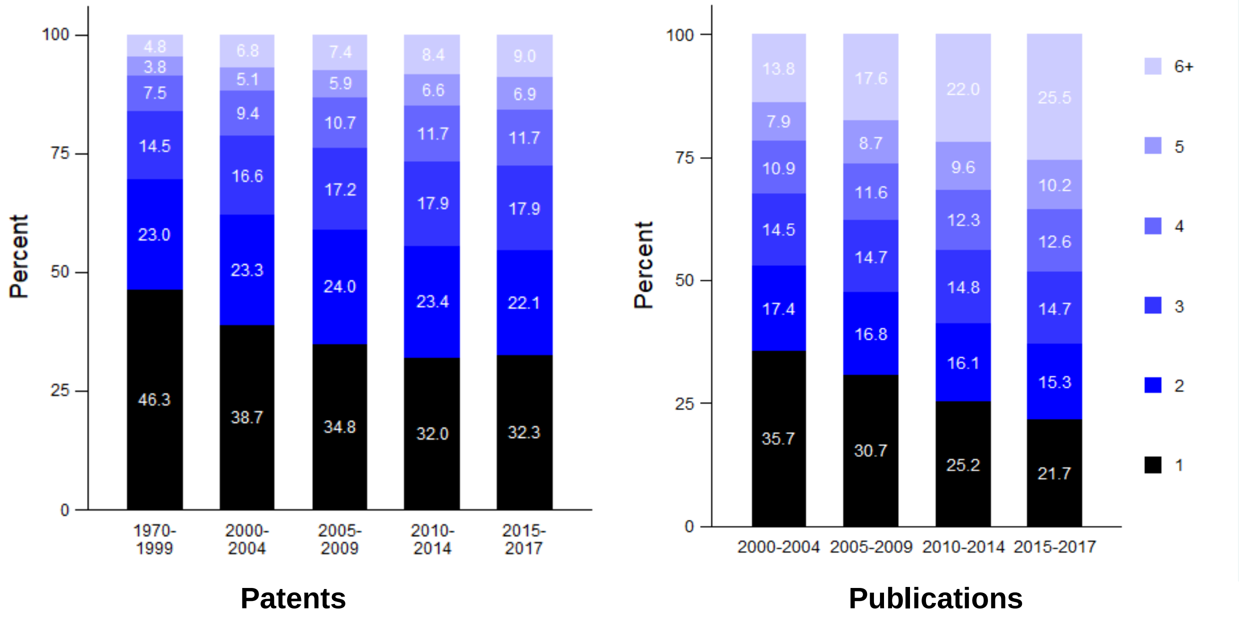
When it comes to examining the cross-country integration of knowledge production, it is useful to refer to the concept of Global Innovation Networks (GINs).[6] A GIN can be defined as a globally organized web of collaborative interactions between organizations (whether firms, universities, international agencies or others), engaged in knowledge production and with a relevant innovation outcome. GINs result from the knowledge-seeking strategy pursued by such organizations, and cannot be reduced to a side effect of market penetration and production outsourcing strategies of MNCs (Castellani et al., 2006; Castellani and Zanfei, 2007).

GINs are kept together by both organization-based linkages (Bathelt et al., 2004; Dunning, 1998) and personal relationships (Lorenzen and Mudambi, 2013), including those resulting from the international mobility of scientists, innovators and entrepreneurs (Breschi et al., 2017; Franzoni et al., 2012; Saxenian, 2006, 2002).

It is important to underline that these increasingly interconnected GINs are happening in the context of an overall increase of collaboration in the production of scientific and technological knowledge. Already in 1998, teams produced the majority of scientific papers. By 2017, *lone wolf* scientists had become half as important as they were 20 years before. The size of the teams is also increasing. In 2017, the average scientific paper required almost two more researchers – on average – than 20 years previously (**Figure 7**). Moreover, the average team size has shifted upward across the board, making teams of six or more scientists the most common in the production of scientific knowledge. Teams collaborating to achieve technological innovations (patents) are smaller but follow a similar increasing trend, with the average team number doubling since the early 1970s. By the mid-2010s, two thirds of inventions were collaborative efforts. All team sizes of inventors are increasing at the expense of single-inventor patents.

[6] See Chaminade et al. (2016).

Figure 7: Inventor and scientific team size, by period



Source: Authors based on PATSTAT, PCT and Web of Science data.

We now examine the importance, geographical distribution, and evolution over time of international patent and publication teams. We geo-localize inventors on the basis of their address, while for authors we consider their affiliation, which we also geo-localize. We consider a team to be international whenever it includes inventors or authors from different countries.[7]

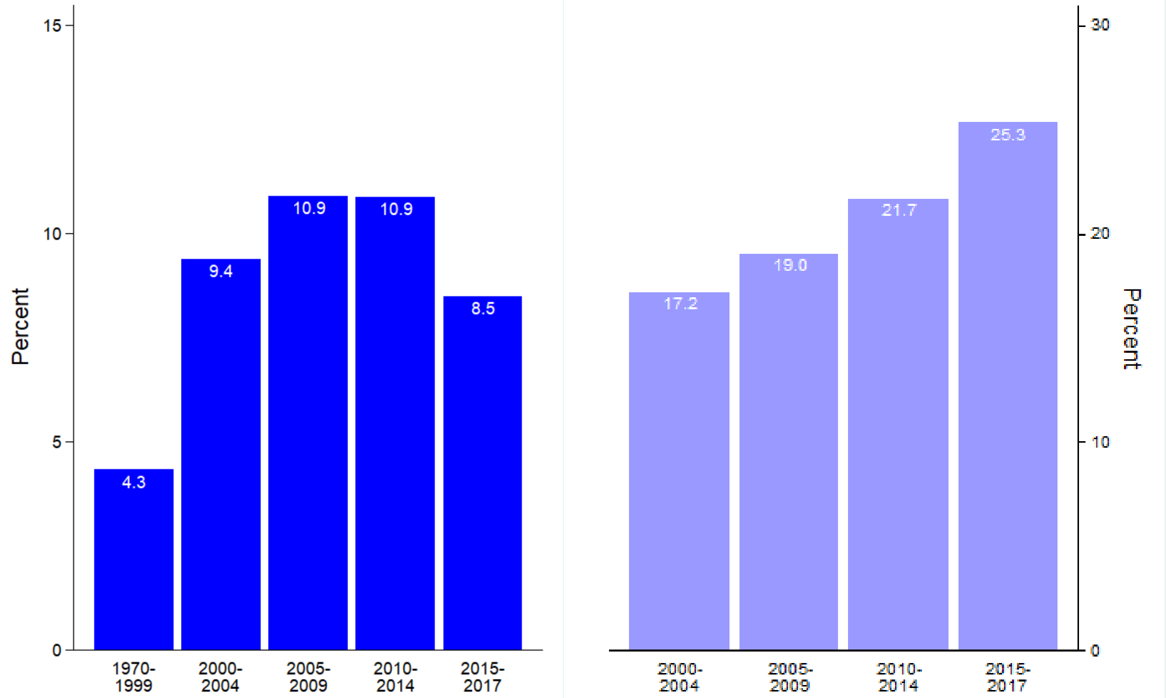
Figure 8 summarizes the main trends, starting in the 1970s for patents and in 1998 for publications. The most striking pattern concerns publications, whose share of internationally co-authored ones has grown from 17% to 25% in less than 20 years.

International co-inventorship is a much less frequent phenomenon, which never went beyond 11% of total patents. Yet, it also exhibits an impressive growth up until the second half of the 2000s, during which it more than doubles. After the Great Recession, the share has become slightly negative. The fact that international teams account for a higher percentage of published scientific articles than of patents indicates that science production is more internationalized than technology production.

We explored if the cooldown of co-inventorship is only an artifact due to computing the share – that is, international co-inventorships still growing, but national teams growing faster. In unreported results, we see that both overall co-inventorships and international co-inventorships have been increasing at a strong pace for most of the past 40 year. However, the number of international co-inventions has a notable slowdown from 2010 onwards, while the total number of co-inventions keeps rising.

[7] In scientific publication there are instances of double affiliation of authors. Unfortunately, one cannot uniquely identify affiliations and authors,

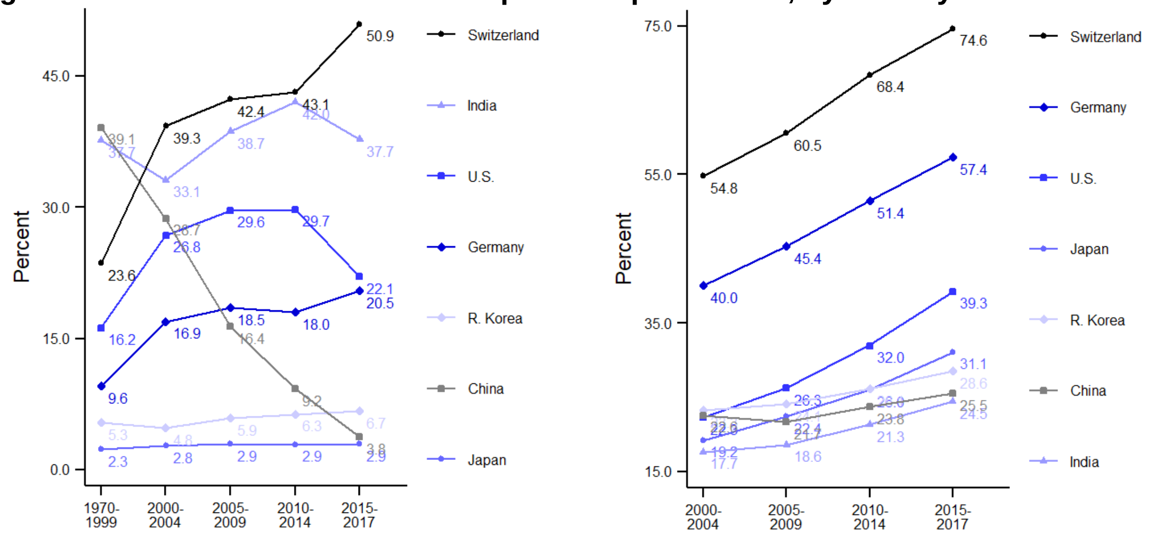
Figure 8: International co-inventorship and co-publishing by country, percent



International co-inventorship **International scientific co-publishing**

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: int. co-inventorship = share of patents with more than one inventor located in at least two countries; int. co-publications = share of scientific articles with more than one affiliation located in at least two countries.

Figure 9: International co-inventorship and co-publication, by country



International co-inventorship **International scientific co-publishing**

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: int. co-inventorship = share of patents with more than one inventor located in at least two countries; int. co-publication = share of scientific articles with more than one affiliation located in at least two countries.

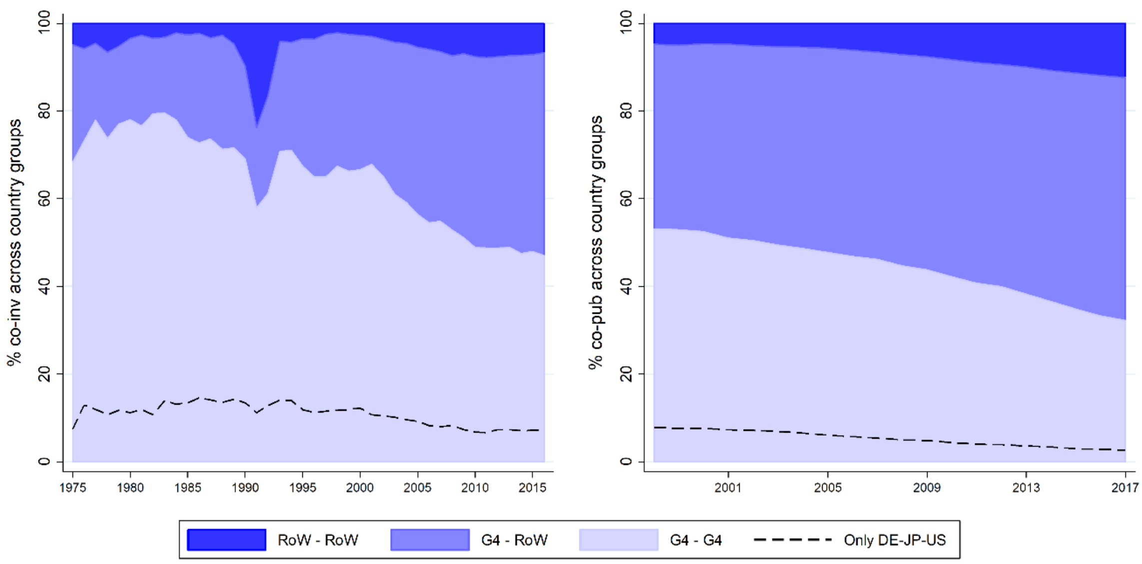
The left panel of **Figure 9** breaks down international patent data by country of the patent applicant, for the top patenting countries worldwide. With the exception of Japan and, to a lesser extent, the Republic of Korea, most top-filing countries show a large international co-inventorship share. The U.S. and Western European countries show a rising trend.

Smaller economies with internationally linked and dense urban and innovation areas – such as Switzerland – are very prone to engage in international collaborations. India also shows a high rate of international co-inventorship. In East Asia's top economies things are different. Up to the mid-1990s, the share of international co-inventorship in China was extraordinarily large, but the volume was small. Thereafter, when the volume of Chinese patenting picked up, the share of international co-inventorship dropped dramatically, although it remained larger than the very low shares of Japan and the Republic of Korea.

Overall, these trends suggest that the globalization of inventive activities mostly concerned the U.S. and Western Europe along with China and India, with China getting less and less self-reliant as it developed its own innovation system (Chaminade et al., 2016; Plechero and Chaminade, 2010). China's dynamics may be in part responsible for the downturn of international co-inventorship after the Great Recession. However, we notice that other big economies, most notably the US, also reduce their share of international co-inventorship after 2010 (France even before that). In their case, it is less likely that this is due to a composition effect, as their national innovation systems were well formed in 2010.

A quick look at country trends for international publication team trends reveals a starkly different picture (right panel of **Figure 9**). Here globalization touches all countries and increase relentlessly. Once again, globalization of science and technology follow different paths.

Figure 10: Co-patenting/co-publishing across country groups (G4 vs Rest of the World)



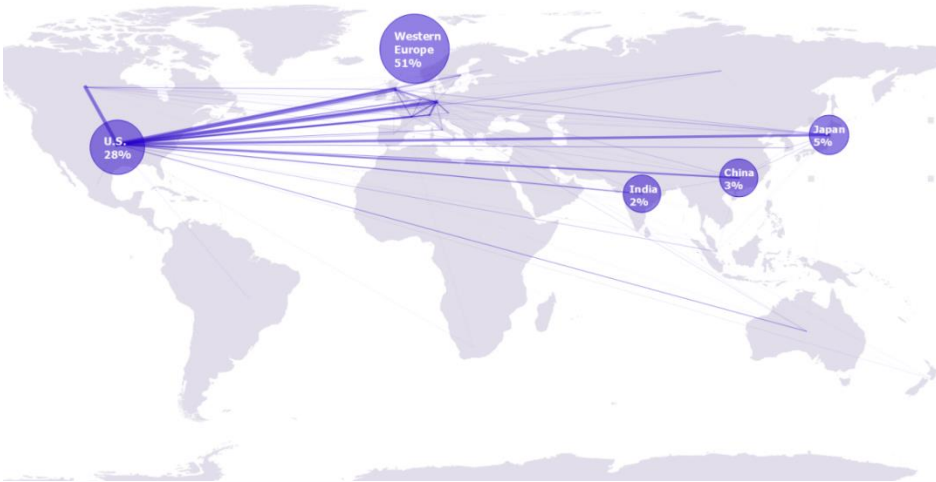
International co-inventorship **International scientific co-publishing**

Source: Authors based on PATSTAT, PCT and Web of Science data.

