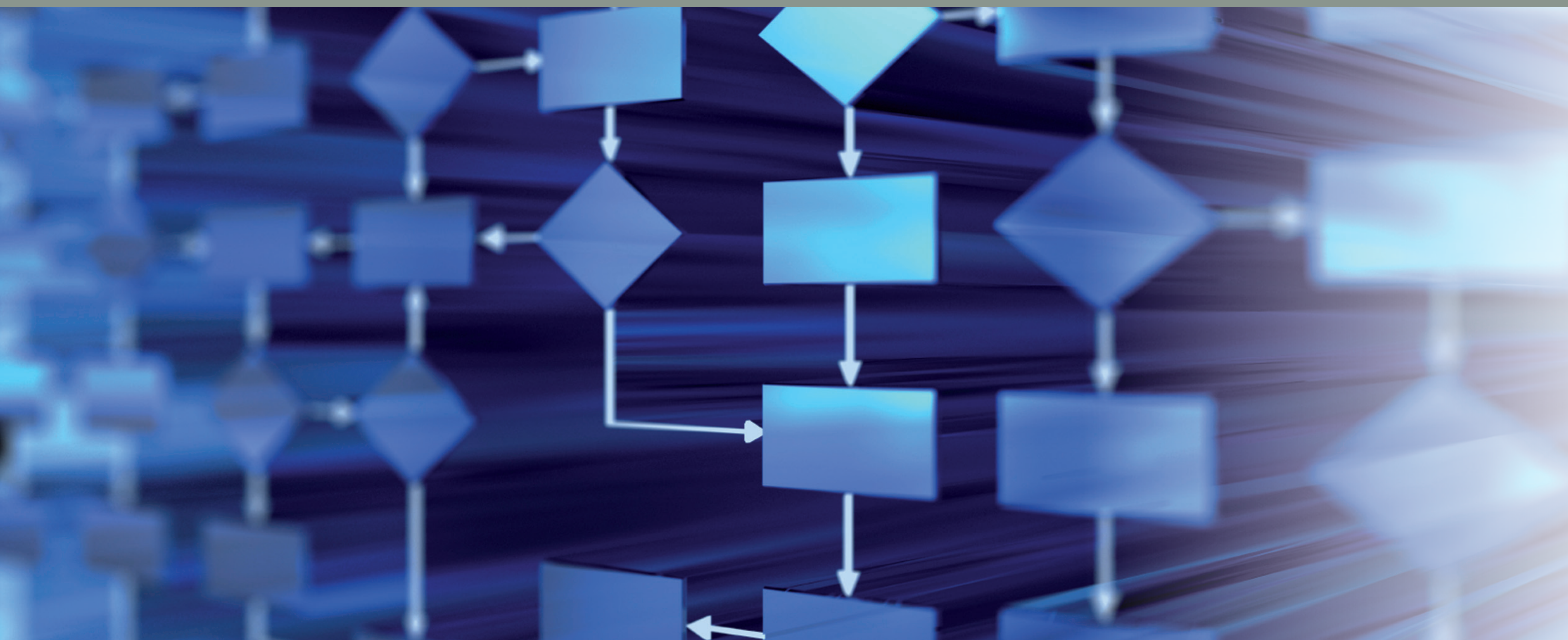
Economic Research Working Paper No. 34

Identifying and ranking the world's largest clusters of inventive activity

Kyle Bergquist
Carsten Fink
Julio Raffo



May 2017

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Identifying and ranking the world's largest clusters of inventive activity

Kyle Bergquist, Carsten Fink, and Julio Raffo [*]

Forthcoming in the 2017 edition of the Global Innovation Index

June 14, 2017

The Global Innovation Index (GII) has traditionally focused on the innovation performance of countries. This focus is rooted in the recognition that innovation outcomes are determined by factors that operate at the level of countries as a whole – such as national policies, laws and institutions, federal spending and cultural ties. The country perspective will continue to be a central focus of the GII. However, it masks important differences in innovation performance within countries, as innovation activities tend to be geographically concentrated in selected clusters linked to a single city or a set of neighboring cities.

Adopting a cluster view opens the door to better understanding the determinants of innovation performance that do not operate at the country level – such as physical and economic geography, sub-national policies and institutions, social networks and local labor market linkages. The GII has long recognized that innovation hubs at the city- or regional level tend to be drivers of innovation performance deserving in-depth analysis.[1] Unfortunately, gaining empirical insight into the comparative performance of individual innovation clusters is challenging. There is neither a generally accepted definition of what constitutes an innovation cluster, nor is there an "off-the-shelf" list of such clusters (see Box X in Chapter 1). In addition, the shape of innovation clusters typically does not correspond to the geographical units for which governments or other entities collect statistical data.

Seeking to overcome these challenges, this chapter presents an empirical approach to identifying and ranking the world's largest clusters of inventive activity on the basis of patent filings. Patent data offer rich information on the locality of innovative activity. Many researchers have already made use of these data to study individual clusters or selected clusters within a particular region.[2] Our approach goes beyond existing work by identifying and ranking innovation clusters on an internationally comparable basis.

The description of our empirical approach proceeds in several stages. We first describe the patent data that underlies our research (Section 1) and explain how we geocoded these data to enable the identification of clusters (Section 2). We then describe the algorithm we employed to map clusters (Section 3). Once identified, we discuss how we measured their size and explore how sensitive the resulting top-100 rankings are to the algorithm's input parameters (Section 4). We finally present the key characteristics of the top-100 clusters as they emerge from patent data (Section 5) and end with a few concluding remarks (Section 5).

---

[*] The authors are Data Analyst, Chief Economist, and Senior Economic Officer, respectively, in WIPO's Economics and Statistics Division. Comments and suggestions from Edward Harris, Ernest Miguelez, Yo Takagi, Sacha Wunsch-Vincent, Maryam Zehtabchi and Hao Zhou are gratefully acknowledged. The views expressed here are those of the authors, and do not necessarily reflect those of the World Intellectual Property Organization or its member states.

[1] See especially the 2013 edition of the GII on the theme of 'Local Dynamics of Innovation'.

[2] See, for example, Boix and Galletto (2007).

1. Description of patent data

Patents protect inventions that are new, involve an inventive step, and are capable of industrial application. Innovators interested in obtaining exclusive rights for their inventions have to formally apply for protection at authoritative offices. The patent records of these offices thus offer rich – and otherwise rare – information on the nature of inventive activity. Nonetheless, it is important to point out at the outset that patent data provide only an incomplete and imperfect perspective on overall innovative activity. The well-known limitations of patent data include the following:[3]

- Patents (mostly) capture technological inventions and thus miss out on non-technological innovations – such as organizational or logistical advances – that can be an important source of productivity gains in the economy.

- Patents do not capture all technological inventions, because inventors can also protect their inventions through trade secrets.

- Some industries use the patent system more intensively than others, depending on the nature of relevant technologies and prevailing business strategies.[4]

- Some patents are more valuable and technologically important than others; indeed, research has pointed to a highly skewed distribution of patent value, with relatively few patents accounting for a high share of the overall value of patents.[5]

These limitations do not mean that patent data cannot usefully inform innovation research. However, they should be kept in mind when interpreting the cluster rankings described in this chapter.