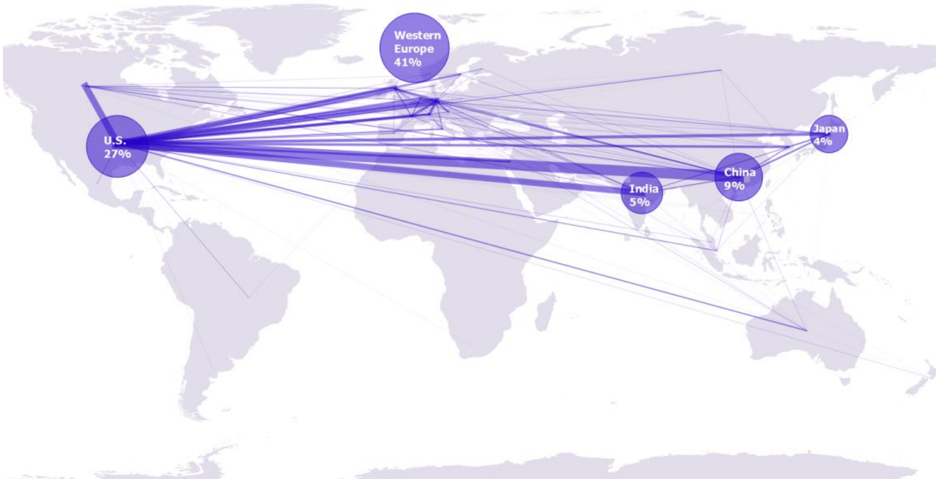
In **Figure 10**, we split the patents and publications by international teams in three groups, namely those whose teams include only inventors or authors from within the U.S., Japan, Canada and Western Europe (as defined above) (G4), inventors or authors from only the rest of the world (RoW), and inventors or authors from both G4 and the rest of the world (RoW) – with a dashed line indicating intra Germany-Japan-US ties.

Figure 11: Concentration and spread of global interactions

International co-inventorship by country pairs

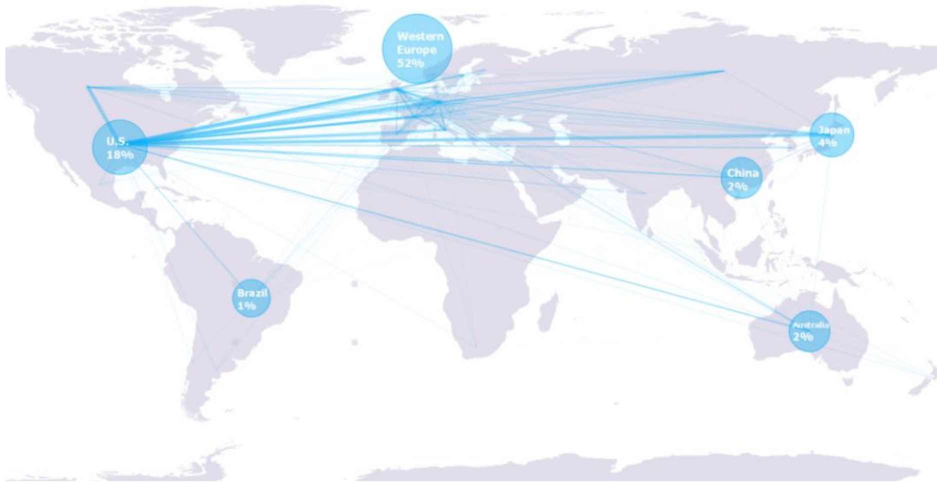


1998-2002

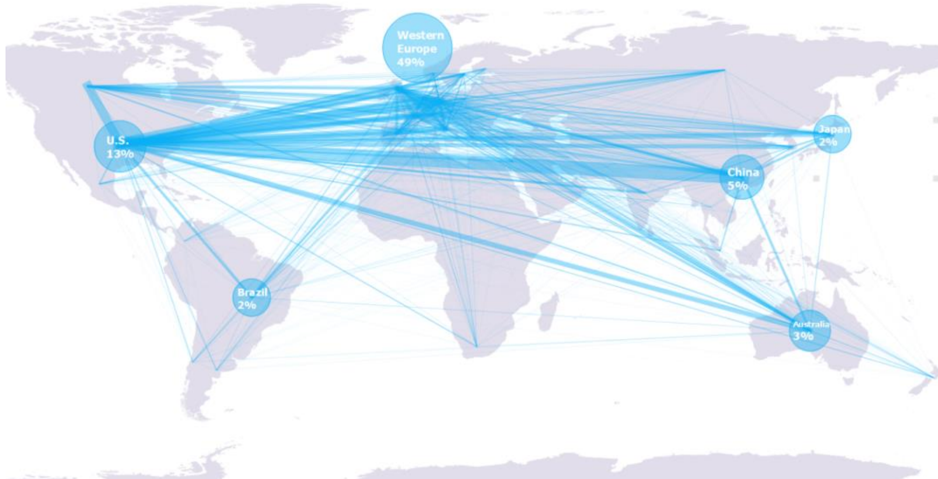


2011-2015

International co-publications by country pairs



1998-2002



2011-2015

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: patents (publications) with more than one inventor (scientists) located in at least two countries. Only top 10% international corridors of each period reported. Bubbles report the share of links only for selected countries and regions.

When examining patents, we notice that, at the beginning of the period analyzed (the 1970s), the G4 countries host the large majority of international teams, almost 80%, followed suit by the combination G4-RoW. But the two groups follow different trends, and G4-RoW collaborations almost catch up with intra-G4 ones by the turn of the century. Teams spanning across countries within the rest of the world remain marginal throughout the period.

Patterns and trends of scientific publications are, once again, rather different. First, the share of intra-G4 international publication teams is way less preponderant and decreases faster than for patents, starting at around 53% in 1998, down to around 32% by the end of the period. The share of teams combining G4 authors and authors from the rest of the world, conversely, increases faster and comes to control over 55% of internationally authored publications. More strikingly, the share of teams comprising authors from the rest of world only follows clearly an upward trend, but remains small compared to the two other.

Figure 11 shows the top-10% largest collaboration corridors, for patents and publications, and for two time-windows. Although concentration of collaborations is decreasing as new stakeholders enter the various collaboration networks, it is only a few countries besides the G4 that explain the trend. “New” entrants such as China, India or Brazil – also mostly link with these economies – take the lion’s share, typically with the U.S. and a few Western European countries, such as U.K. and Germany.

As for collaborations not involving any G4 country, the main players are China, India, Singapore and, to a lesser extent, Australia, Brazil, Argentina, Mexico and South Africa – all of which have increased their participation in the subnetwork, although mostly for scientific co-publication. But their connections still mostly involve one of the big three – particularly the U.S. and Europe – rather than another non-core location.

Overall, the scientific and technological international collaboration trends suggest that the globalization of inventive activities mostly concerned the U.S. and Western Europe along with China and India.

4. Local analysis: Who belongs to the Global Innovation Network?

4.1 City hubs versus niche clusters

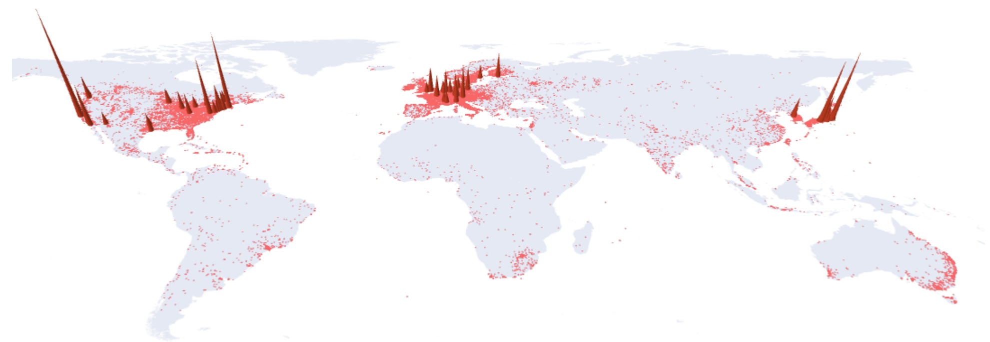
Clusters are the archetypical form of agglomeration/concentration of any type of economic activity, including knowledge production. As defined by Porter (2000), clusters are “geographic concentrations of interconnected companies, specialized suppliers, service providers, firms in related industries, and associated institutions (e.g., universities, standards agencies, trade associations) in a particular field that compete but also cooperate” (Porter, 2000, p. 16).

The intellectual origins of the cluster idea go back as long as Alfred Marshall’s contributions on why industries agglomerate. He defined what went on to be known as Marhsallian districts as an “industry concentrated in certain localities” (Marshall, 1920, p. 268), which he could clearly observe in places like Birmingham, Bedfordshire, Buckinghamshire, Manchester or Sheffield during the second half of the XIX century.

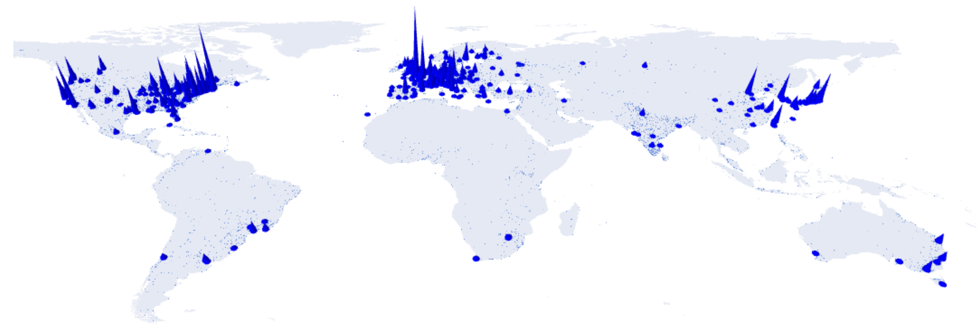
Figure 12: Sticky knowledge in space

Patents per small administrative area

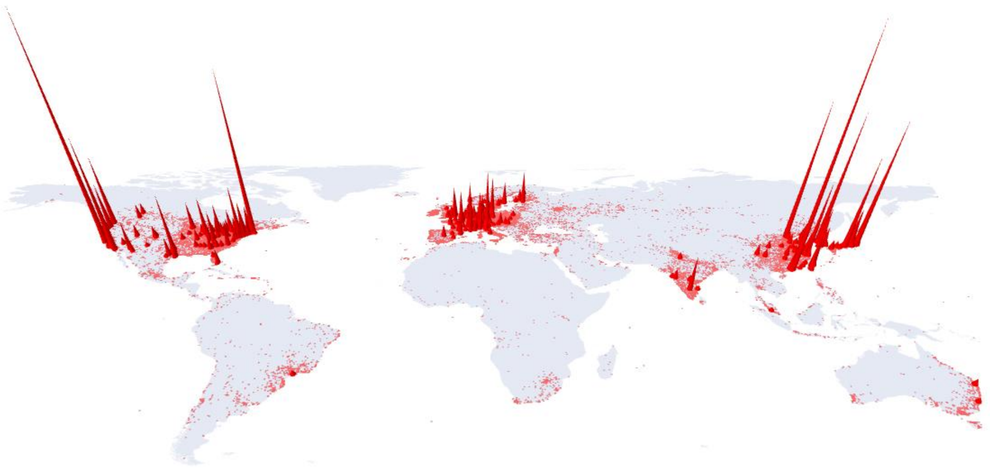
Publications per small administrative area



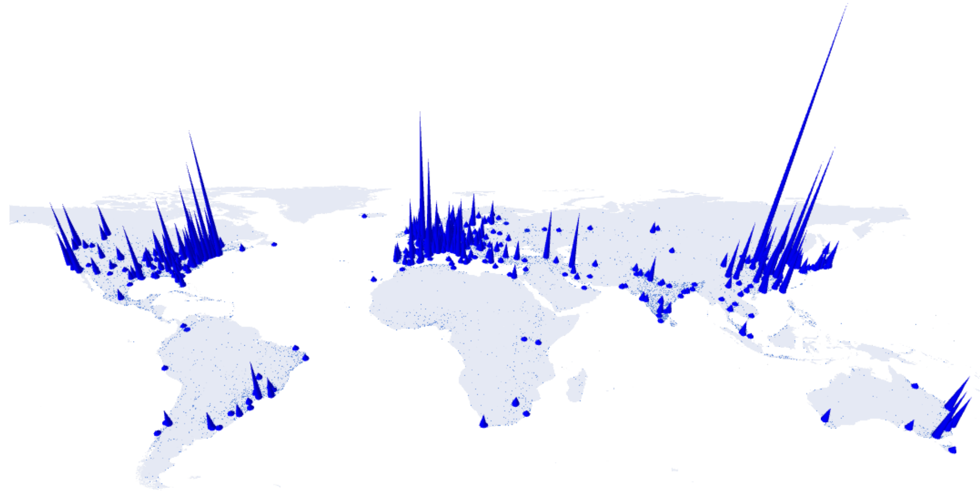
1996-2000



1996-2000



2011-2015



2011-2015

Source: Authors based on PATSTAT, PCT and Web of Science data.

While Marshallian districts still play a key role in many national economies, they may miss the inter-industry connections that characterizes and make successful more diversified clusters (Porter, 2000, p. 21), such as those hosted by large cities and metropolitan areas (Glaeser et al., 1992). The compact nature of cities facilitates sharing and communication, leading to increasing returns of agglomeration economies (Lucas Jr, 1993). In fact, as Jacobs (1969) already noticed, innovation is mostly an urban phenomenon. Jacobs (1969) stressed that, while Marshallian externalities in clusters/industrial districts are mostly intra-industry, the crucial type of spillovers are across industries, allowing cross-fertilization of ideas.

Innovation-based clusters are only a particular case of business clusters, in where the rate of innovation, and innovation of high added value, moves extraordinarily fast. The key element was, once again, the speed by which knowledge and information spread to firms located in the cluster, much faster than firms located outside, and this happened much faster in Silicon Valley compared to Boston. Engel and del-Palacio (2009) and Engel (2015) summarized the concept of clusters of innovation as “global economic *hotspots* where new technologies germinate at an astounding rate and where pools of capital, expertise, and talent foster the development of new industries and new ways of doing business” (Engel, 2015, p.36).

In this section, we explore how various types of agglomerations of inventors and/or scientists have prospered or declined along with the internationalization of science and technology. In the following one we examine to what extent they are connected and participate to the Global Innovation Network.

Figure 12 shows a preliminary view of the distribution of patents and publications, for two different time windows, on a spiky map. Patents and publications are grouped across the smallest available administrative areas within each country, as provided by GADM maps (https://gadm.org). By comparing the two time periods within each figure, we get an immediate sense of both the uneven distribution of inventive and scientific activities within each country, as well as of the emergence of new locations in the 2000s. At the same time, we cannot detect any clear substitution pattern: not only the emerging locations do not displace the incumbent ones, but most of the latter exhibit a marked increase in the number of patents or publications they produce. This holds both at the international level and, at first sight, at the national level.

The administrative units in **Figure 12**, however, do not distinguish between types of agglomeration. Nor they are entirely comparable across countries, at least for two reasons. First, some agglomeration may encompass more than one administrative unit, or sit across them. Second, the spatial and population size of administrative units vary across countries (see Annex 4 for details). For example, the Chinese counties (the smallest units in the country administrative system) are bigger than both U.S. counties and European NUTS3 regions. And even within each country, administrative units’ size can vary considerably, for historic reasons.

For these reasons, we both geolocalize with high precision (exact latitude and longitude) all patents and publications in our dataset, and apply a DBSCAN clustering algorithm to identify a multitude of agglomerations worldwide (Ester et al., 1996). In a nutshell, the algorithm consists on inspecting a small area around each patent or publication (a polygon), searching for other patents or publications therein, and then aggregate all significantly overlapping polygons, irrespective of the administrative units into which they fall (more details in Annex 4). We obtain two types of agglomerations, which we identify by running our DBSCAN algorithm:

1. Global Innovation Hotspots (GIH), which we obtain by considering at once all scientific fields for publications or technological fields for patents. They are large knowledge production centers, capturing the most innovation-dense geographical areas of the world in terms of scientific articles or patent families per square kilometer (km). By definition, these areas are internationally comparable and

geographically distinct. The same scientific publication or patent density determines the same hotspot anywhere in the world, although the threshold is different for scientific publication and patent data. No patent or scientific publication address can be in two hotspots at the same time.

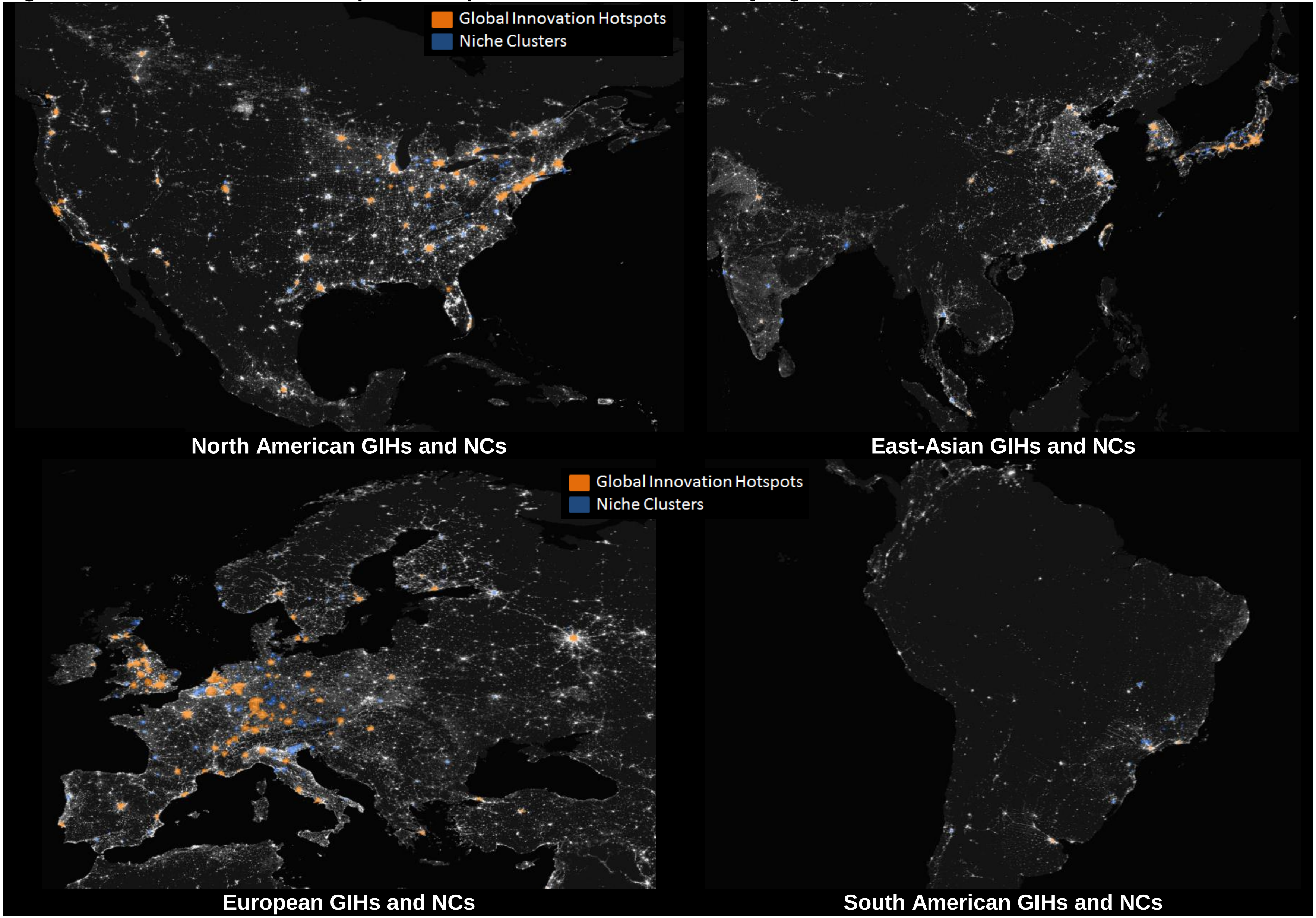
2. Niche Clusters (NC): which we obtain, after identifying the GIHs, by considering all patents and publications not yet assigned to any of them, and treat them separately by technology and field. We created them in order to avoid biases arising from some scientific or technological fields being overrepresented in the scientific publication and patent data, respectively. The NCs capture areas with high innovation density in one or more specific scientific publication or patenting fields, and that otherwise have not met the criteria to be a GIH. The resulting clusters are also distinct geographical areas, as the overlapping clusters for different fields are consolidated into one cluster. But they are only internationally comparable within their specific scientific or technological 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.

Despite being measured independently from administrative units, most GIHs and NCs fall within the largest and/or most prosperous urban areas of the world. However, innovation is more concentrated than both general economic activity and population. **Figure 13** places a large number of GIHs (orange) and NCs (blue) on the maps of selected continents or subcontinents, with night light as background so to compare their presence with actual economic activity. Innovation follows a similar pattern of economic agglomeration, but it does not overlap entirely with it – witness the several bright spots that do not take any orange or blue color.

Only 30 hotspots in 16 different countries are responsible for the creation of almost 70% of the patents and around 50% of the scientific publications produced. More generally, very little inventive and scientific activity is produced outside the GIHs and NCs we identify, and even less outside the few countries hosting them. Indeed, there are more than 160 countries not hosting any hotspot or niche cluster. The skewness exists also in these less innovation dense areas, as most of the knowledge is produced in few urban dense areas. Just 30 agglomerations located in only 24 different countries produce around 64% of patents and 61% of scientific publication outcomes within these non-innovation dense countries (**Table 2)**. Despite the concentration in these few agglomerations, the gap with the top hotspots is huge. The volume of patents and scientific publication of the top 30 agglomerations is only 0.4% and 4% of those of the top 30 hotspots, respectively.

Figure 13: Global Innovation hotspots and specialized niche clusters, by region



Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: Night light data from the U.S. National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center.

Table 2: Concentration of patenting and publishing among GIH and NC, and among less innovation dense countries, 1998-2017

Top 30 hotspots (as share of all GIHs in the world)		
Hotspots (%)	30	(17.2%)
Countries (%)	16	(47.1%)
Patents (%)	3,234,850	(69.2%)
Scientific articles (%)	10,987,971	(47.8%)
Top 30 agglomerations in non-innovation dense countries		
Agglomerations (%)	30	(5.0%)
Countries (%)	24	(14.4%)
Patents (%)	11,491	(64.1%)
Scientific articles (%)	484,689	(61.0%)

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: Only data from 1998 to 2017 reported. Top 30 is calculated separately for patent and publication data. Top 30 agglomerations in non-innovation dense countries are based on the same methodology described for GIHs (annex 4).

Table 3: Largest GIH and NC, by number of patents

GIH		NC		GIH		NC	
Name	**199195**	**Name**	**199195**	**Name**	**2011115**	**Name**	**2011115**
Tokyo	105393	Friedrichshafen	1433	Tokyo	234232	Suzhou	3182
Osaka	38165	Bern	1284	Seoul	87494	Friedrichshafen	2981
Nagoya	14523	Würzburg	1181	Osaka	78153	Cheonan	2800
Frankfurt	14300	Okayama	1024	Shenzhen-HK	57321	Würzburg	2412
Paris	13068	Rosenheim	806	SJ-San Francisco	51957	Qingdao	2165
Köln-Dusseldorf	12010	Kōfu	768	Nagoya	43088	Hartford	2147
Seoul	9227	Toyama	645	Beijing	30394	Bern	2069
Stuttgart	8338	Iwakuni	633	New York City	25009	Bielefeld	1850
SJ-San Francisco	6725	Schwäbisch G.	620	Boston	21918	Hyderabad	1806
London	6698	Bielefeld	612	Frankfurt	21233	Kaohsiung	1795

Source: Authors based on PATSTAT, PCT and Web of Science data.

Table 4: Largest GIH and NC, by number of publications

GIH		NC		GIH		NC	
Name	**199802**	**Name**	**199802**	**Name**	**2011115**	**Name**	**2011115**
Tokyo	184035	Kanazawa	14058	Beijing	282885	Changsha	36715
Washington-Baltimore	148130	Iowa City	13496	New York City	249036	Changchun	33989
New York City	142845	State College	13391	Tokyo	238489	Harbin	33425
London	124926	Strasbourg	13040	Boston	231877	Hefei	32268
Boston	123169	Ithaca	12958	Washington-Baltimore	228408	Jinan	31194
Paris	117831	Marburg	12657	London	210969	Chongqing	28437
SJ-San Francisco	106614	College Station	12263	Paris	184104	Kuala Lumpur	28413
Osaka	94043	Saint Petersburg	12203	Seoul	183014	Gent	27504
Los Angeles	77606	Brentwood	12147	SJ-San Francisco	174205	Padova	26429
Amsterdam	71431	Padova	11886	Amsterdam	143607	Cairo	24657

Source: Authors based on PATSTAT, PCT and Web of Science data.

The U.S. hosts the majority of GIHs and NCs (25%), followed by Germany (12.9%), Japan (9.4%), China (6.8%), the U.K. (4.9%) and France (4.3%). By continent, Europe concentrates 40.5% of GIH and NC, followed by North America (28%), Asia (25%), Latin America (2.9%), Oceania and Pacific (2.7%), and Africa (1%). Their average area is around 1,704.9 square km (3,013.44 square km for GIH and 977.45 square km for NC), and the average population is 1,949,350 inhabitants in 2015 (3,794,694 for GIH and 923,504.4 for NC). On average, they have produced 12,881.7 patents from 1976 to 2015 (31,782.22 for GIH and 2,636.6 for NC), and 67,496.09 publications from 1998 to 2018 (31,782.22 for GIH and 2,636.6 for NC). **Table 3** and **Table 4** show the top-10 GIH and NC, in two different time windows, separately for patents and for scientific publications (longer lists and time periods are available in the annex). Clusters are labelled according to the name of the largest city they host (as measured by the number of inhabitants in 2005), with some manual double-checking (data from www.geonames.org).

While GIHs usually coincide with large urban areas, not all such areas host a GIH, or at least a major one. GIHs in Beijing, London, Los Angeles, New York, Seoul, and Tokyo concentrate a large amount of both patents and scientific articles, those related to Buenos Aires, Delhi, Istanbul, Mexico City, Moscow, Sao Paulo, and Tehran have a much smaller scale. Others, such as Cairo, Bangkok, Kolkata, and Chongqing, conversely, host a NC. Finally, a certain number of highly populated metropolitan areas – such as Jakarta, Karachi or Manila – do not host any knowledge agglomeration at all.

Table 5: 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

Source: Authors based on PATSTAT, PCT and Web of Science data.

On the other hand, some less dense urban areas in high-income and innovative countries host a large number of important NCs. **Table 5** and **Table 6** rank, respectively, GIH and NC by their share and patents (publications) in selected fields. While, as expected, GIHs always come on top, some NCs (shaded in grey) do better than many GIHs in certain technological or scientific fields). For instance, the NC

Rennes produce more patents in Audio-visual technologies than other GIH in that field, such as Amsterdam, Singapore, Ann Arbor, Stockholm, Cambridge or Copenhagen. The NCs Bern or Mumbai do better in biopharma patents than GIHs such as Singapore, Taipei, Nurnberg, Rochester or even Bengaluru.

The NC Ithaca, in the US, do better in producing publications in applied biology compared to other big GIH, such as Brisbane, Köln-Dusseldorf, Munich, Seattle, Frankfort, Milan or Guangzhou. In Earth Sciences, the NCs Canberra, Honolulu or Santiago de Chile are better positioned than Grenoble, Newcastle, or Austin. All this indicates that, despite NC being, on average, smaller than most GIH, they can and do better than the latter ones in certain fields, in which the NC might be relatively more specialized.

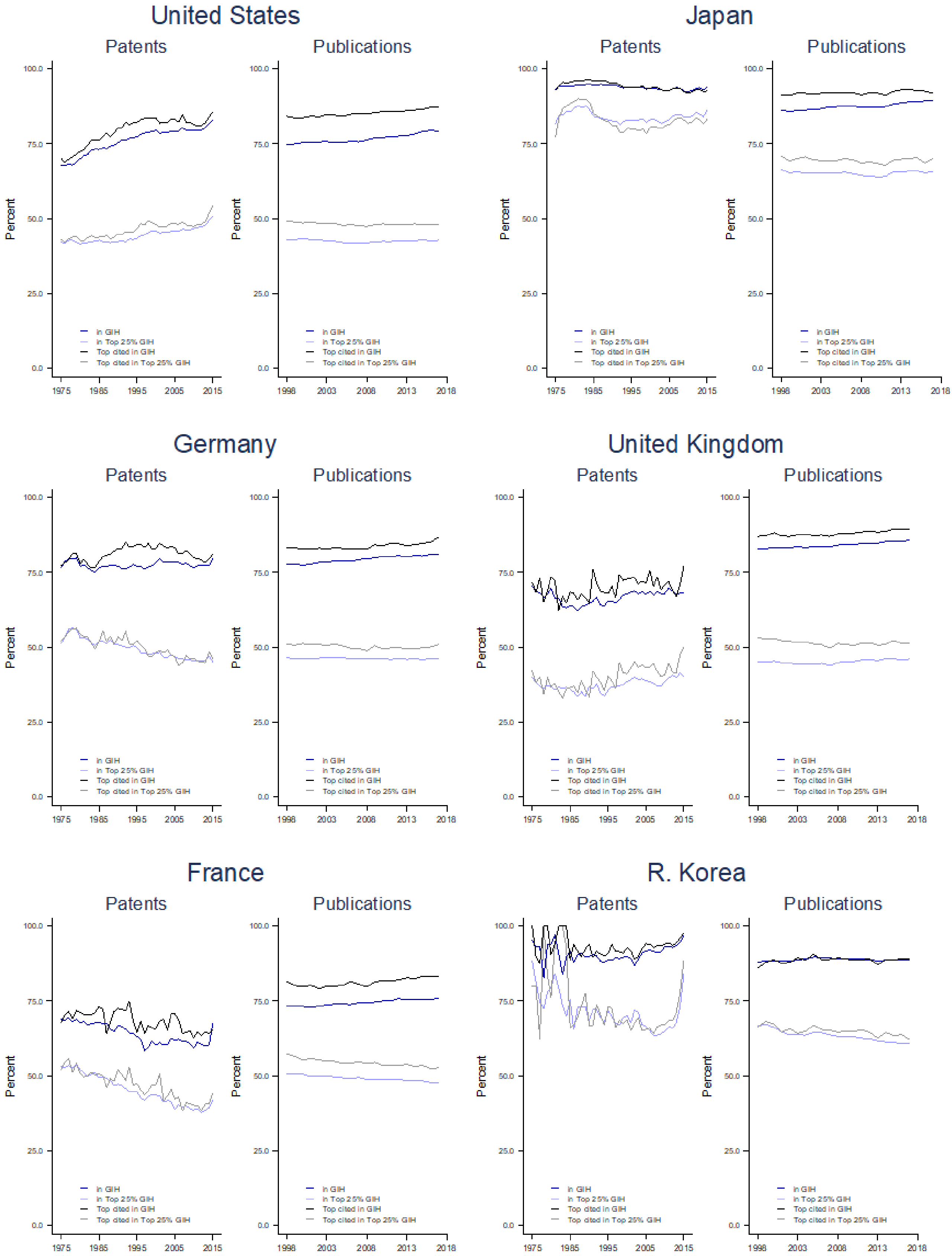
Table 6: Ranking GIH and NC, for selected fields, scientific publications

Applied Biology			Earth Sciences		
Position	Share pub.	Cluster name	Position	Share pub.	Cluster name
1	2.45%	Tokyo	1	3.15%	Beijing
2	1.93%	Washington-Baltimore	2	2.46%	Washington-Baltimore
3	1.87%	Beijing	3	2.09%	Paris
4	1.49%	London	4	2.00%	Tokyo
5	1.48%	New York City	5	1.87%	San Jose-San Francisco
6	1.35%	Paris	6	1.65%	Los Angeles
7	1.17%	Boston	7	1.52%	Boston
8	1.15%	Osaka	8	1.48%	New York City
9	1.13%	San Jose-San Francisco	9	1.35%	Amsterdam
....			....		
31	0.63%	Ithaca	57	0.45%	Canberra
32	0.62%	Mexico City	58	0.44%	Honolulu
33	0.62%	Brisbane	59	0.43%	Santiago
34	0.61%	Köln-Dusseldorf	60	0.43%	Nijmegen
35	0.61%	Munich	61	0.42%	Bologna
36	0.60%	Montpellier	62	0.42%	Grenoble
37	0.60%	Seattle	63	0.41%	Prague
38	0.58%	Nanjing	64	0.41%	Newcastle
39	0.57%	Mannheim	65	0.40%	State College
40	0.55%	Milan	66	0.40%	São Paulo
41	0.54%	Guangzhou	67	0.40%	Austin

Source: Authors based on PATSTAT, PCT and Web of Science data.

When it comes to examining the behavior of knowledge agglomerations in relation to the growing internationalization of innovation, we notice immediately that, in the majority of the high-income and emerging countries, a limited number of GIHs and NCs continued to produce the same share or even more of the knowledge produced nation-wide.