For our investigation, we rely on patents published between 2010 and 2015 under the Patent Cooperation Treaty (PCT) System, which is operated by the World Intellectual Property Organization (WIPO). The PCT is an international cooperation agreement that patent applicants use when they seek patent protection internationally. It came into force in 1978. By 2010, the System had 142 members that together accounted for more than 98 percent of national and regional patent filings worldwide.[6] In a nutshell, by filing a patent application under the PCT, applicants can delay the decision in which countries they would like to pursue exclusive rights for their inventions. In addition, the patent receives a first evaluation, which similarly helps applicants in their subsequent patent filing decisions.[7]

Our reliance on PCT filing data is motivated by two reasons. First, the PCT System applies one set of procedural rules to applicants from around the world and collects information based on uniform filing standards. This reduces potential biases that would arise if one were to collect similar information from different national sources applying different rules and standards. Second, PCT applications are likely to capture the commercially most valuable

---

[3] See UKIPO (2015) for a practical guide on the value and limitations of patent information for empirical analysis and WIPO (2011) for additional background on the economics of the patent system.

[4] See, for example, Hall and Ziedonis (2001).

[5] See, for example, Gambardella *et al* (2008).

[6] The four largest economies that were not party to the PCT System in 2010 were Saudi Arabia, Argentina, Venezuela, and Pakistan. Saudi Arabia acceded in 2013. An applicant from a non-member state can still file a PCT application, if there is a co-applicant from a member state. However, non-membership generally has a negative effect on the participation of applicants from non-members in the System, which one should keep in mind when interpreting the rankings presented in this chapter. The 98 percent coverage figure is an estimate based on national patent filing statistics available in WIPO's IP Statistics Data Centre (http://ipstats.wipo.int).

[7] See WIPO (2016) for a more detailed description of the PCT System.

inventions. Patenting is a costly process and the larger the number of jurisdictions in which a patent is sought, the greater the patenting cost. An applicant will only seek a patent internationally if the underlying invention generates a sufficiently high return – higher than for patents that are only filed domestically.[8]

On the downside, not all patent applications for which applicants pursue protection internationally go through the PCT system, and not every PCT application will eventually result in a granted patent.[9] Systemic differences in PCT use across countries, industries and applicants may thus introduce a measurement bias, which – again – should be kept in mind when interpreting our cluster rankings.

2. Geocoding PCT inventor addresses

Between 2011 and 2015, approximately 950,000 applications were published under the PCT system. Each of these applications lists the names and addresses of the inventor(s) responsible for the invention described in the application. In total, these amount to 2.7 million addresses.

Previous work using patent data assigned inventors to districts, primarily on the basis of the postal codes included in their addresses.[10] However, this approach biases the identification and measurement of clusters, due to the so-called Modifiable Areal Unit Problem (MAUP) – the choice of district boundaries exerting a strong influence on the shape and size of clusters.[11] The MAUP bias would be compounded in our case, because we seek to identify clusters on an internationally comparable basis and the geographical units associated with postal codes, for example, differ substantially across countries.

For this reason, we geocoded inventor addresses at a higher level of accuracy – ideally at the rooftop level – using Google Maps' Application Program Interface (API). While the quality of the API's returns varied, we were able to obtain highly accurate geo-coordinates for most countries.[12] Table 1 presents a summary of the geocoding results for the top PCT-filing countries. If Google Maps could not identify a specific geocode associated with an address, it typically returned an approximate area where that address is found. Extrapolating this information we were able to categorize our results into different accuracy scores. For most countries, more than two-thirds of the returned geocodes were within a 100m accuracy radius and more than 90 percent of the returns were within a 25km ratio – the accuracy threshold we employed for geocodes to be used for identifying clusters.[13] Since patent applications can list more than one inventor, the share of PCT filings covered by accurate geocodes is even higher.

[8] For other empirical investigations relying on PCT data, see Miguelez and Fink (2013) and Lax-Martínez *et al* (2016).

[9] In 2015, so-called PCT national phase entries accounted for 57 percent of non-resident patent filings worldwide (WIPO, 2016). However, this figure understates the 'market share' of the PCT, as it does not account for PCT applications that do not see any subsequent national phase entry.

[10] See, for example, Maraut *et al* (2008).

[11] See Oppenshaw (1983) for the seminal discussion of the MAUP.

[12] For some jurisdiction, this required fine-tuning the address feeds – mainly by progressively removing information that seemingly confused the API's address matching algorithm, such as the applicant name or outdated postal codes.

[13] The choice of this threshold partly reflects the reporting categories of the Google Maps API and the choice of cluster density parameters, as described in the next section.

Table 1: Summary of geocoding results

	Addresses with sufficiently accurate geocodes (%)			Share of PCT filings covered by accurate geocodes (%)
	<100m	<10km	<25km	
Australia	84.6	96.6	97.3	97.9
Austria	92.5	97.6	98.9	99.1
Belgium	54.8	93.0	95.4	96.3
Canada	78.3	95.6	95.9	96.8
China	25.4	60.8	94.9	94.9
Denmark	92.2	94.1	94.1	95.5
Finland	85.3	92.1	93.0	95.2
France	85.2	93.3	94.2	96.8
Germany	96.8	97.9	97.9	98.7
Hungary	90.1	91.4	91.4	94.5
India	60.6	76.7	77.5	85.2
Israel	64.8	79.2	86.9	80.1
Italy	83.5	85.4	85.4	88.3
Japan	81.7	89.9	89.9	91.3
Malaysia	76.0	79.8	79.8	83.2
Netherlands	96.9	99.4	99.5	99.5
Norway	86.8	94.4	94.9	95.5
Republic of Korea	34.7	78.6	89.4	89.3
Russian Federation	54.5	90.2	93.6	96.1
Singapore	78.1	79.0	79.0	84.5
Spain	66.1	96.0	98.8	98.8
Sweden	91.2	92.0	92.0	94.8
Switzerland	83.7	97.7	98.2	98.5
United Kingdom	70.7	97.5	97.8	98.2
United States	83.0	91.7	97.5	98.1

3. Density-based cluster identification

Researchers have used a variety of methods to identify clusters out of spatial raw data, depending on the nature of the data and the hypothesized forces giving rise to clustering. These methods range from pure visual identification to different kinds of technical algorithms.[14]

Having considered alternative options, we adopted the density-based algorithm for discovering clusters originally proposed by Ester *et al* (1996) – also referred to as the DBSCAN algorithm. In a nutshell, this algorithm groups together points with many nearby neighbors on the basis of pre-defined density parameters. Two reasons determined this choice. First, this algorithm can account for noise points not belonging to any cluster. This is important for our dataset, as patenting can occur outside of any innovation cluster – by, say, single “garage inventors”. Second, we are interested in descriptively measuring the innovation output of different localities, while initially being agnostic about what precisely drives the formation of these clusters. The DBSCAN algorithm allows us to flexibly map clusters across countries with varying physical and economic geography on the basis of the same density criteria.

We perform the DBSCAN algorithm on the geocoded inventor locations. In doing so, we treated multiple listings of the same address – for example, due to a single inventor being listed in multiple patent applications – as separate data points.

[14] For a recent overview, see Sharma *et al* (2016).

The DBSCAN algorithm requires two input parameters: the radius of the cluster-identifying circle around any given data point; and the minimum number of data points within that circle required for them to be counted toward a cluster. The choice of these input parameters critically determines the shape and size of identified clusters. We tested various combinations of input parameters with three guiding criteria. First, we focused on identifying the world's largest innovation clusters, which calls for a relatively high-density threshold. Second, we visually inspected the resulting clusters to evaluate the extent to which they correspond to our preconceived notions of existing clusters. Third, we made use of co-inventor relationships to evaluate the fit of the identified clusters. In particular, we gave preference to parameters that minimized the share of co-inventors outside the identified cluster but located within 160km of the cluster midpoint.