Figure 14: Share of patenting and publishing in clusters, by selected countries



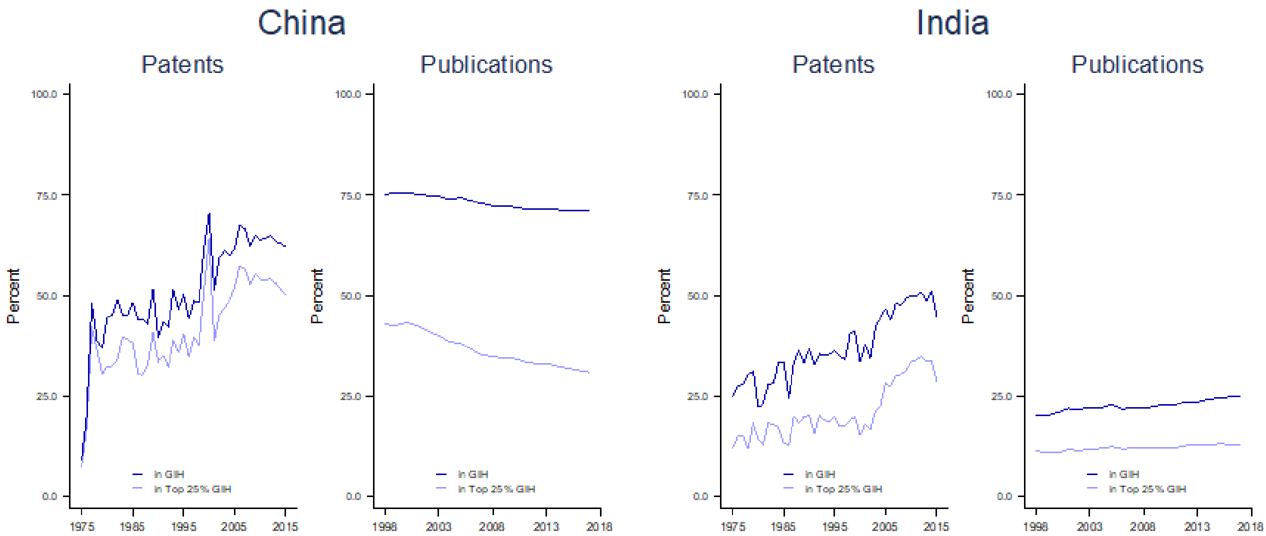
Source: Authors based on PATSTAT, PCT and Web of Science data.

In **Figure 14** we look at the cumulative patents that GIHs host over time (by country). We draw blue lines for the total GIHs and for the GIHs in the top 25% of size rankings. We see, for instance, that the share of patents produced in GIHs in the U.S. goes from 63% to 71% (71% and 78% respectively if we were adding NCs). We also draw grey lines for the distribution of the top-10% most cited patents per year and technology. We clearly observe that highly-cited, highly-valuable patents are systematically more concentrated than the average patents. We see similar patterns of increasing concentration for R. Korea, the U.K. and, especially, China and India (**Figure 15**, for China and India, where top-cited patents concentration is removed). The evolution in Germany is more stable (though with a slight increase in recent years), and stable or even decreasing in France.

Lighter blue and grey lines show which part of this share is attributable to the top-25% GIHs, per country. The share of the most productive clusters has increased over time in all countries shown, except for France and Germany.

The right panels of **Figure 14** and **Figure 15** reproduce these same figures for the case of scientific publications. Again, we do not see a spatial spread of scientific activities over time in most countries (except for France and, especially, China), and even see slight increases in recent years for some countries.

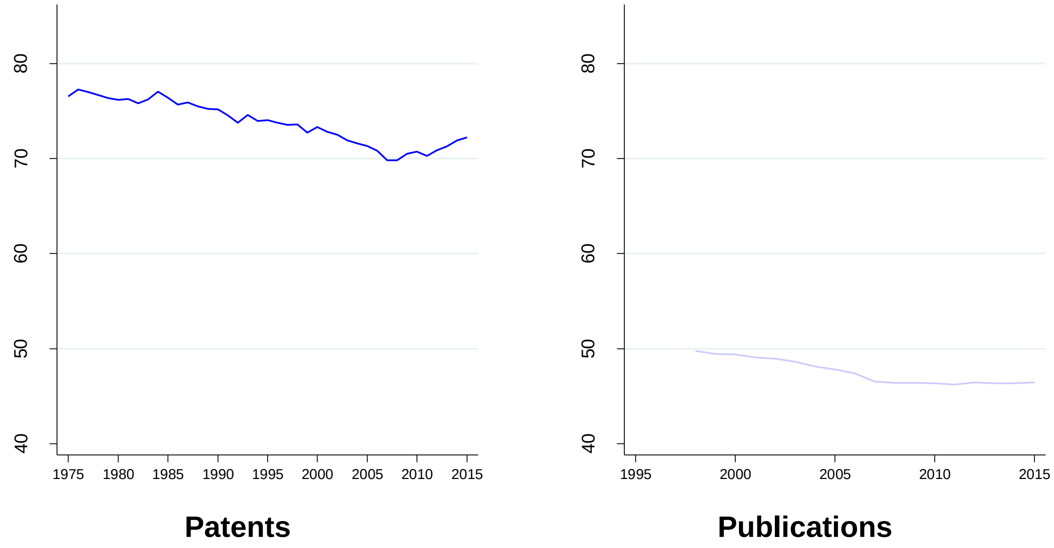
Figure 15: Share of patenting and publishing in clusters, China and India



Source: Authors based on PATSTAT, PCT and Web of Science data.

In **Figure 16**, we look at the time evolution of unequal contribution of clusters. In this case, both GIHs and NCs are included in the analysis. As shown in **Figure 16** (left panel), despite only 10% of GIHs and NCs (49) concentrate 70-80% of all patent production, the concentration has gone down from almost 80% in the 1980s to around 70% at the end of the period. There is a slight increase of concentration in the last years, though one would need to wait some extra years to see whether there is indeed a change of tendency. Right panel finds the same decreasing trend for publications, but starting from way less concentration levels (less than 50% throughout the whole period).

Figure 16: Concentration of patenting and publishing among clusters



Source: Authors based on PATSTAT, PCT and Web of Science data.

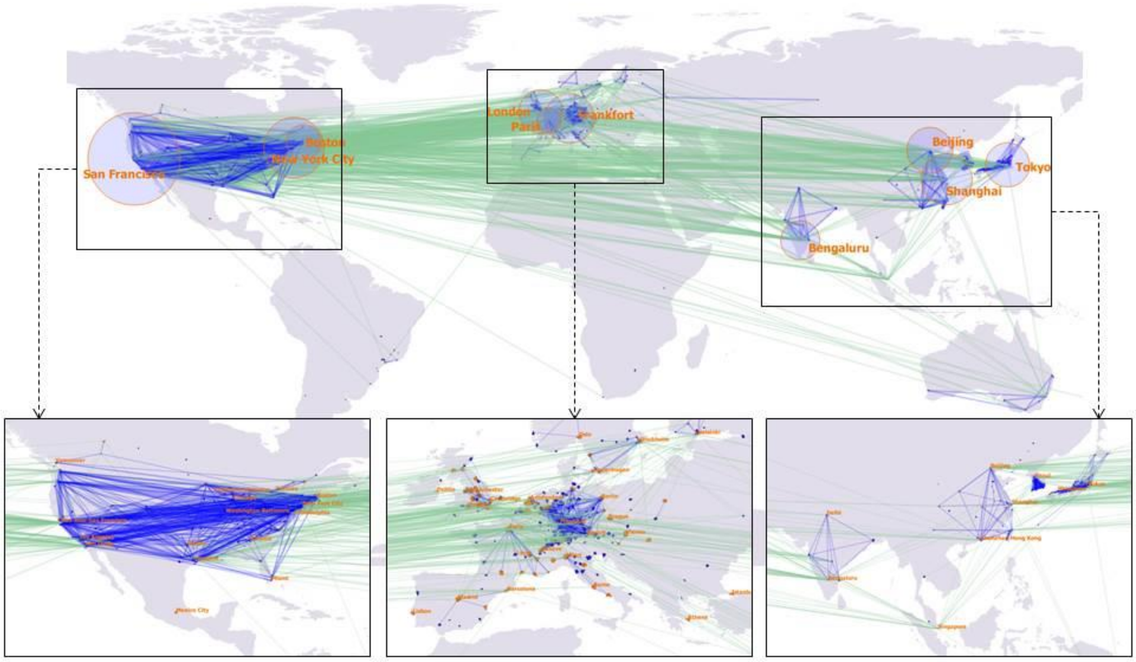
4.1 A global network of GIH and NC

The GIHs and NCs we detected form a veritable GIN. **Figure 17** reports GIHs as orange nodes and NCs as blue ones, linked one another by national ties (blue) and international ones (green). The thickness of the links represents the amount of bilateral collaborations between clusters, measured using co-inventorship data. In these maps, only the top-10% largest links are depicted, between 2011 and 2015.

We notice a thick web of ties between clusters in the US, Europe and Asia, as well as among clusters within countries (especially within the US). However, it is difficult to appreciate the structure of the relations of the network in such a zoomed-out map. For this reason, we zoom-in below.

In the US, it seems that large GIHs tend to collaborate both nationally and internationally. Smaller GIHs or NCs seem to be more specialized in national interactions. In Europe a similar pattern can be detected, with certain clusters in each country acting as gatekeepers that connect the national innovation system to the GIN. Clear examples can be found in France, with Paris connecting other French cities with the rest of the world, followed at some length by Lyon and Grenoble or the UK, with London being the central actor. Germany shows some hierarchical structure too, though access points to the GIN are more numerous. In China, this hierarchical structure is also evident, with Shanghai, Beijing, and Shenzhen acting as the top gatekeepers. For completeness, **Figure 18** repeats the same analysis for publication data.

Figure 17: GIH, NC, and the Global Innovation Network, patent data 2011-2015



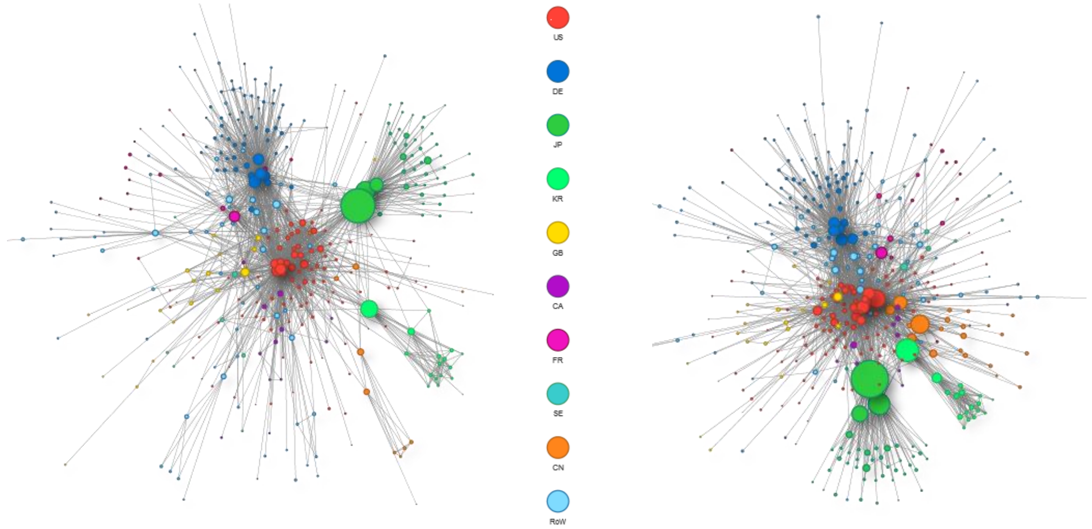
Source: Authors based on PATSTAT, PCT and Web of Science data.

Figure 18: GIH, NC, and the Global Innovation Network, scientific publication data 2011-2015

Source: Authors based on PATSTAT, PCT and Web of Science data.

When drawing proper network graphs, as in **Figure 19** (in two time periods), we can appreciate the position of individual GIHs or NCs in the network. An innovation agglomeration is more "central" within this global network the more international connections it concentrates. The nodes are located in space depending on their connectivity: better connected nodes (in terms of the number of other nodes to which the focal node connects) are more central, and their size is determined by the total number of patents (rescaled). As can be seen, large U.S. clusters are at the center of the graph, which means they are the most connected and central. The center of the picture also hosts other GIHs which are arguably highly connected, such as Tokyo, London, Shanghai, Beijing, Seoul, Paris, etc. Smaller clusters in each country are connected to the big and highly connected ones in their own country, showing this hierarchical pattern we discussed early (one can clearly see this for Japan, R. Korea and the U.K.). On the other hand, several hotspots that are larger or similar in size to the top U.S. agglomerations – for example, Tokyo – fail to occupy the same kind of central position in the global network.

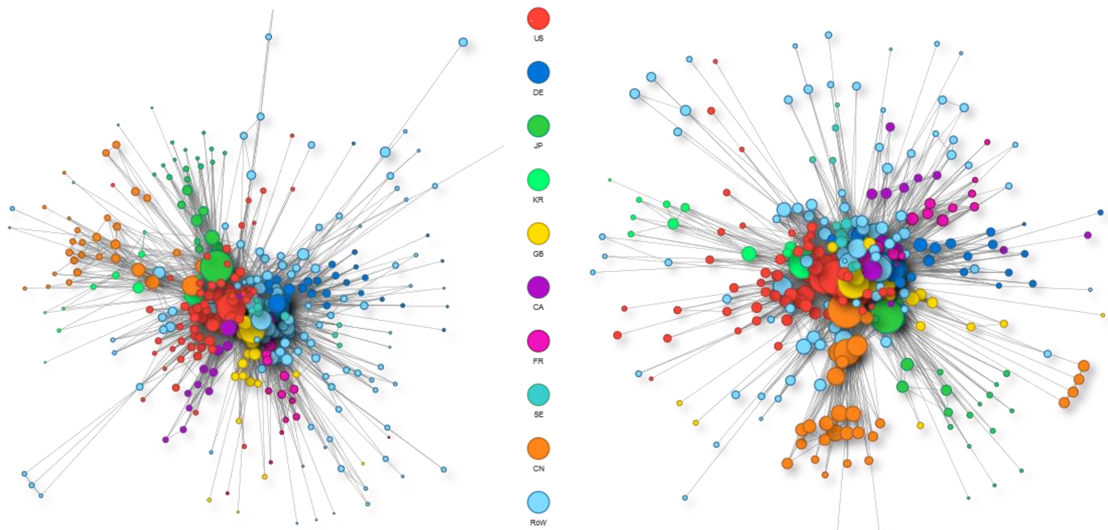
Figure 19: GIH and NC positions in the GIN, patents



Patent co-invention network 2001-2005 | **Patent co-invention network 2011-2015**

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: Only the world's 10% largest links reported. Bubble size reflects patent volume. Bubbles positioned according to their network centrality.

Figure 20: GIH and NC positions in the GIN, publications



Publication co-invention network 2001-2005 | **Publication co-invention network 2011-2015**

Source: Authors based on PATSTAT, PCT and Web of Science data. Notes: Only the world's 10% largest links reported. Bubble size reflects patent volume. Bubbles positioned according to their network centrality.

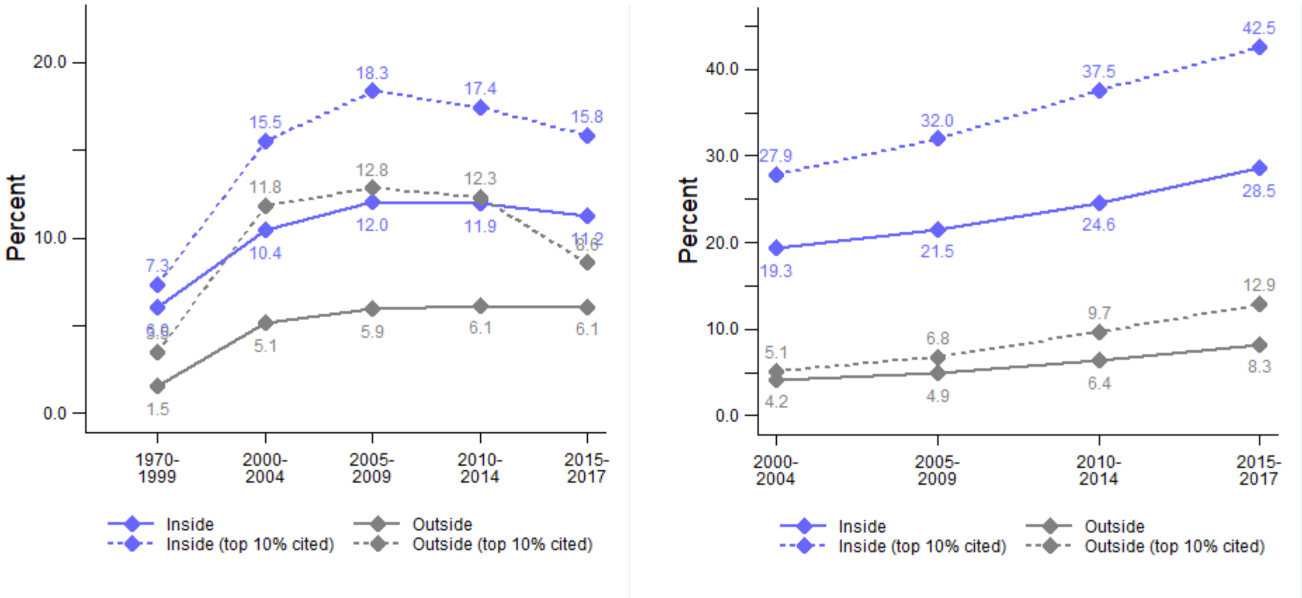
The scientific network (for the two periods) is slightly different (**Figure 20**). Many other nodes from the Rest of the World are as big and important in the network (they are located in the center of the graphs) as countries of the 9 countries highlighted. Moreover, nodes from many countries are located in the center, which indicates that the network of clusters in scientific production is way less hierarchical.