On the basis of these criteria, we settled on baseline input parameters of 13km (radius) and 2000 (minimum number of data points), corresponding to a density of approximately 5 listed inventors per square kilometer.[15] With these parameter values, the DBSCAN algorithm identified 162 clusters in 25 countries.

While most clusters were geographically separated from one another, a few of them were contiguous to one another.[16] In order to decide whether to merge these clusters into one, we again made use of co-inventor relationships. In particular, we calculated the share of a cluster's co-inventors belonging to all the other clusters as well as to two noise categories – namely, co-inventors located within and beyond 80km of the cluster midpoint not belonging to any other cluster. We then merged two clusters if two conditions were met for at least one of the clusters: first, the minimum distance between any two points of the two clusters was less than 5km; and second, the neighboring cluster accounted for the largest share of co-inventors among all clusters plus the two noise categories. This procedure led to the merger of eight clusters, so that we ended up with 154 clusters for our ranking.[17]

[15] Since DBSCAN relies on latitude and longitude coordinates to calculate the distance between two points, the second (inverse) geodetic problem implies somewhat shorter distances the further away those points are from the equator.

[16] The presence of contiguous clusters partly reflects the nature of the DBSCAN algorithm, as it has difficulties accounting for natural obstacles – such as rivers or train tracks – that cut through a cluster. Imperfect geocodes – say, those with only a 25km accuracy radius – may compound this problem as they often lead to the same geocode covering a large number of listed inventors. Our choice of a relatively large radius (13km) for DBSCAN minimizes but does not completely overcome these problems.

[17] In particular, we merged Alzenau with Frankfurt–Mannheim, Karlsruhe with Frankfurt–Mannheim, Bonn with Cologne–Dusseldorf, two separate Houston clusters, Södertälje with Stockholm, Takasaki with Tokyo–Yokohama, and Tsukuba with Tokyo–Yokohama. In addition, we merged Cheongju with Daejeon. Although Daejeon was only the second largest co-inventing cluster for Cheongju after Seoul, this largely reflects the strong presence of the Seoul cluster in the Republic of Korea. Indeed, all other identified clusters in the Republic of Korea feature Seoul as the largest co-inventing cluster (see Table 3). It is also worth pointing out that the merging of clusters had a negligible influence on the overall ranking of clusters, as at least one of the merging entities was always small in size.

Table 2: Cluster ranking

Rank	Cluster name	Territory(ies)	Number of PCT filings	
1	Tokyo–Yokohama	Japan	94,079	map
2	Shenzhen–Hong Kong	China/Hong Kong (China)	41,218	map
3	San Jose–San Francisco, CA	United States	34,324	map
4	Seoul	Republic of Korea	34,187	map
5	Osaka–Kobe–Kyoto	Japan	23,512	map
6	San Diego, CA	United States	16,908	map
7	Beijing	China	15,185	map
8	Boston–Cambridge, MA	United States	13,819	map
9	Nagoya	Japan	13,515	map
10	Paris	France	13,461	map
11	New York, NY	United States	12,215	map
12	Frankfurt–Mannheim	Germany	11,813	map
13	Houston, TX	United States	9,825	map
14	Stuttgart	Germany	9,528	map
15	Seattle, WA	United States	8,396	map
16	Cologne–Dusseldorf	Germany	7,957	map
17	Chicago, IL	United States	7,789	map
18	Eindhoven	Netherlands/Belgium	7,222	map
19	Shanghai	China	6,639	map
20	Munich	Germany	6,578	map
21	London	United Kingdom	6,548	map
22	Tel Aviv	Israel	5,659	map
23	Daejeon	Republic of Korea	5,507	map
24	Stockholm	Sweden	5,211	map
25	Los Angeles, CA	United States	5,027	map
26	Minneapolis, MN	United States	4,422	map
27	Portland, OR	United States	4,146	map
28	Nuremberg–Erlangen	Germany	4,049	map
29	Irvine, CA	United States	3,965	map
30	Berlin	Germany	3,632	map
31	Zurich	Switzerland/Germany	3,615	map
32	Philadelphia, PA	United States	3,172	map
33	Plano, TX	United States	3,147	map
34	Helsinki–Espoo	Finland	3,045	map
35	Singapore	Singapore	2,996	map
36	Basel	Switzerland/France/Germany	2,804	map
37	Raleigh–Durham, NC	United States	2,775	map
38	Hitachi	Japan	2,648	map
39	Copenhagen	Denmark	2,613	map
40	Hamamatsu	Japan	2,496	map
41	Washington, DC	United States	2,491	map
42	Cincinnati, OH	United States	2,481	map
43	Bengaluru	India	2,479	map
44	Sydney	Australia	2,380	map
45	Rotterdam–The Hague	Netherlands	2,235	map
46	Atlanta, GA	United States	2,162	map
47	Montreal, QC	Canada	2,124	map
48	Toronto, ON	Canada	2,094	map
49	Austin, TX	United States	2,089	map
50	Lyon	France	2,063	map

Rank	Cluster name	Territory(ies)	Number of PCT filings	
51	Wilmington, DL	United States	2,046	map
52	Barcelona	Spain	2,003	map
53	Regensburg	Germany	2,001	map
54	Brussels–Leuven	Belgium	1,994	map
55	Cambridge	United Kingdom	1,984	map
56	Grenoble	France	1,969	map
57	Moscow	Russian Federation	1,915	map
58	Milan	Italy	1,909	map
59	Hamburg	Germany	1,870	map
60	Melbourne	Australia	1,799	map
61	Madrid	Spain	1,796	map
62	Malmö	Sweden	1,737	map
63	Guangzhou	China	1,670	map
64	Indianapolis, IN	United States	1,596	map
65	Lausanne	Switzerland/France	1,580	map
66	Ottawa, ON	Canada	1,560	map
67	Hartford, CT	United States	1,540	map
68	Busan	Republic of Korea	1,470	map
69	Gothenborg	Sweden	1,461	map
70	Rochester, NY	United States	1,414	map
71	Vienna	Austria	1,403	map
72	Phoenix, AZ	United States	1,378	map
73	Vancouver, BC	Canada	1,362	map
74	Heidenheim–Aalen	Germany	1,352	map
75	Cleveland, OH	United States	1,346	map
76	Boulder, CO	United States	1,319	map
77	Yokkaichi	Japan	1,318	map
78	Haifa	Israel	1,298	map
79	Salt Lake City, UT	United States	1,293	map
80	Ann Arbor, MI	United States	1,289	map
81	Pittsburgh, PA	United States	1,283	map
82	Aachen	Germany/Netherlands/Belgium	1,279	map
83	Shizuoka	Japan	1,241	map
84	Buhl	Germany	1,223	map
85	Hangzhou	China	1,213	map
86	Albany, NY	United States	1,184	map
87	St. Louis, MO	United States	1,138	map
88	Oxford	United Kingdom	1,134	map
89	Baltimore, MD	United States	1,089	map
90	Daegu	Republic of Korea	1,085	map
91	Amsterdam	Netherlands	1,063	map
92	Kuala Lumpur	Malaysia	1,049	map
93	Clermont-Ferrand	France	1,041	map
94	Nanjing	China	1,030	map
95	Mumbai	India	1,012	map
96	Pune	India	1,006	map
97	Shikokuchuo	Japan	995	map
98	Toulouse	France	991	map
99	Hannover	Germany	979	map
100	Suzhou	China	956	map

Notes: The number of PCT filings refers to the 2011-2015 period. It represents the inventor fractional count of patents associated with a cluster, as explained in the text.

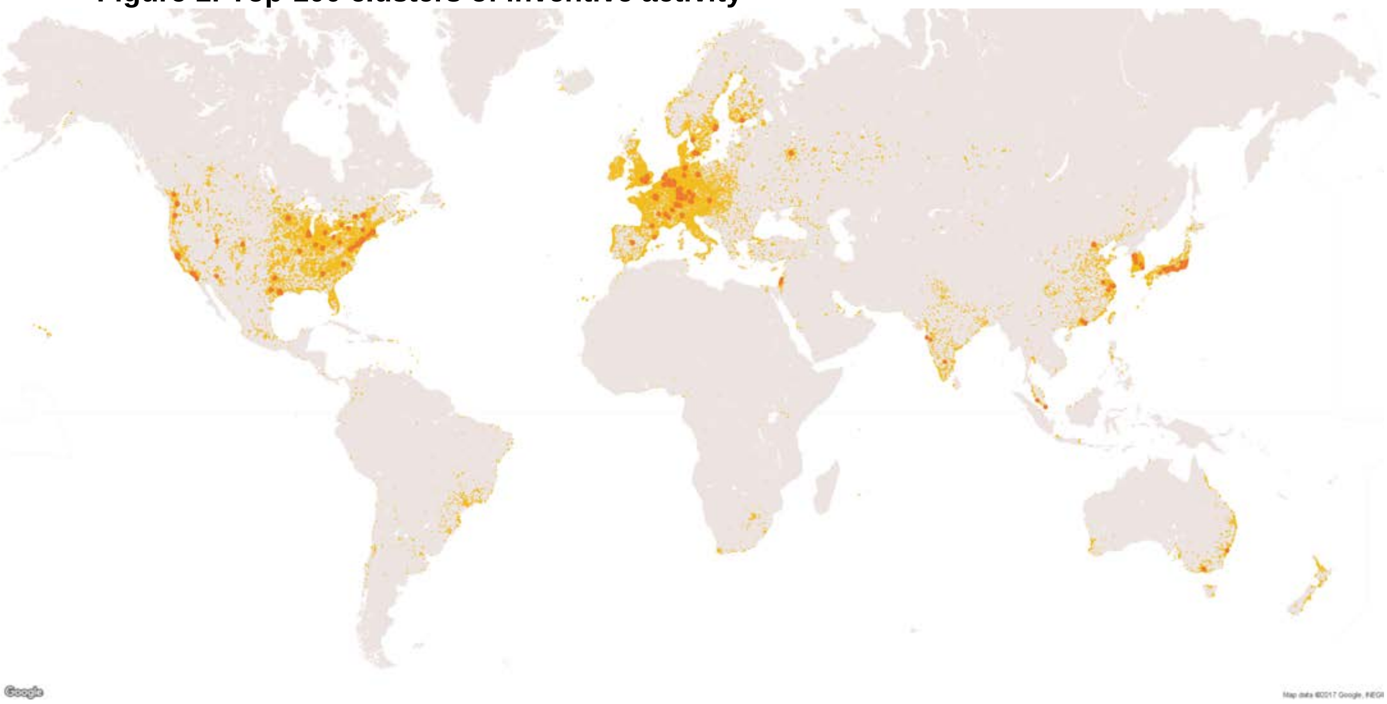
4. Measuring cluster size and sensitivity analysis

We measured the size of the identified clusters by the number of PCT applications associated with the data points in a given cluster. In doing so, we adopted a fractional counting approach, whereby counts reflect the share of a patent's inventors present in a particular cluster.[18] For example, a patent which lists three inventors in cluster A and one inventor in cluster B would contribute 0.75 to cluster A and 0.25 to cluster B.[19]

Table 2 presents the resulting ranking of the top-100 clusters.[20] The top 100 clusters account for 59.0 percent of all PCT filings in the period under consideration. We named clusters according to the main city or cities covered by the cluster. Tokyo–Yokohama – with a wide margin – emerges as the top-ranking cluster, followed by Shenzhen–Hong Kong, San Jose–San Francisco, Seoul, and Osaka–Kobe–Kyoto. These five clusters alone account for 23.9 percent of all PCT filings.

Figure 1 depicts the location of the top-100 clusters on a world map, also showing the 'raw' inventor address data points. Figures 2-4 offer zoomed-in regional perspectives and Figures 5-7 depict the shapes of the top-3 clusters.[21]

Figure 1: Top-100 clusters of inventive activity



[18] As alternative size measures, we also tested the simple count of listed inventors belong to a given clusters, and the (non-fractional) number of patents associated with those inventors. The resulting rankings correlated closely with the one relying on the fractional count for the top 35 clusters, though it led to several sizeable rank shifts for the remaining clusters that overall showed smaller differences in size scores. We only report rankings relying on fractional patent counts, as we feel it is the conceptually most appropriate size measure.

[19] Our fractional counts ignore inventors for which we obtained inaccurate geocodes (>25km). For example, if a patent has three inventors and the geocode for one inventor is inaccurate, we assigned 0.5 scores to the two inventors with accurate geocodes. However, given the small share of listed inventors and patents affected (see Table 1), the resulting measurement bias is likely to be small.

[20] The focus on the top-100 clusters reflects our choice of cluster identification input parameters, which are mainly geared to capture the world's largest clusters of inventive activity.

[21] Note that the visualization of the Shenzhen–Hong Kong cluster is somewhat misleading, as the relatively less accurate geocoding results for China (see Table 1) imply that many Chinese addresses are associated with the same geocode; in fact inventors located in Shenzhen account for a far higher share of cluster points than inventors located in Hong Kong.