4.2 The globalization of GIH and NC

As discussed earlier, R&D-intensive MNCs have prioritized knowledge-seeking foreign expansion strategies; to a lesser extent, the same applies to large universities and public research organizations. Such strategies put them into contact with GIHs and NCs around the world, and at the same time provide links between them. As MNCs wishes to tap into global knowledge and talent pools, knowledge clusters become the most important location factor of innovation-oriented MNCs (Cantwell et al., 2010; McCann and Mudambi, 2005).

But the relation between GIHs, NCs, MNCs and research organizations work in two ways. Successful clusters cannot be self-sufficient in terms of the knowledge base they draw upon, and the organizations within them (whether firms, universities, public laboratories or other actors) deliberately build international links to complementary pools of knowledge abroad, which would not otherwise be available locally (Awate and Mudambi, 2017; Bathelt et al., 2004; Lorenzen and Mudambi, 2013; Turkina and Van Assche, 2018).

Figure 21: Percentage of international patent and publication teams, inside vs outside GIHs and SNCs



Patents **Scientific publications**

Source: WIPO based on PATSTAT, PCT and Web of Science data. Notes: Shares of the total patents and scientific publications, respectively.

To what extent the internationalization process we examined in section 3 did impact on the GIN of GIHs and NCs? In **Figure 21** we split patents in two groups

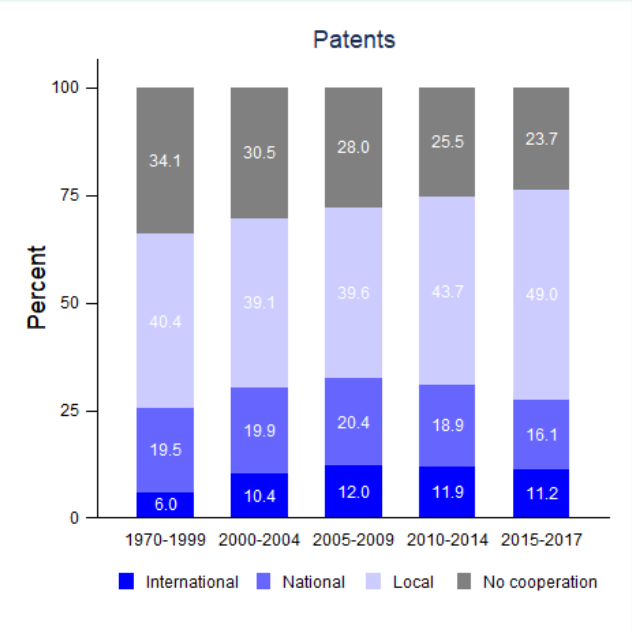
– the patents produced by clusters, which we defined as those whose teams count at least one inventor address in a GIH or NC,

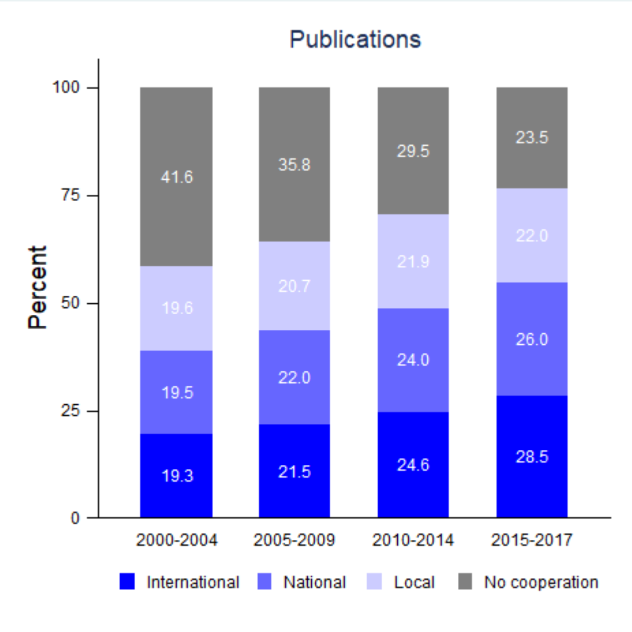
– the patents produced outside clusters (no inventor address to be found in any cluster)

We see that patents produced in GIHs and NCs tend to be more internationalized than those produced outside them. The difference is even starker when focusing only on highly cited patents (10% more cited patents, per year and technology). At the same time, though, the trend inversion of recent years is more pronounced for clusters' patents, which would point to some evidence that more complex knowledge needs to be produced among geographically close researchers.

Figure 21, right panel, repeats the same approach, but for scientific publications. Scientific publications produced in clusters are, on average, more internationalized that their non-cluster counterparts. Again, the pattern for highly-cited publications is the same. We see, however, larger globalization for the case of publications (and larger differences between clustered and non-clustered publications than what we see for patents), and we do not observe the decrease on internationalization of teams observed for patents.

Figure 22: GIHs' and SNCs' share of co-inventorship and co-publication interactions, by partner location





Patents **Publications**

Source: WIPO based on PATSTAT, PCT and Web of Science data.

Figure 22 looks at the globalization of GIH and NC. We calculate the share of either patents or publications produced in them in which:

1. the team includes at least one member located outside both the cluster and the country to which the cluster belongs (international)
2. the team includes at least one member located outside the cluster, but no foreign ones (national)
3. the team includes only local (within-cluster) members (local)
4. The team is formed by a single inventor/author

Several patterns are worth reporting. The percentage of scientific and inventive output in these innovation-dense agglomerations that does not involve any local, national or international collaboration has decreased. Inventions with a single inventor went from one-third in the 1970s and 1980s to less than a quarter. Scientific publication by a sole author went from more than 40 percent in the early 2000s to less than 25 percent in the second half of the 2010s. The more the hotspots and niche clusters collaborate, the denser the network of knowledge they create.

In other respects, the picture differs depending on whether it is scientific or inventive output. For patents, the share of local-only teams is larger than that of national and international ones, while this is not the case for scientific publications. Nevertheless, for scientific publications, international co-publication has continuously grown faster than national and local collaborations. The same trend is observed for patents from the early 1980s until the second half of the 2000s.

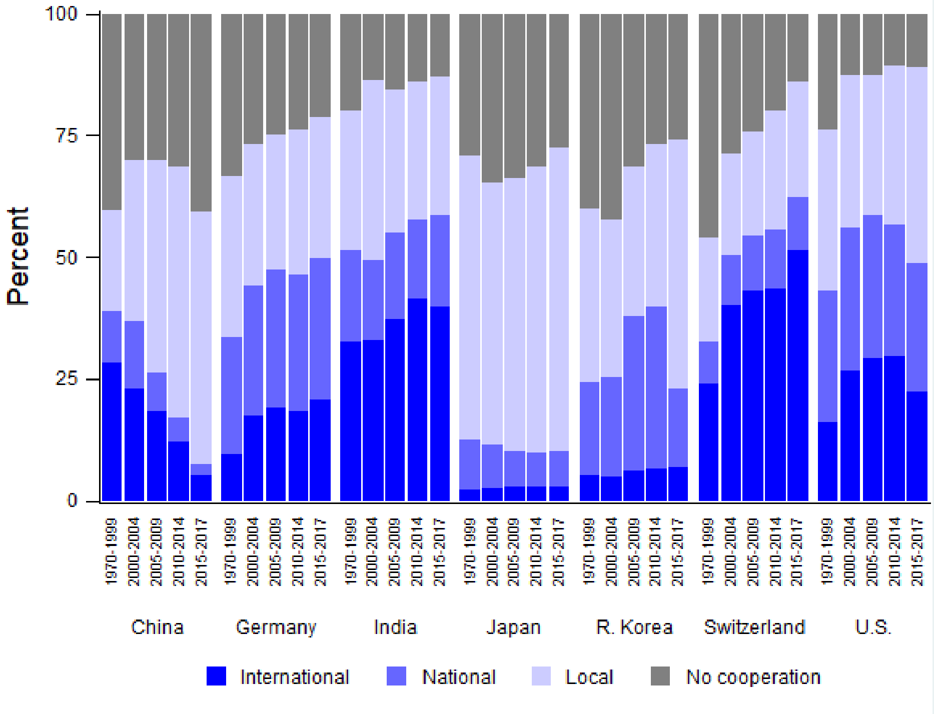
Since around 2005, however, there has been a fresh rise in the share of local-only patents. The change coincides with a slowdown in the pace of globalization and internationalization generally, as reflected in slower growth of trade, FDI flows and financial integration. It also coincides with a decrease in the share of patents generated by teams that are national but not just local. The explanation for the latter could be that part of the slowdown in the globalization of knowledge creation and innovation has to do with the rise of local hotspots rather than with the development of new national innovation systems. For scientific publication, however, the trend has been one of continued internationalization.

Figure 23 provides details of the same trends at the country level, for a selection of high-income and emerging economies, respectively for patents and publications. We consider five time periods for patents, finishing in 2017; and four time periods for publications, from 2000 to 2017.

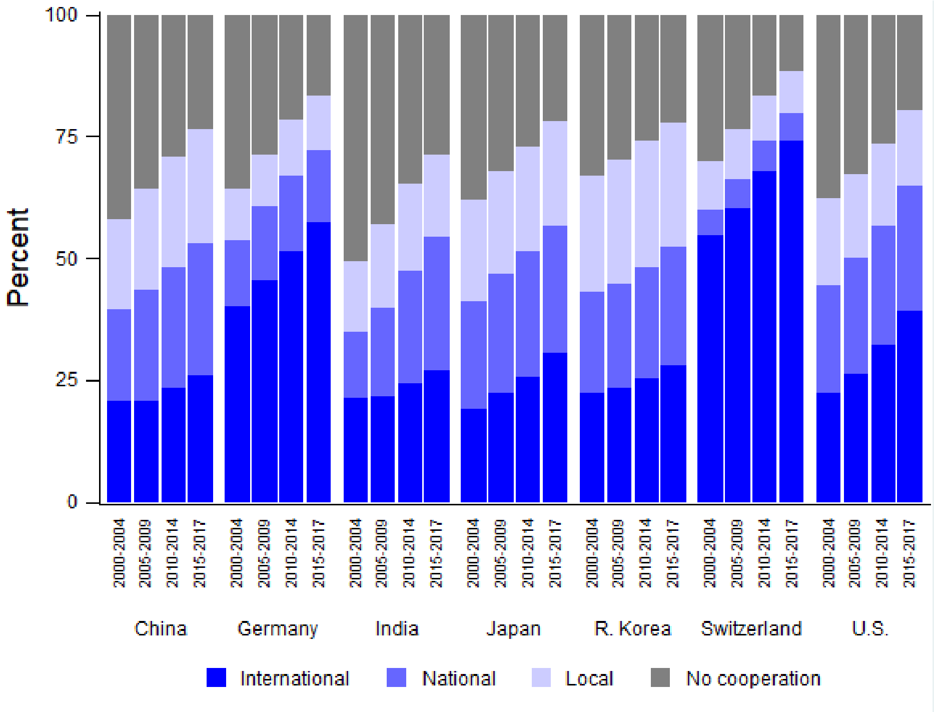
The degree of clusters' internationalization varies considerably across countries, very much in line with what we observed in sections 2 and 3, with countries like India or Switzerland being most open, and R. Korea and Japan at the other extreme. In line with evidence in sections 2 and 3, we notice a loss of weight of international teams for countries like China in recent years. However, in the majority of the cases the pattern in the last years is that of stagnation or slight growth. We see several cases in which the share of national teams is reduced, like China, and the U.S. to some extent. Interestingly, the slight decrease in international teams is absorbed by local teams, whose share is larger for some countries in the last period.

As reported above, we do not see downturn or stagnation in international co-publication patterns (rather the opposite) – **Figure 22**. Moreover, we see that in all countries the increase in international teams is absorbed by local teams or individual authors, that see their shares systematically reduced in all countries analyzed.

Figure 23: GIHs' and NCs' share of co-inventorship and co-publication interactions, by partner location, selected countries



Patents



Publications

Source: WIPO based on PATSTAT, PCT and Web of Science data.

5. Conclusions

In this paper, we have exploited a rich dataset of patent applications and scientific publications in order to answer several questions concerned with two current phenomena on the way knowledge is produced and shared worldwide: its geographical spread at the international level and its spatial concentration in few worldwide geographical hotspots. We recap the main messages of the paper as follows.

First, we indeed observe that the production of patents and scientific publications has spread geographically to several countries, and has not kept within the traditional knowledge producing economies (Western Europe, Japan and the US). This meant already an achievement, as knowledge-related phenomena such as patenting, scientific production, R&D investments and so on have been always way more concentrated than other pillars of globalization, such as trade or FDI. This has started to change, and concentration of the former has started to converge towards the latter.

Second, in recent years we see an upturn degree of concentration in some variables (notably, R&D expenditures and patents, particularly those highly-cited and belonging to high-tech sectors). The same occurred in some technological fields and for highly cited patents in the 1980s. Both processes could be related to the appearance of certain breakthrough, highly-cited technologies in those years, that remain concentrated in central countries at the beginning, and only disseminate after some time and contribute to dispersion.

Yet, by the end of the period, Western Europe, Japan and the U.S. concentrated around 70% of internationally-oriented patent activity, and more than 50% of all scientific activity, which is a lot. In fact, it seems that the large majority of knowledge production spreading is due to a handful of emerging economies, notably China. In the meanwhile, large areas of the world, notably in Africa and Latin America, are left out of the process of knowledge globalization.

Part of this partial geographical spread of knowledge activities is due to the setting up of GINs and other knowledge networks, first toward more traditional innovative countries, and then towards emerging economies too. However, networks among innovation-core countries dominate, with an increasing share of them between the latter and few emerging economies (with innovation networks among countries outside the Triad being relatively marginal). This is less the case when it comes to scientific publications, where the role of emerging economies, or even networks among non-traditional hubs are on the raise.

Overall, we have seen an increasing degree of globalization of knowledge production (that is, the formation of international teams in producing ideas). Interestingly, we see some stagnation of co-inventorship networks in recent years (which we do not see in publications). Part of this slight decrease is due to China and other emerging economies. As they start competing with more advanced economies in producing innovation, their need to rely in global networks is reduced – thus contributing negatively to the globalization of innovation.

Yet, this does not seem to explain the whole picture, as some other large innovation economies see their share of international teams to stagnate. One possible explanation is the decline of multilateralism after the Great Recession, which could be behind this downturn. An alternative explanation is the increasing degree of complexity of innovation, which requires teams to be less geographically spread out.

Another important message of this paper relates to the geographical distribution of knowledge production within countries (both in traditional knowledge producers as well as in emerging ones). Despite the increasing worldwide spread of knowledge production, we do not see the same spreading process within countries, and even we see some increased concentration in some of them. This may have, of course, important distributional consequences within countries, which will need to be addressed properly in the near future.

Not only these areas, which we have identified as GIHs and NCs, keep concentrating a larger amount of idea production. We have provided evidence that they also concentrate a large and increasing connectivity, within their own country to other hotspots, and across countries through GINs. Again, this goes against lagging behind areas of these countries, that not only produce less innovation, but also lack the necessary connectivity to the outside world to avoid being lock-in in non-innovative development paths. In fact, we have seen that when it comes to external patent collaborations, clusters' globalization has never slowed down: team members external to the cluster are increasingly located abroad, whether in other clusters or not. This reinforces the message that the main reason behind the overall reversal of globalization is the increasing self-sufficiency of many clusters, not a comeback of national innovation systems or the rise of new ones.