Figure 2: Selected top-100 clusters in East Asia

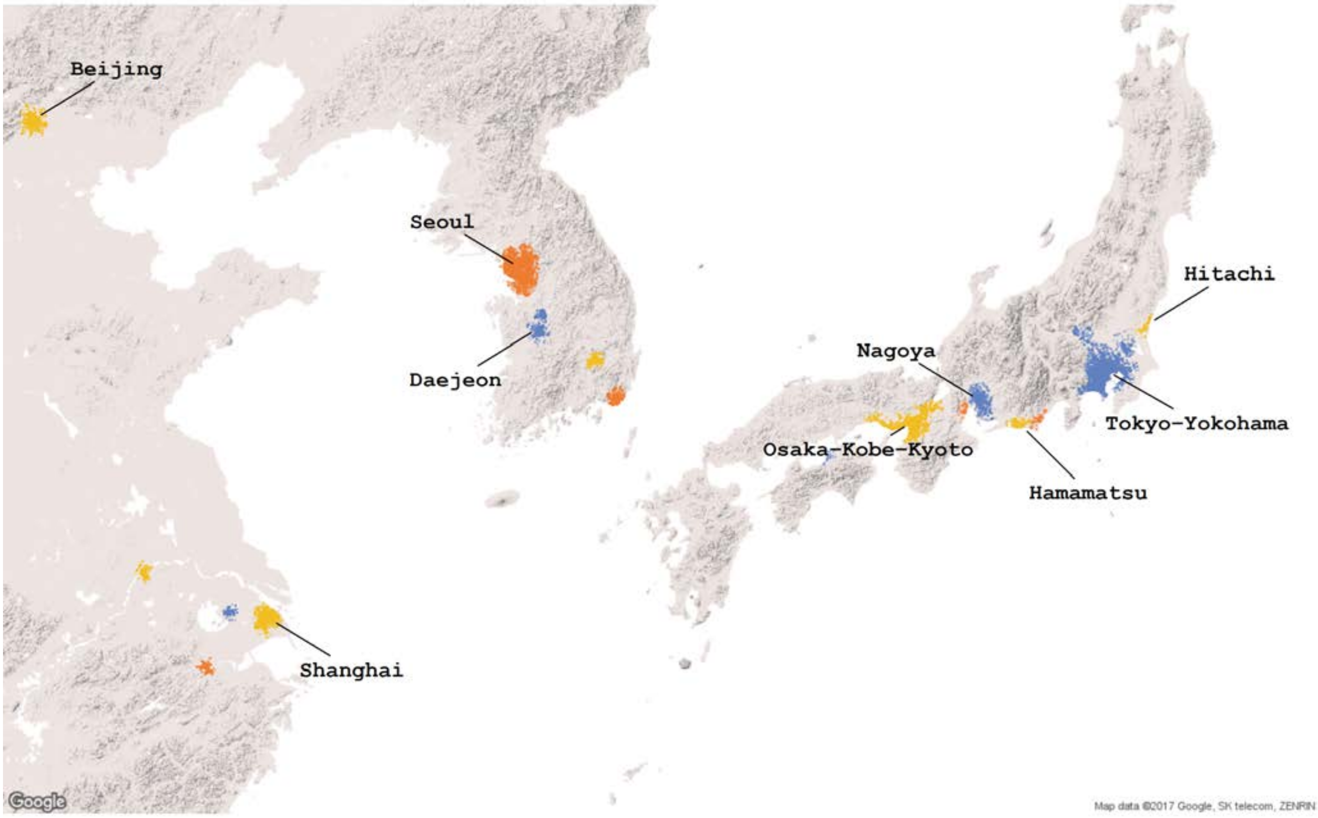


Figure 3: Selected top-100 clusters in Europe

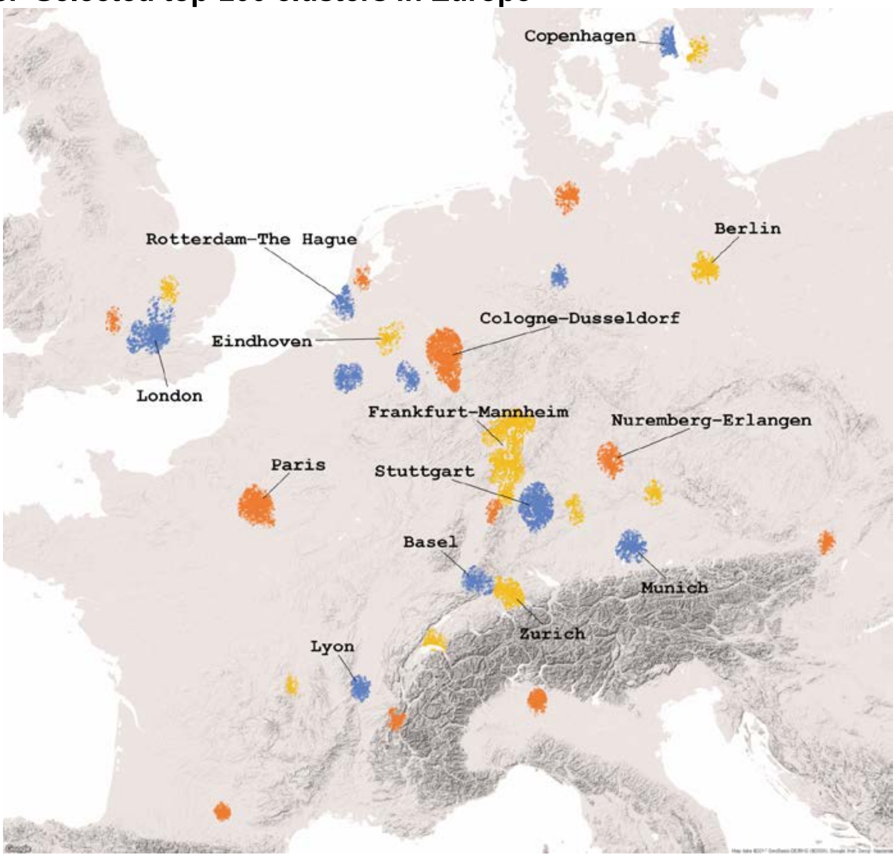
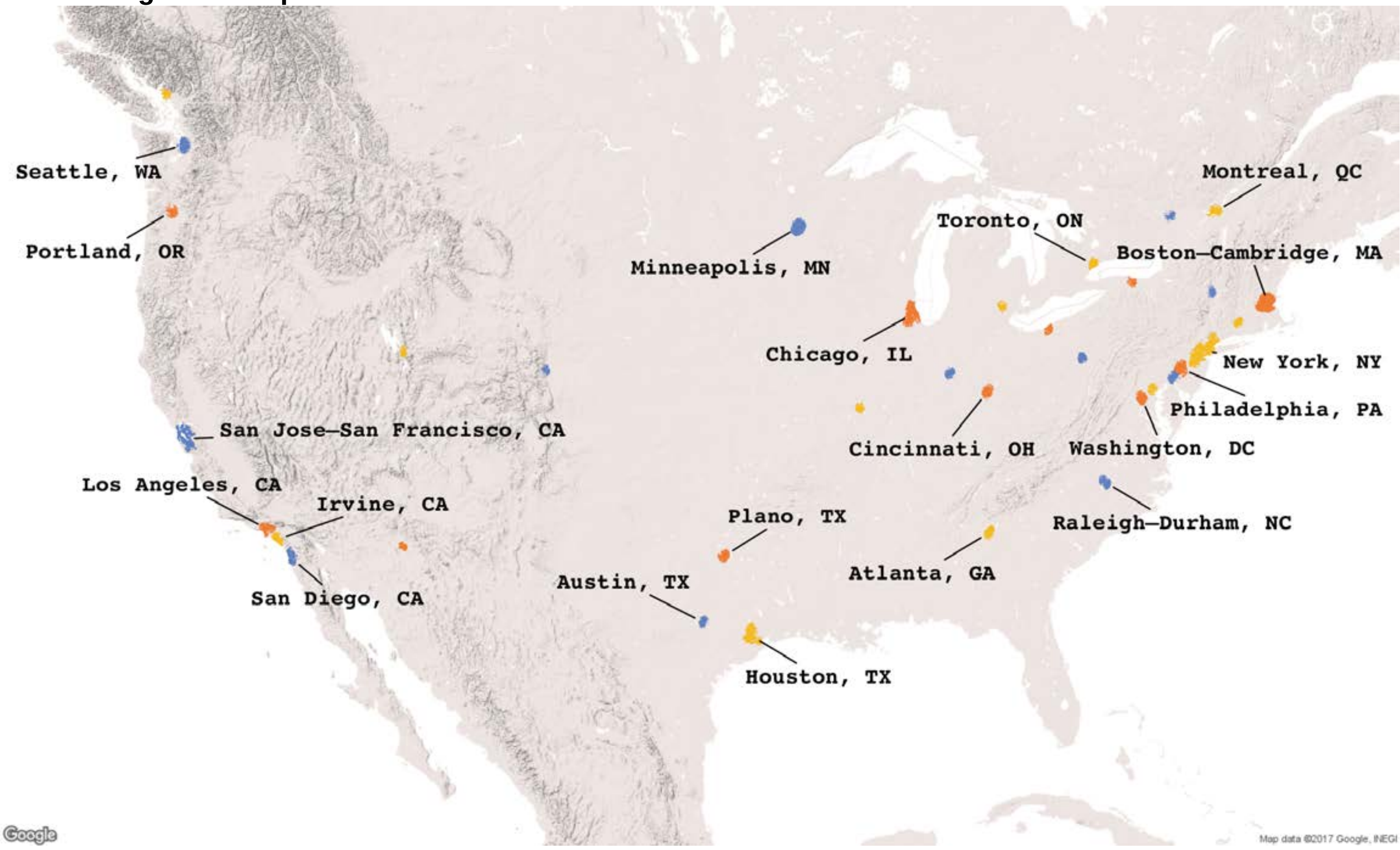


Figure 4: Top-100 clusters in North America



The distribution of clusters across countries is highly uneven. Seven countries feature four or more clusters in the top-100: the United States (31), Germany (12), Japan (8), China (7), France (5), Canada (4), and the Republic of Korea (4). An additional 16 countries feature between one and three clusters.[22] Among middle-income economies and other than China, India features three clusters, and Malaysia and the Russian Federation each feature one. The top-100 does not feature any cluster from Latin America and the Caribbean, Sub-Saharan Africa, and Northern Africa and Western Asia.

The distribution of clusters within countries is also uneven. Notably, in the case of the United States, less than half of the 50 states feature a cluster, while California, New York, and Texas each feature three or more. Finally, note that several clusters span more than one territory – most notably, the cluster located in the tri-border region around Basel.

[22] This count of clusters assigns multi-territory clusters to the territory accounting for the largest share of PCT filings. Note that an additional two countries – Norway and Hungary – feature clusters that do not rank among the top-100.

Figure 5: Tokyo-Yokohama cluster

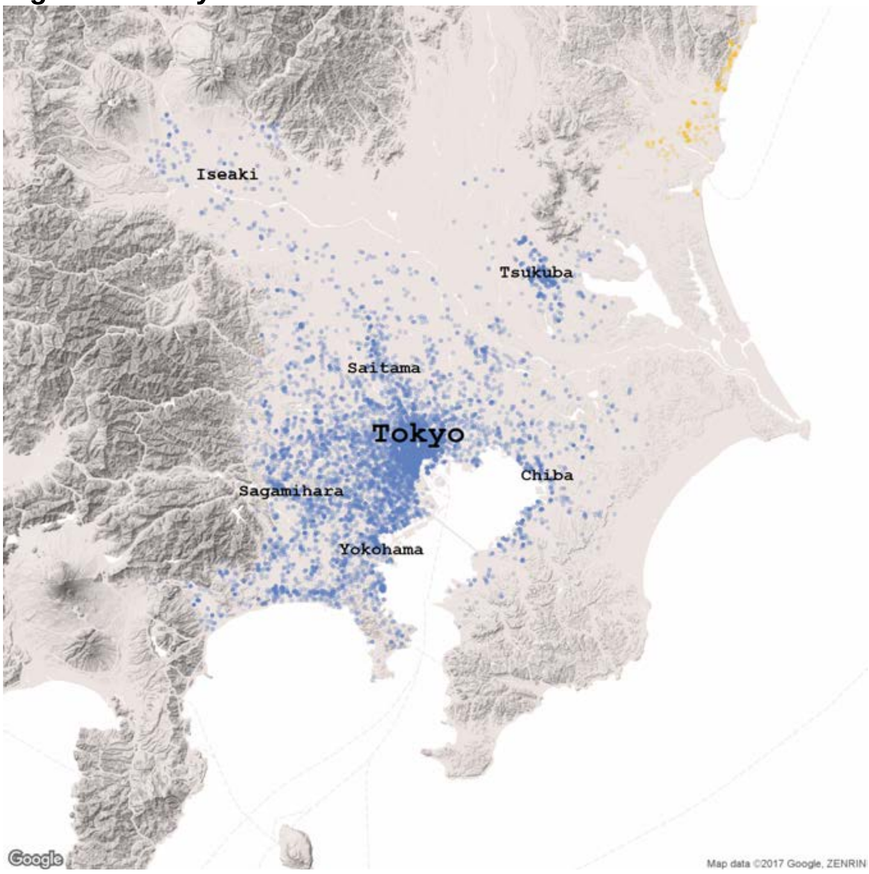
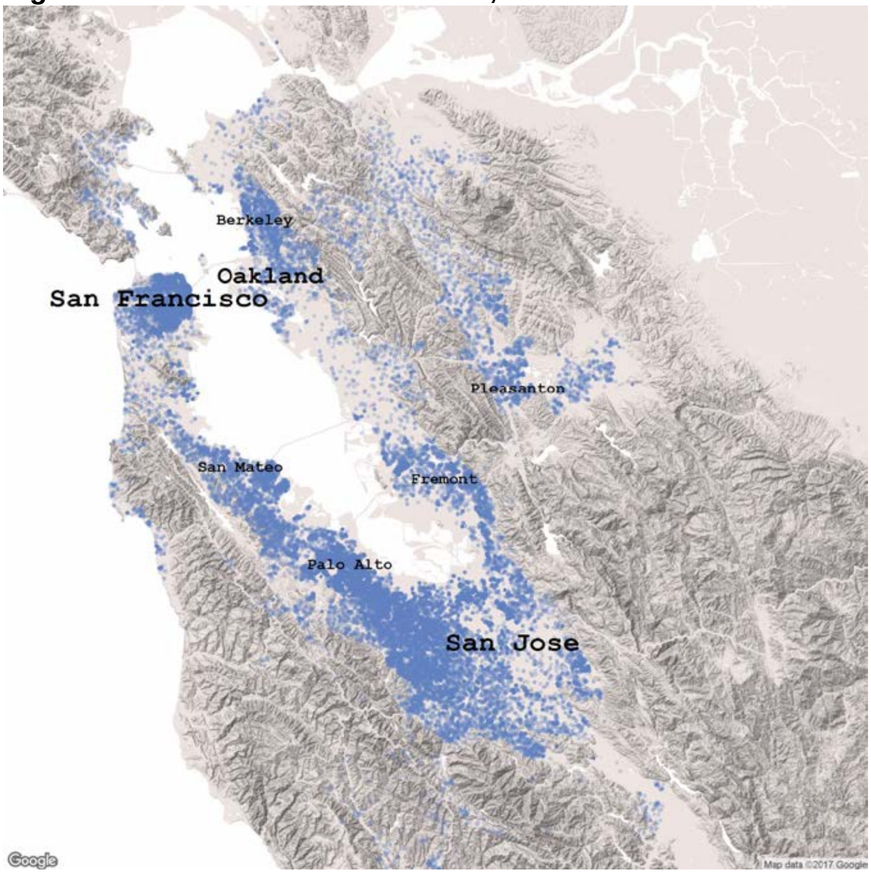


Figure 6: Shenzhen–Hong Kong cluster

Figure 7: San Jose–San Francisco, CA cluster



How sensitive is the ranking presented in Table 2 to different cluster-identifying input parameters? We tested different combinations of input parameters and compared the results to our baseline results. Two important insights emerged. First, while different input parameters influence the exact shape and size of the clusters, the resulting rankings were for the most part similar, with clusters only moving up or down a few ranks, especially for the top-30.[23] Tokyo–Yokohama always emerged as the top cluster. Second, two prominent (sets of) clusters were particularly sensitive to the chosen input parameters: New York and Frankfurt–Mannheim either emerged as broad clusters – as shown in Table 2 – or were divided into smaller clusters associated with the main population centers within those two clusters. These included Trenton, NJ, Newark, NJ, and Armonk, NY for the former, and Wiesbaden, Mannheim–Heidelberg, and Karlsruhe for the latter. Once divided, the smaller clusters had lower ranks, though Frankfurt and New York typically remained within the top-30.