Finally, we have provided some evidence that, even within GIHs, inequalities arise, with few of them being truly innovation and connectivity hubs, for both technology developments and scientific advancements, and others having a more subordinate position to these hubs (usually with one or few global hotspots per country, and other clusters connecting internationally through the former one).

References

Alcácer, J., Zhao, M., 2016. Zooming in: A practical manual for identifying geographic clusters. Strategic Management Journal 37, 10–21. https://doi.org/10.1002/smj.2451

Alcácer, J., Zhao, M., 2012. Local R&D Strategies and Multilocation Firms: The Role of Internal Linkages. Management Science 58, 734–753. https://doi.org/10.1287/mnsc.1110.1451

Amendolagine, V., Chaminade, C., Guimón, J., Rabellotti, R., 2019. Cross-border knowledge flows through R&D FDI: Implications for low-and middle-income countries. Lund University, CIRCLE-Center for Innovation, Research and Competences in ….

Awate, S., Larsen, M.M., Mudambi, R., 2012. EMNE catch-up strategies in the wind turbine industry: Is there a trade-off between output and innovation capabilities? Global Strategy Journal 2, 205–223.

Awate, S., Mudambi, R., 2017. On the geography of emerging industry technological networks: the breadth and depth of patented innovations. Journal of Economic Geography 18, 391–419.

Bathelt, H., Malmberg, A., Maskell, P., 2004. Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. Progress in Human Geography 28, 31–56. https://doi.org/10.1191/0309132504ph469oa

Bergquist, K., Raffo, J., Fink, C., 2017. Identifying and ranking the world's largest clusters of inventive activity (No. 34), WIPO Economic Research Working Papers. World Intellectual Property Organization - Economics and Statistics Division.

Branstetter, L., Li, G., Veloso, F., 2014. The Rise of International Co-invention (NBER Chapters). National Bureau of Economic Research, Inc.

Breschi, S., Lissoni, F., Miguelez, E., 2017. Foreign-origin inventors in the USA: testing for diaspora and brain gain effects. J Econ Geogr 17, 1009–1038. https://doi.org/10.1093/jeg/lbw044

Cantwell, J., Dunning, J.H., Lundan, S.M., 2010. An evolutionary approach to understanding international business activity: The co-evolution of MNEs and the institutional environment. Journal of International Business Studies 41, 567–586.

Castellacci, F., Archibugi, D., 2008. The technology clubs: The distribution of knowledge across nations. Research Policy, Special Section Knowledge Dynamics out of Balance: Knowledge Biased, Skewed and Unmatched 37, 1659–1673. https://doi.org/10.1016/j.respol.2008.08.006

Castellani, D., Castellani, D., Zanfei, A., 2006. Multinational firms, innovation and productivity. Edward Elgar Publishing.

Castellani, D., Zanfei, A., 2007. Multinational companies and productivity spillovers: is there a specification error? Applied Economics Letters 14, 1047–1051. https://doi.org/10.1080/13504850600706503

Chaminade, C., De Fuentes, C., Harirchi, G., Plechero, M., 2016. The geography and structure of global innovation networks: global scope and regional embeddedness, in: Handbook on the Geographies of Innovation. Edward Elgar Publishing.

Crescenzi, R., Iammarino, S., Ioramashvili, C., Rodríguez-Pose, A., Storper, M., 2019. The Geography of Innovation: Local Hotspots and Global Innovation Networks. WIPO Economic Research Working Paper No. 57. Geneva: WIPO.

D'Agostino, L.M., Santangelo, G.D., 2012. Do overseas R&D laboratories in emerging markets contribute to home knowledge creation? Management International Review 52, 251–273.

Delgado, M., Porter, M.E., Stern, S., 2016. Defining clusters of related industries. J Econ Geogr 16, 1–38. https://doi.org/10.1093/jeg/lbv017

Dunning, J.H., 1998. Location and the multinational enterprise: A neglected factor. Journal of International Business Studies 29, 45–66.

Duranton, G., Overman, H.G., 2005. Testing for Localization Using Micro-Geographic Data. Rev Econ Stud 72, 1077–1106. https://doi.org/10.1111/0034-6527.00362

Ellison, G., Glaeser, E.L., 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. Journal of Political Economy 105, 889–927. https://doi.org/10.1086/262098

Engel, J.S., 2015. Global Clusters of Innovation: Lessons from Silicon Valley [WWW Document]. URL https://journals.sagepub.com/doi/abs/10.1525/cmr.2015.57.2.36 (accessed 7.31.19).

Engel, J.S., del-Palacio, I., 2009. Global networks of clusters of innovation: Accelerating the innovation process. Business Horizons 52, 493–503.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Presented at the Kdd, pp. 226–231.

Feldman, M.P., 1994. The geography of innovation. Springer Science & Business Media.

Foley, C.F., Kerr, W.R., 2013. Ethnic Innovation and U.S. Multinational Firm Activity. Management Science 59, 1529–1544. https://doi.org/10.1287/mnsc.1120.1684

Franzoni, C., Scellato, G., Stephan, P., 2012. Foreign-born scientists: mobility patterns for 16 countries. Nat Biotech 30, 1250–1253. https://doi.org/10.1038/nbt.2449

Gerybadze, A., Merk, S., 2014. Globalisation of R&D and host-country patenting of multinational corporations in emerging countries. International Journal of Technology Management 64, 148–179.

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in Cities. Journal of Political Economy 100, 1126–52.

Ikeuchi, K., Kazuyuki, M., Ryuichi, T., Naotoshi, T., 2017. Measuring Science Intensity of Industry using Linked Dataset of Science, Technology and Industry (No. 17056), Discussion papers. Research Institute of Economy, Trade and Industry (RIETI).

Jacobs, J., 1969. The economy of cities. Vintage Books.

Jaffe, A.B., de Rassenfosse, G., 2017. Patent citation data in social science research: Overview and best practices. Journal of the Association for Information Science and Technology 68, 1360–1374. https://doi.org/10.1002/asi.23731

Kerr, W.R., Kominers, S.D., 2014. Agglomerative Forces and Cluster Shapes. The Review of Economics and Statistics 97, 877–899. https://doi.org/10.1162/REST_a_00471

Krishna, V.V., Patra, S.K., Bhattacharya, S., 2012. Internationalisation of R&D and Global Nature of Innovation: Emerging Trends in India. Science, Technology and Society 17, 165–199. https://doi.org/10.1177/097172181101700201

Li, G.-C., Lai, R., D'Amour, A., Doolin, D.M., Sun, Y., Torvik, V.I., Yu, A.Z., Fleming, L., 2014. Disambiguation and co-authorship networks of the U.S. patent inventor database (1975–2010). Research Policy 43, 941–955. https://doi.org/10.1016/j.respol.2014.01.012

Lorenzen, M., Mudambi, R., 2013. Clusters, Connectivity and Catch-up: Bollywood and Bangalore in the Global Economy. J Econ Geogr 13, 501–534. https://doi.org/10.1093/jeg/lbs017

Lucas Jr, R.E., 1993. Making a miracle. Econometrica: Journal of the Econometric Society 251–272.

Marshall, A., 1920. Principles of Economics: Abridged Edition, Abridged edition. ed. Cosimo Classics, New York.

McCann, P., Mudambi, R., 2005. Analytical differences in the economics of geography: the case of the multinational firm. Environment and Planning A 37, 1857–1876.

Moreira, A. and M.Y. Santos (2007). Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. In *Proceedings of the Second International Conference on Computer Graphics Theory and Applications (GRAPP 2007),* Barcelona, March 8–11. INSTICC Press. ISBN 978-972-8865-71-9, pp. 61–68.

Moreno, R., Miguélez, E., 2012. A Relational Approach to the Geography of Innovation: A Typology of Regions. Journal of Economic Surveys 26, 492–516. https://doi.org/10.1111/j.1467-6419.2012.00727.x

Morrison, G., Riccaboni, M., Pammolli, F., 2017. Disambiguation of patent inventors and assignees using high-resolution geolocation data. Scientific Data 4, 170064. https://doi.org/10.1038/sdata.2017.64

OECD, 2011. How's life?: measuring well-being. OECD Paris.

Patel, P., Pavitt, K., 1991. Large Firms in the Production of the World's Technology: An Important Case of "Non-Globalisation." Journal of International Business Studies 22, 1–21. https://doi.org/10.2307/155237

Plechero, Monica & Chaminade, Cristina, 2010. "Different competences, different modes in the globalization of innovation?. A comparative study of the Pune and Beijing regions," Papers in Innovation Studies 2010/3, Lund University, CIRCLE - Center for Innovation, Research and Competences in the Learning Economy.

Porter, M.E., 2000. Location, competition, and economic development: Local clusters in a global economy. Economic development quarterly 14, 15–34.

Rassenfosse, G. de, Kozak, J., Seliger, F., 2019a. Geocoding of worldwide patent data. Sci Data 6, 1–15. https://doi.org/10.1038/s41597-019-0264-6

Rassenfosse, G. de, Kozak, J., Seliger, F., 2019b. Geocoding of worldwide patent data (No. 19–458), KOF Working papers. KOF Swiss Economic Institute, ETH Zurich.

Reddy, P., 1997. New trends in globalization of corporate R&D and implications for innovation capability in host countries: a survey from India. World development 25, 1821–1837.

Saxenian, A., 2006. The New Argonauts: Regional Advantage in a Global Economy. Harvard University Press.

Saxenian, A., 2002. Silicon Valley's New Immigrant High-Growth Entrepreneurs. Economic Development Quarterly 16, 20–31. https://doi.org/10.1177/0891242402016001003

Sorenson, O., Rivkin, J.W., Fleming, L., 2006. Complexity, networks and knowledge flow. Research policy 35, 994–1017.

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

Useche, D., Miguelez, E., Lissoni, F., 2019. Highly skilled and well connected: Migrant inventors in cross-border M&As. J Int Bus Stud. https://doi.org/10.1057/s41267-018-0203-3

Webb, C., Dernis, H., Harhoff, D., Hoisl, K., 2005. Analysing european and international patent citations.

Yin, D., Kazuyuki, M., 2018. Inventor Name Disambiguation with Gradient Boosting Decision Tree and Inventor Mobility in China (1985-2016) (No. 18018), Discussion papers. Research Institute of Economy, Trade and Industry (RIETI).

Annex 1: Main data sources

Patent data

The patent data used in this report cover all patent documents – granted or not – filed from 1970 to 2017 in all patent offices worldwide and available in the European Patent Office's (EPO) PATSTAT database and WIPO's Patent Cooperation Treaty (PCT) collections. The unit of analysis is the first filing for a set of patent documents filed in one or more countries and claiming the same invention. Each set containing one first and, potentially, several subsequent filings is defined as a patent family. In the analysis, patent families are split into those oriented internationally and those oriented only domestically. Internationally oriented patent families refer to applicants seeking patent protection in at least one jurisdiction other than their country of residence. These include patent families containing only patent documents filed at the EPO or through the PCT. Conversely, domestic patent families refer only to filings in a home country, for instance, a Japan-based applicant filing only at the Japan Patent Office. The raw data come the European Patent Office's (EPO) Worldwide Patent Statistical Database (PATSTAT, April 2019) and WIPO's Patent Cooperation Treaty (PCT) collections. In the analyzed period (1970-2017), these sources account for 49,286,675 first patent filings and 26,626,660 subsequent patent filings, totaling 75,913,322 patent applications from 168 different patent offices.

Unless otherwise stated, the report makes use of international patent families only as the unit of analysis for all patent statistics reported. This relates mostly to the incomplete coverage of the domestic only patents (and utility models) of many national collections in PATSTAT. While the top national and international offices are usually well covered – namely USPTO, JPO, KIPO, CNIPA, EPO and WIPO – some other offices have limited coverage in PATSTAT. For instance, the coverage in PATSTAT of national collection data from some top 20 patent offices – such as India, Indonesia, Iran (Islamic Republic of), Mexico and Turkey – is limited. As a result, the report makes use of the information of 8,955,990 international patent families containing 35,582,650 different patent applications.

Scientific publication data

The scientific publication data used in this report comes from 27,726,805 records published from 1998 to 2017 in the Science Citation Index Expanded (SCIE) of the Web of Science (WoS), the citation database operated by the Clarivate Analytics company. The analysis focuses on 23,789,354 observations referring only to scientific articles, conference proceedings, scientific abstracts, and data papers. Scientific articles constitute the bulk of the resulting dataset.

Annex 2: Technological and scientific fields

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.[13] The criteria to group the fields was: (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; 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,

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

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; .

Annex 3: Geolocalization methodology

Patent data

As far as possible, the geocoding – attributing the geographical coordinates to a given location – relates to the inventors' address based on the best available data source within a patent family. The data work relies on the research efforts and generosity conducted by many others. In particular, it relies on geocoded patent data from Yin et al. (2018), Ikeuchi et al. (2017), Li, et al. (2014), de Rassenfosse et al (2019), Morrison et al. (2017) and PatentsView (www.patentsview.org, accessed March 2019). Many addresses are geocoded at a very precise level – i.e. street or block – but others only at the postal code or other sub-city level. To remain internationally comparable, but also due to the limited coverage of inventors' addresses in some national collections, the clustering analysis (see annex 4) relies only on internationally oriented patents.

For patents, 87 percent of the international patent families filed from 1976 to 2015 were geocoded.[14] Most of the non-geocoded cases had no usable address information.
As far as possible, the geocoding was applied to the inventors' addresses by using the most complete and reliable data source available within each patent family. In addition, the data were enriched with exiting geocoded patent data (see Yin et al., 2018; Ikeuchi et al., 2017; Li, et al., 2014; de Rassenfosse et al, 2019; Morrison et al., 2017).
All these sources and our own geocoding using ESRI or Geonames and geocoded postal codes official national sources were analyzed and consolidated to get the best possible geocoded data for each patent family. When there was more than one source for a given patent family, the following order of priority was given: (1) sources having information from the inventor (inventor principle); (2) sources having more inventors' addresses covered (coverage principle); (3) sources with the best geocoding resolution (resolution principle); (4) sources closest to the address country – e.g. entrusting Chinese addresses to CNIPA data, Japanese addresses to JPO data, etc. – (local principle); and (5) manually check and ad-hoc selection when two or more sources were still available. As a result, many inventor's addresses were geocoded at a precise level – i.e. street or block – but others only at the postal code or other sub-city level. Patent families containing more offices are more likely to be geocoded and at higher quality. This is another reason why the report relies only on international patent families.

Scientific publication data

The report assumes that research conducted for any publication takes place at the institutions and organizations to which the authors declare their affiliation. Virtually all of these locations were geocoded at the postal code or sub-city level using Geonames and geocoded postal codes official national sources. In the case of authors with more than one affiliation in the same publication, all different addresses are considered.
97 percent of all the available affiliation addresses were geocoded at the postal code or sub-city level.

[14] Patents in the entire period 1970-2017 were also geocoded, but the lack of complete addresses is more severe before 1976 and after 2015.

Annex 4: Identifying clusters from geo-referenced data

In some countries, the size gap between the different administrative areas is quite important. In the U.S. for example, one jumps from States (whose size is also very heterogeneous) down to counties. This is a well-known issue in spatial analysis, which goes under the name of modifiable areal unit problem (MAUP), which produces statistical and visual biases when aggregating point-based measures of spatial phenomena. Results and conclusions, both exploratory and confirmatory, can change considerably depending on the administrative area chosen as unit of analysis.

A related problem arises when one aims to carry out international comparisons of the agglomeration of economic and innovative activities, such as in our case. The administrative areas are country-specific and hardly comparable across countries. For example, the Chinese smallest units are counties, which tend to be larger than U.S. counties as well as than European NUTS3 regions. Even within Europe, a relatively homogeneous multi-country region, comparisons are difficult. The average size of a U.K. NUTS2 region is 6.581 square km. This same average jumps to 24.726 square km for France and 31.157 square km for Spain.

Next, administrative borders do not often coincide with the limit of the agglomerated economic activity one aims to measure (Alcácer and Zhao, 2016). It could be that a given economic activity encompasses a territory smaller than a given administrative areas, while there is nothing of that activity in the rest of its geographical space. It could be that a given administrative area hosts two or more distinct agglomerations, which we would not be able to identify if we rely only in administrative units. It could even be sometimes that the relevant agglomeration of economic and innovative activities span across county, region, State or even country borders, making the use of formal administrative areas misleading. Confounding different agglomerations within the same administrative unit inflicts further damage to the analysis when the two agglomerations have distinctive characteristics, such as different specialization patterns. For example, we may confuse two specialized agglomerations, respectively specialized in chemicals and electronics; or a classic Marshallian districts, highly specialized in a traditional or high-tech sector, with Jacobian cities, whose strength lies in its diversification and cross-industry spillovers. To the extent that we are interested in the role played by both types of agglomerations in the internationalization and globalization of science and technology, this confusion is highly undesirable.

A vast literature exists, which proposes different solutions to the MAUP, all of which consists in identifying agglomerations from georeferenced micro-data (Alcácer and Zhao, 2016, 2012; Delgado et al., 2016; Duranton and Overman, 2005; Ellison and Glaeser, 1997; Ester et al., 1996; Kerr and Kominers, 2014). We follow a similar approach here. In particular, we use all foreign-oriented patents from PATSTAT, from 1976 to 2015 to identify patent agglomerations, and Clarivate's Web of Science, from 1998 to 2017, to identify publication agglomerations.

We use DBSCAN (Density-based spatial clustering of applications with noise), an organically defined cluster identification approach based on the density-based algorithms, well-known in the geography literature (Ester et al., 1996). In a nutshell, DBSCAN is a well-known data clustering algorithm for geo-referenced spatial raw data, which is particularly suited for capturing agglomerations with an irregular shape (Ester et al., 1996).

Density-based spatial clustering of applications with noise

Density-based spatial clustering of applications with noise (in short, DBSCAN) is a well-known data clustering algorithm that is commonly used to identify clusters from spatial raw data. The algorithm was first proposed by Ester et al. (1996) who demonstrated how density-based algorithms are more suitable than partitioning and hierarchical methods to capture clusters with an irregular shape.

This feature makes DBSCAN algorithm particularly suited to our purpose, given the type of data we use, since we want to be able to identify clusters which may develop geographically across countries and very different urban areas, without being limited by administrative boarders.

Based on a set of coordinates, DBSCAN groups together points that are close to each other based on a distance measure and a minimum number of points. It also allows marking as outliers the points that are in low-density regions and thus are not included in any cluster. Also, this second feature is particularly important in our datasets, because we want to allow patenting and publication to still occur outside any cluster.

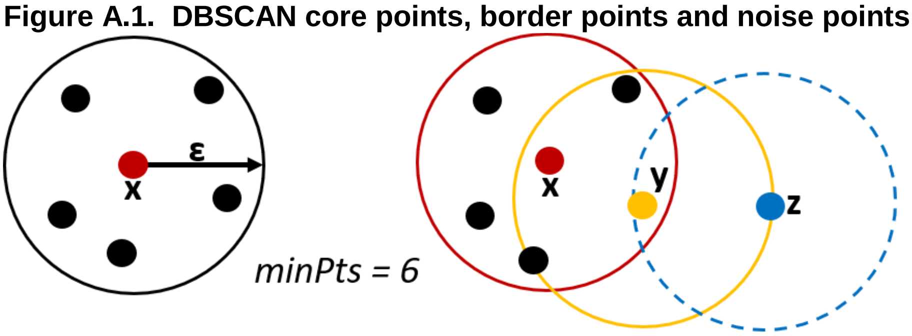
DBSCAN divides a dataset into subgroups of high density regions. The two parameters required for DBSCAN are epsilon (ε) and minimum amount of points required to form a cluster (minPts). The parameter epsilon defines the radius of neighborhood around a point x. It's called the ϵ-neighborhood of x. The parameter minPts is the minimum number of neighbors within ε radius.

We can imagine each data point having a circle with radius ε drawn around it. Using ε and minPts, we can classify each data point as:

- Core point: a point that has at a number of other points greater than or equal to minPts within its ε radius;
- Border point: a point is within the ε radius of a core point but has less than the minimum number of other points (minPts) within its own ε radius;
- Noise point: a point that is neither a core point nor a border point.

The figure below shows the different types of points (core, border and outlier points) using minPts = 6. Here x is a core point because $\text{neighbors}_\epsilon(x)=6$, y is a border point because $\text{neighbors}_\epsilon(y)<\text{minPts}$, but it belongs to the ϵ-neighborhood of the core point x. Finally, z is a noise point. Each core point forms a cluster together with the points that are reachable within its ε radius.

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



Source: Authors based on Ester et al. (1996) and https://www.datanovia.com/en/lessons/dbscan-density-based-clustering-essentials/ (accessed November 8, 2019).

Choosing DBSCAN parameters

As mentioned above, DBSCAN algorithm requires two input parameters: epsilon, which specifies how close points should be to each other to be considered a part of a cluster; and minPts, which specifies how many neighbors a point should have to be a core point of a cluster. The choice of these two parameters is arbitrary and critically determines the results of the cluster identification both in terms of size and shape of the clusters.

After some visual inspection of the data, and given the less precise geocoding of scientific publications (postal code, and not street) we set two different epsilon parameters for, respectively, patent and publication data. For publication data we calculated the average commuting distance to work in OECD countries (23 Km) (OECD, 2011), while for patent data we relied on a smaller radius, the 13 Km, as in Bergquist et al (2017).

To decide minPts, and after we have chosen the value for epsilon, we proceed as follows:

- We identify how many points lie within each point's epsilon-neighborhood. In order to do so, we used GIS software to create a buffer zone with epsilon radius around each coordinate point (i.e., buffer zones of 23/13 km radius around each coordinate point, respectively for scientific publications and patents) and counted each point's neighbors within the buffer – coordinate points weighted by the number of actual points in each coordinate.

- Once we have counted the number of neighbors, we run DBSCAN setting the minPts parameter at the median of the distribution of the number of neighbors. 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.

The algorithm is run separately for patents and scientific publications. In particular, we use all foreign-oriented patents from PATSTAT, from 1976 to 2015 to identify patent agglomerations, and Clarivate's Web of Science, from 1998 to 2018, to identify publication 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 overlap (more details below), identifying separately global clusters of patents and global clusters of publications.

Next, we also take care of fine-tuning our algorithm in order to make it suitable to identify both specialized and diversified agglomerations of patents and publications. The above method is repeated for 25 sub-samples of the publication and patent data that do not lie into the above identified clusters, which refer to 12 scientific fields and 13 technological ones, respectively.[15]

Figure A.8. below shows the distribution of the number of neighbors and the median value selected as minPts parameter for publication and patent data.[16] As can be seen,

[15] We also calculate minPts at the country level, which allows us to identify country specific innovation clusters. For highly innovative regions (such as the North America, Eastern Asia, Western and Northern Europe), the country threshold of minPts will be higher than the global one, thus only the denser areas will be identified as clusters, while the opposite is true for less innovative areas.

[16] The mean value is showed for comparison purposes, as well as because it is used as minPts to determine the clusters in technological fields 5100 and 5200.

thresholds are quite different across fields, reflecting the large heterogeneity among them.

Based on different ways to choose the parameters necessary to run DBSCAN, we have uncovered two types of agglomerations:

1. Global Innovation Hotspots (GIH): these are agglomerations that emerge from DBSCAN when using all the data across fields and technologies pooled (separately for publications and foreign-oriented patents). They are therefore large knowledge centers, either in patents or in scientific publications (normally in both)

2. Niche Clusters (NC): these are agglomerations emerging when running DBSCAN separately for each technology and field, respectively for patents and publications, to the patents and publications that do not belong to GIH. As the parameters chosen to run DBSCAN depend on the total number of patents/publications in a given field, with this second method some niche clusters that do not qualify to be GIH are identified.

Figure A.2. minPts value for publication and patent data

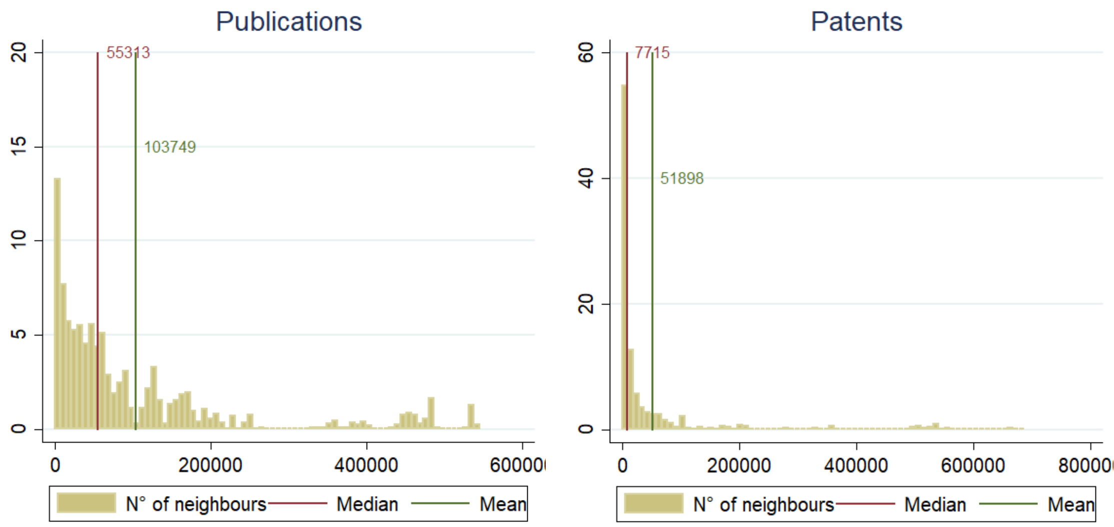
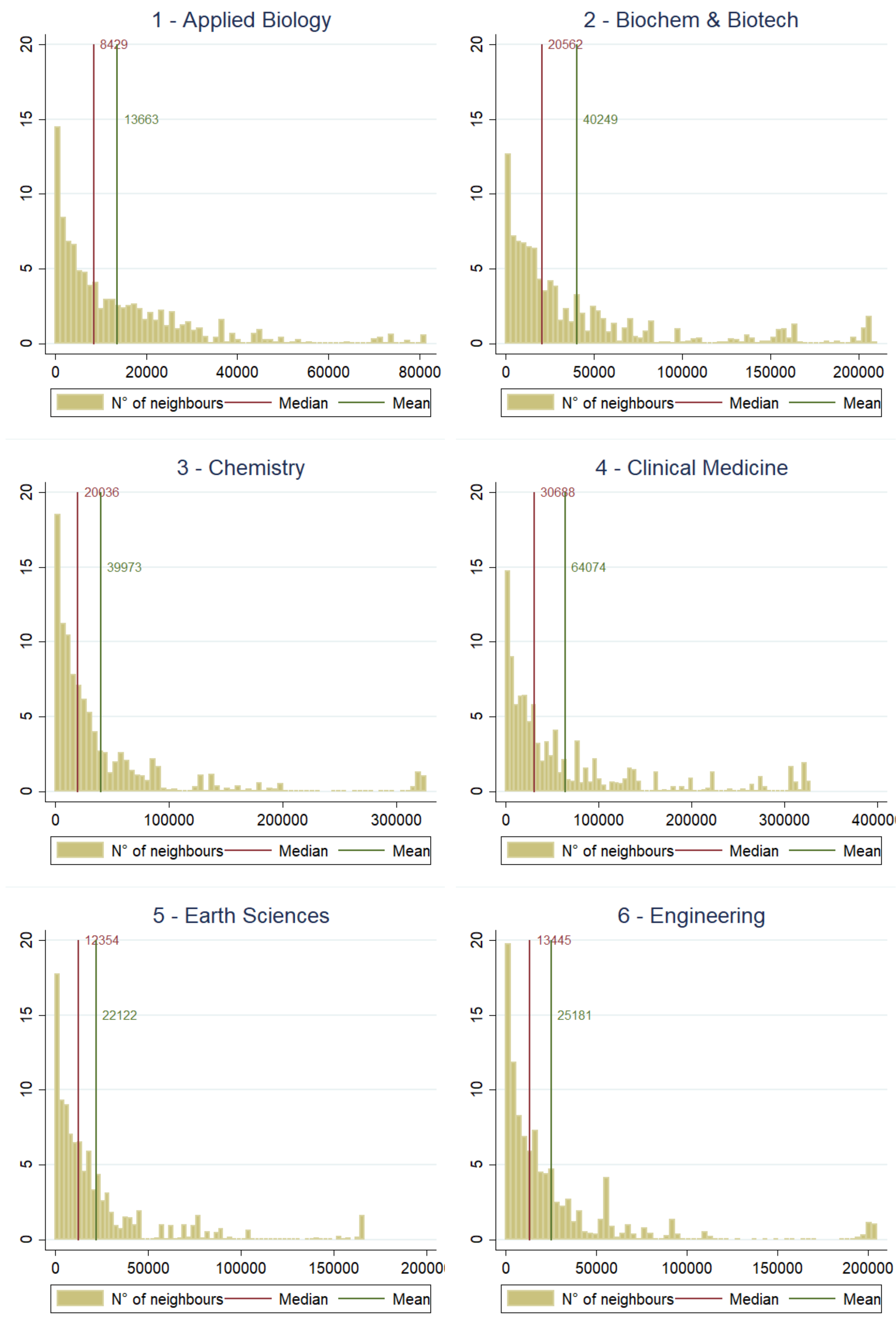