[23] For this sensitivity analysis, we ignored extreme parameter values that led to counter-intuitive results – such as mega-clusters spanning several hundred kilometers,

5. Cluster characteristics

As already mentioned, patent data provide rich information on the nature of inventive activity and we can exploit these data further to characterize our top-100 clusters. In particular, Table 3 presents the largest patent filing entity, the main field of technology, the share of universities and public research organizations (PROs), the largest co-inventing cluster, and the share of women inventors associated with each cluster.

For most clusters, the largest patent applicant is a company, though for several of them it is a university – most notably the Massachusetts Institute of Technology for the 8th ranked Boston–Cambridge cluster. Interestingly, several companies constitute the top applicant for more than one cluster. Ericsson stands out as the largest applicant in five different clusters. Siemens and Intel each appear as the top applicant in four different clusters.

There are pronounced differences in the share of PCT filings accounted for by a cluster's top applicant. For many clusters, this share stands below 10 percent, suggesting a high degree of applicant diversity. For others, this share is higher, pointing to a more concentrated distribution of applicants within clusters. Most notably, Philipps accounts for 85 percent of the 18th ranked Eindhoven cluster, suggesting a cluster largely revolving around a single company.

Cluster diversity is also reflected in the share of the main technological field associated with a cluster's patent filings. For example, the 2nd ranked Shenzhen–Hong Kong cluster has a strong focus on digital communications, with around 41 percent of patent filings falling into this technology field. By contrast, the top ranked Tokyo–Yokohama cluster appears significantly more diversified, with its main technology field – electrical machinery, apparatus, and energy – accounting for only 6.3 percent of PCT filings. The most prominent technology field among the top-100 clusters is medical technology – accounting for the top field in 17 clusters – followed by digital communication (16), pharmaceuticals (15), and computer technology (12). Overall, 18 different technology fields – out of a total of 35 – feature as the top field in at least one cluster.

Interesting variation also exists in the prominence of universities and public research organizations (PROs) among the top-100 clusters. For some clusters – in particular, Baltimore, Daejeon, Grenoble, Kuala Lumpur, and Singapore – universities and PROs account for more than one-third of PCT filings. In many others, inventive activity largely occurs in companies, with academic institutions accounting for negligible filing shares. Interestingly, many clusters featuring medical technology or pharmaceuticals as their top field have relatively high university and PRO shares, underlining the importance of science linkages in these two fields.

How do the top-100 clusters connect to one another? One way of answering this question is to look at co-inventors located outside a cluster's borders and specifically in the remaining 99 clusters. On this basis, Table 3 identifies a cluster's most important partner cluster – defined as the cluster accounting for the largest share of external co-inventors. At least two interesting insights emerge. First, distance and cluster size – the classic gravity variables of economists – can in many cases explain the identity of the top partner cluster. For example, Tokyo–Yokohama is the top partner cluster for all other clusters in Japan and Seoul is the top partner cluster for all other clusters in the Republic of Korea. Second, the San Jose–San Francisco cluster emerges as by far the most collaborative cluster, emerging as the top partner in 24 cases, including 6 clusters located outside of the US.

The value of the top partner's share of external co-inventors captures the diversity of partner clusters. The low share for San Jose–San Francisco confirms the high degree of partner diversity for this cluster. Conversely, many clusters in Japan and the Republic of Korea show high shares, pointing to a more confined set of partners – possibly influenced by language barriers.

Finally, the last column in Table indicates the share of women inventors among all inventors located in a particular cluster. As can be seen, women inventors account for less than one-third of all inventors across all clusters. However, there is substantial variation in the extent of women participation; among the top-10 clusters alone the share ranges from 5.6 percent for Nagoya to 28.9 percent for Shenzhen–Hong Kong. Overall, the patterns shown largely reflect prior insights on the participation of women inventors: clusters in China and the Republic of Korea tend to be relatively more gender equal, as are clusters for which the main field of technology is either pharmaceuticals or biotechnology (see Lax-Martínez *et al*, 2016).

6. Concluding remarks

This chapter described an empirical approach towards identifying and measuring the size of the world's largest clusters of inventive activity on the basis of international patent filings. It provides a fresh perspective on the spatial agglomeration of innovative activity, relying on a globally harmonized set of criteria.

Notwithstanding the measurement progress offered by our approach, it is important to view the analysis presented here as a first step in a longer term effort to better capture innovative activity at the subnational level. Our approach relies exclusively on patent data, which are an imperfect metric for inventive activity and an even less perfect metric for innovative activity more broadly. In addition, while the identification and ranking of clusters is reasonably robust to different input parameter choices, the rankings should be used with due caution. Aside from Tokyo's top rank, they are best interpreted as orders of magnitude, with clusters moving up and down a few ranks depending on meaningful parameter choices.

For the future, we aim to improve and broaden the analysis presented here in at least three ways. First, we will seek to obtain more empirical insights into the forces giving rise to clustering and use these insights to refine our cluster identification approach. Second, we will analyze clusters at the level of specific technologies and industries, not least because key cluster forces – such as local labor market linkages – are often industry or technology specific. Finally, we will try to include other measures of innovative activity in the analysis – such as scientific publications and the performance of universities and firms – to obtain a more complete picture of innovation taking place across the world's largest clusters.

References

Boix, R. and V. Galletto. (2009). “Innovation and Industrial Districts: A First Approach to the Measurement and Determinants of the I-District Effect.” *Regional Studies*, 43(9). 1117-1133.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. (1996). “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, 226–231.

Miguelez, E. and C. Fink. (2013). “Measuring the International Mobility of Inventors: A New Database.” Economic Research Working Paper No. 8, WIPO.

Gambardella, A., D. Harhoff, and B. Verspagen. (2008). “The Value of European Patents.” *European Management Review*, 5(2), 69-84.

Hall, B.H. and R.H. Ziedonis. (2001). “The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995.” *The Rand Journal of Economics*, 32(1), 101-128.

Lax-Martínez, G.L., J. Raffo, and K. Saito. (2016). “Identifying the Gender of PCT Inventors.” Economic Research Working Paper No. 33, WIPO.

Maraut, S., H. Dernis, C. Webb, V. Spieza, and D. Guellec. (2008). “The OECD REGPAT Database: A Presentation.” Science, Technology, and Industry Working Papers, No. 2008/2, OECD.

Openshaw S. (1984). *The Modifiable Areal Unit Problem*. Geobooks, Norwich, England.