Figure A.3. minPts value for publication data by scientific field



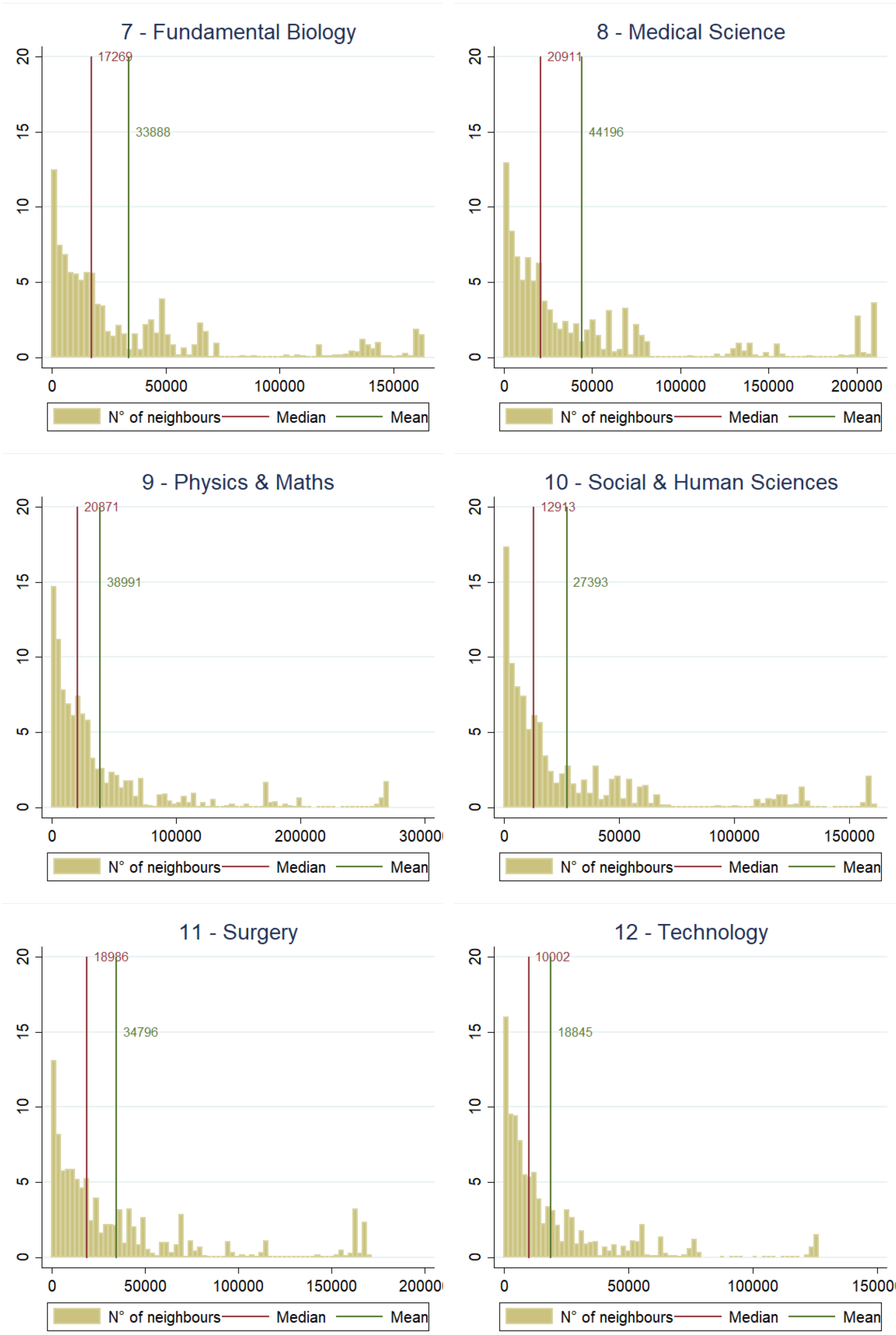
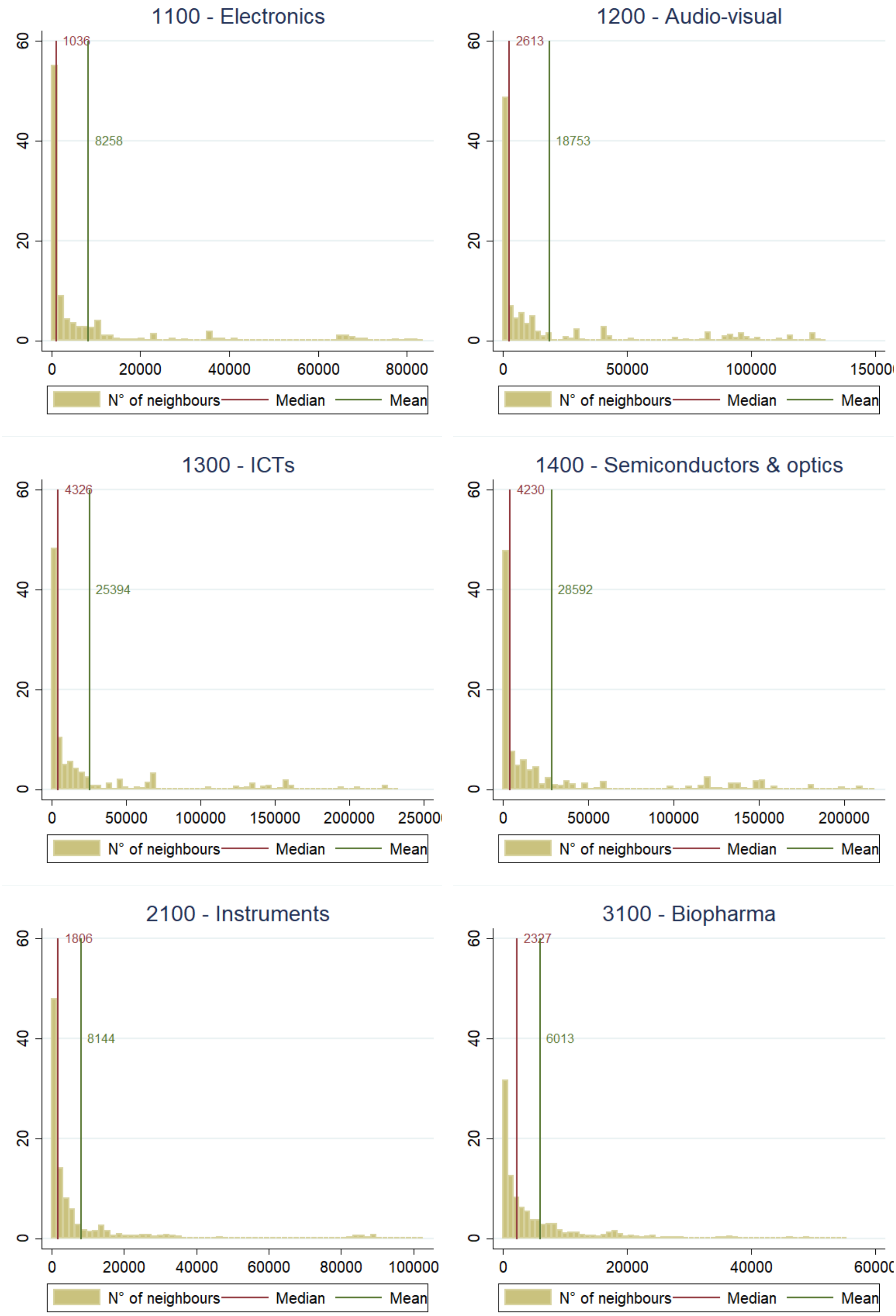
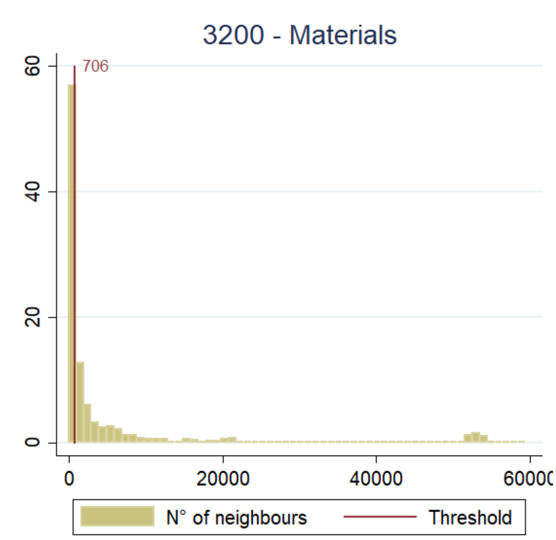
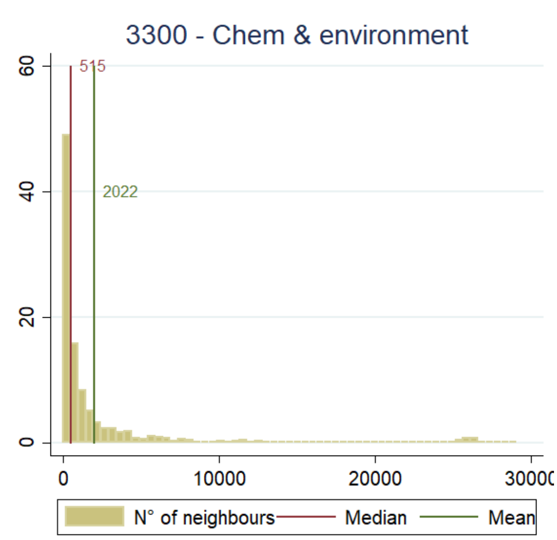
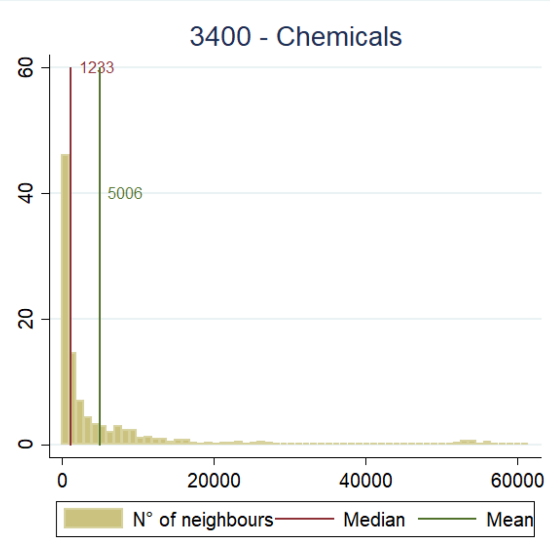

7 - Fundamental Biology
17269
33888
0
50000
100000
150000
N° of neighbours
Median
Mean
8 - Medical Science
20911
44196
0
50000
100000
150000
200000
N° of neighbours
Median
Mean
9 - Physics & Maths
20871
38991
0
100000
200000
300000
N° of neighbours
Median
Mean
10 - Social & Human Sciences
12913
27393
0
50000
100000
150000
N° of neighbours
Median
Mean
11 - Surgery
18986
34796
0
50000
100000
150000
200000
N° of neighbours
Median
Mean
12 - Technology
10002
18845
0
50000
100000
150000
N° of neighbours
Median
Mean

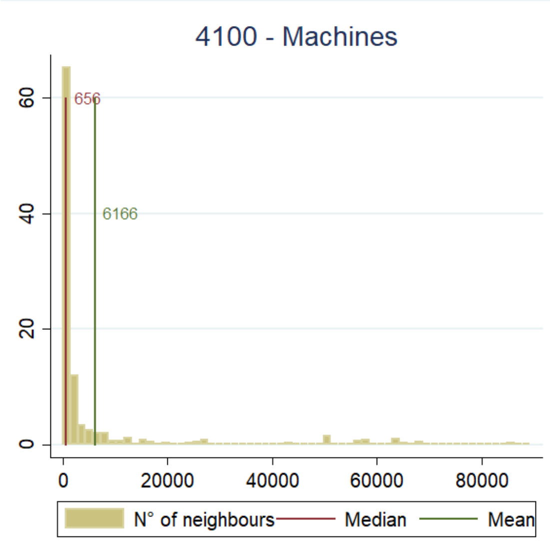
Figure A.4. minPts value for patent data by technological field

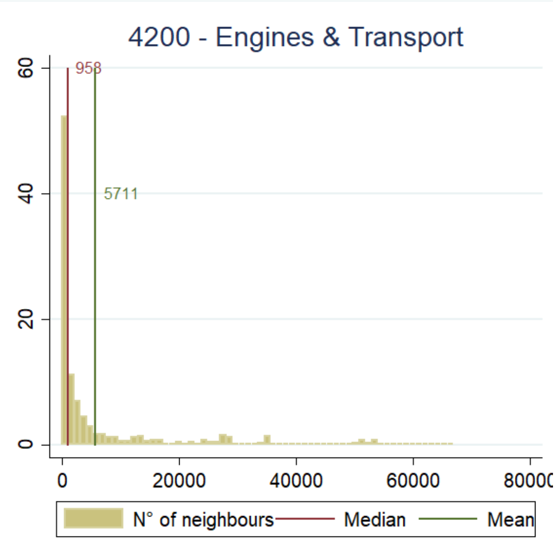


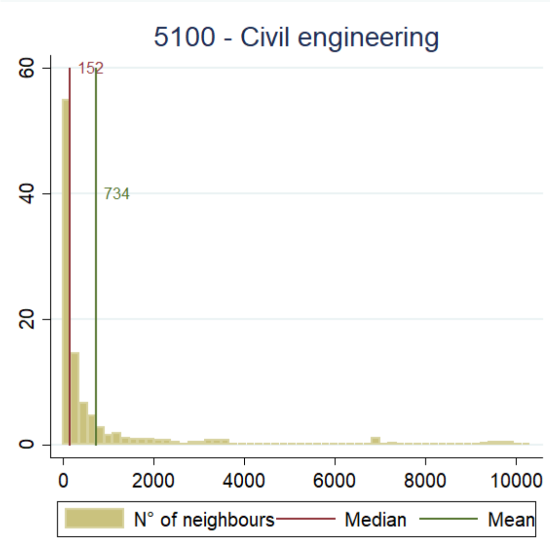

3200 - Materials
706
0
20
40
60
0
20000
40000
60000
N° of neighbours
Threshold

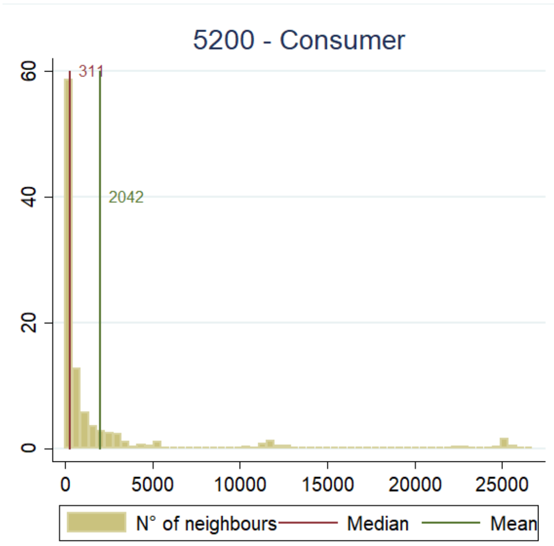

3300 - Chem & environment
515
2022
0
20
40
60
0
10000
20000
30000
N° of neighbours
Median
Mean


3400 - Chemicals
1233
5006
0
20
40
60
0
20000
40000
60000
N° of neighbours
Median
Mean


4100 - Machines
656
6166
0
20
40
60
0
20000
40000
60000
80000
N° of neighbours
Median
Mean


4200 - Engines & Transport
958
5711
0
20
40
60
0
20000
40000
60000
80000
N° of neighbours
Median
Mean


5100 - Civil engineering
152
734
0
20
40
60
0
2000
4000
6000
8000
10000
N° of neighbours
Median
Mean



Delimiting clusters' borders

We explain in more detail in this section how the different layers (clusters identified separately in patents and publications, and across the 25 fields – 12 scientific fields and 13 technological ones) that partially or totally overlap are merged, and how the borders of our final list of GIH and NC are defined.

First, once each coordinate point is assigned to a cluster-layer, their boundaries are defined by means of the convex and concave hull algorithms. The convex hull in a set X of points in the Euclidean space is the smallest convex set that contains all the X points. That is to say, it is a polygon encompassing all the X points, with straight, short lines connecting the outer points in space among X. Any point inside the outer shape belongs to the polygon. The convex hull however does not correctly handle concave shapes, and some points may be assigned to belong to a given polygon, while they are not. The concave hull starts from the convex hull, but removes the concave areas that do not have any point inside (Moreira and Santos, 2007). Different ways to remove these areas exist. We make sure to never go beyond cutting more than 25% of the original polygon, in order to ensure a meaningful shape of the polygon, especially when these are based in relatively few coordinate points.

Figure A.5. Convex and concave hull

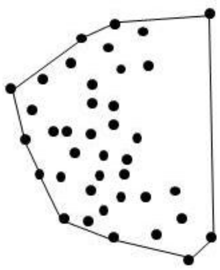


Figure A.11.a. Convex hull algorithm. It does not represent the area occupied by the set of points.

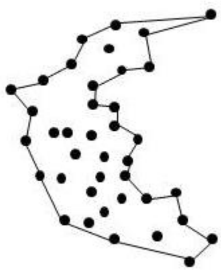


Figure A.11.b. Concave hull.

Thus, we compute all concave hulls to establish the boundaries of the polygons that encloses all the points belonging to a single-layer cluster. In most of the cases, as the identified clusters have more than two spatial points, the output is a polygon. However, when the cluster encompasses only a single point or a line (two points), a buffer of 13km is added, in order to transform it into a polygon and facilitate its visualization (this only happens around 5 cases between GIH and NC; one critical example being Teheran, where all its publications are geo-referenced in the same single coordinate point).

Layers identified from the concave hull after using patent and publication pooled data are merged to create the final boundaries of the GIH. In this step we only merge polygons from the patent layer (A) with polygons from the scientific publication layer (B) and vice versa, as by definition the polygons within each of the previous layer do not intersect each other, in order to create a new layer of polygons (C). In order to make this join, we follow four main criteria, as follows:

1. If a polygon (a), such that (a) is contained in A, intersects any polygon (b), such that (b) is contained in B, and the intersected area is equal or larger than 5% the area of (a), then (a) is joined with (b) and a new polygon (c) is created, such that c belongs to the new layer C. This process is repeated for the polygons (b) that were not initially contained in C, when the intersected area with (a) is equal or larger than 5% the area of (b).
2. If a polygon (a) intersects any polygon (b) and the intersected area is less than 5% the area of (a) or (b), then (a) and (b) are considered as different polygons. The intersected area is assigned to the polygon with the larger area, so the polygons (a’) and (b’) are created in the layer C.
3. If a polygon (a) does not intersect any polygon (b), then (a) and (b) are added to the layer C.
4. As a polygon from a layer (a) can intersect more than one polygon from layer (b) and vice versa, we repeat the procedure described in steps 1) and 2) using a threshold of 20% instead of 5%, for the polygons in layer c that intersects other polygons from the same layer.

For the NC, we follow a similar procedure as for GIH, but we first compute the union for each type of cluster no matter their field, before creating the final polygons. More specifically, after creating single layers for each of the type-fields (step 1), we create layer by type (a layer enclosing the 13 patent-field layers and another enclosing all the 12 scientific publication layers), applying the same four criteria iteratively. In other words, for each type of layer (patent or scientific publications) we begin merging two field-layers (i.e. the first two categories of the fields) and then we add to this output each of the remaining individual field-layers sequentially. It should be noticed that the polygons of the specialized clusters do not intersect the polygons of the hotspots, as for identifying specialized clusters we only consider the geographical points that are outside the hotspots’ boundaries.

By definition, the resulting areas: (1) are internationally comparable, i.e. the same scientific publication or patent (specialized) density would have determined the same hotspot (cluster) anywhere in the world; (2) can have different scientific and technological density, i.e. hotspots and niche clusters need only scientific publication or patent high concentration, but not necessary both; (3) have different specialization density, i.e. niche clusters are defined with lower density thresholds than hotspots; (4) are distinct geographical areas, i.e. the polygons are non-overlapping within and across hotspots and niche clusters; and, (5) have non-predefined boundaries, i.e. hotspots and niche clusters can have different sizes and include more than one city, state/province or country.