Sharma, A., R.K. Gupta, and A. Tiwari. (2016). “Improved Density Based Spatial Clustering of
Applications of Noise Clustering Algorithm for Knowledge Discovery in Spatial Data.” *Mathematical Problems in Engineering*, 2016.

UKIPO. (2015). *The Patent Guide*. Second Edition. The Intellectual Property Office, Newport, United Kingdom.

WIPO. (2016). *Patent Cooperation Treaty Yearly Review*. WIPO, Geneva, Switzerland.

WIPO. (2011). *World Intellectual Property Report: the Changing Face of Innovation*. WIPO, Geneva, Switzerland.

Table 3: Cluster characteristics

Rank	Cluster name	Largest applicant: Name	Largest applicant: Share of PCT filings	Main field of technology: Name	Main field of technology: Share of PCT filings	Share of universities & PROs	Largest co-inventing top-100 cluster: Name	Largest co-inventing top-100 cluster: Share of co-inventors	Share of women inventors
1	Tokyo–Yokohama	Mitsubishi Electric	6.4	Electr. mach., appar., energy	6.3	2.9	Osaka–Kobe–Kyoto	22.8	8.5
2	Shenzhen–Hong Kong	ZTE Corporation	32.4	Digital communication	41.2	1.2	Beijing	11.7	28.9
3	San Jose–San Francisco, CA	Google	6.5	Computer technology	18.3	3.4	Portland, OR	5.3	15.0
4	Seoul	LG Electronics	16.6	Digital communication	10.4	10.8	Daejeon	34.6	27.5
5	Osaka–Kobe–Kyoto	Murata Manufacturing	10.4	Electr. mach., appar., energy	8.3	4.2	Tokyo–Yokohama	51.3	8.6
6	San Diego, CA	Qualcomm	56.1	Digital communication	23.6	3.1	San Jose–San Francisco, CA	14.8	16.9
7	Beijing	BOE Technology Group	14.1	Digital communication	22.6	19.0	San Jose–San Francisco, CA	12.2	31.3
8	Boston–Cambridge, MA	M.I.T.	6.1	Pharmaceuticals	12.4	16.6	San Jose–San Francisco, CA	6.7	17.4
9	Nagoya	Toyota	42.4	Transport	13.0	1.9	Tokyo–Yokohama	41.2	5.6
10	Paris	L'Oréal	7.7	Transport	8.1	9.6	Lyon	4.5	18.9
11	New York, NY	IBM	4.2	Pharmaceuticals	10.9	12.4	San Jose–San Francisco, CA	5.8	20.0
12	Frankfurt–Mannheim	BASF	19.7	Organic fine chemistry	7.2	4.3	Stuttgart	7.8	13.4
13	Houston, TX	Halliburton	12.9	Civil engineering	25.1	5.2	New York, NY	4.0	11.6
14	Stuttgart	Robert Bosch	47.7	Engines, pumps, turbines	11.3	2.3	Frankfurt–Mannheim	12.6	4.8
15	Seattle, WA	Microsoft	41.9	Computer technology	34.6	4.2	San Jose–San Francisco, CA	16.8	13.2
16	Cologne–Dusseldorf	Henkel	7.7	Basic materials chemistry	7.1	2.4	Frankfurt–Mannheim	10.5	13.7
17	Chicago, IL	Illinois Tool Works	11.6	Digital communication	7.4	5.5	San Jose–San Francisco, CA	4.8	13.1
18	Eindhoven	Philips Electronics	84.9	Medical technology	17.9	0.9	Rotterdam–The Hague	7.2	12.0
19	Shanghai	Alcatel Lucent	4.3	Digital communication	9.5	11.4	New York, NY	6.3	30.2
20	Munich	Siemens	11.7	Transport	8.0	4.4	Nuremberg–Erlangen	4.4	9.3
21	London	Unilever	6.1	Digital communication	7.2	7.6	Cambridge	7.9	14.7
22	Tel Aviv	Intel	4.1	Computer technology	12.8	8.9	Haifa	22.3	13.5
23	Daejeon	LG Chem	19.8	Electr. mach., appar., energy	10.7	33.9	Seoul	68.6	27.3
24	Stockholm	Ericsson	44.1	Digital communication	26.8	0.5	San Jose–San Francisco, CA	6.2	10.3
25	Los Angeles, CA	University of California	8.4	Medical technology	9.5	21.2	San Jose–San Francisco, CA	12.1	15.0
26	Minneapolis, MN	Medtronic	14.1	Medical technology	32.7	4.0	San Jose–San Francisco, CA	4.4	12.1
27	Portland, OR	Intel	49.1	Computer technology	20.0	2.5	San Jose–San Francisco, CA	24.8	14.0
28	Nuremberg–Erlangen	Siemens	41.5	Electr. mach., appar., energy	11.5	8.3	Munich	8.1	4.7
29	Irvine, CA	Allergan	8.0	Medical technology	21.7	3.0	Los Angeles, CA	13.9	12.7
30	Berlin	Siemens	12.7	Electr. mach., appar., energy	8.5	12.6	Cologne–Dusseldorf	11.8	11.6
31	Zurich	ABB Technology	6.3	Medical technology	6.4	8.0	Basel	10.2	10.4
32	Philadelphia, PA	University of Pennsylvania	8.8	Pharmaceuticals	15.9	19.1	New York, NY	16.5	19.6
33	Plano, TX	Halliburton	17.1	Civil engineering	15.3	4.6	San Jose–San Francisco, CA	8.3	11.9
34	Helsinki–Espoo	Nokia	21.0	Digital communication	19.6	2.7	Beijing	6.4	14.0
35	Singapore	A*STAR	15.3	Medical technology	4.9	35.5	San Jose–San Francisco, CA	6.8	23.0
36	Basel	Hoffman-La Roche	10.6	Organic fine chemistry	13.1	3.0	Zurich	16.2	16.0
37	Raleigh–Durham, NC	Cree	11.1	Pharmaceuticals	9.3	19.7	Frankfurt–Mannheim	6.9	15.7
38	Hitachi	Hitachi	32.4	Electr. mach., appar., energy	19.9	0.5	Tokyo–Yokohama	86.3	7.1
39	Copenhagen	Novozymes	10.4	Biotechnology	11.1	11.9	Malmö	7.2	17.2
40	Hamamatsu	NTN Corporation	25.1	Transport	11.5	3.3	Tokyo–Yokohama	43.1	6.6

Rank	Cluster name	Largest applicant		Main field of technology		Share of universities & PROs	Largest co-inventing top-100 cluster		Share of women inventors
		Name	Share of PCT filings	Name	Share of PCT filings		Name	Share of co-inventors	
41	Washington, DC	US Department of HHS	11.6	Pharmaceuticals	14.7	15.6	San Jose–San Francisco, CA	7.5	19.4
42	Cincinnati, OH	Procter & Gamble	33.3	Medical technology	25.7	4.1	Frankfurt–Mannheim	4.7	14.6
43	Bengaluru	Hewlett-Packard	9.2	Computer technology	17.7	3.3	San Jose–San Francisco, CA	11.6	14.8
44	Sydney	University of Sydney	4.5	Medical technology	8.8	10.8	Melbourne	10.0	12.5
45	Rotterdam–The Hague	TNO	12.2	Other special machines	5.6	22.4	Amsterdam	8.4	11.2
46	Atlanta, GA	Georgia Tech Research	7.1	Medical technology	11.0	9.4	San Jose–San Francisco, CA	4.6	19.0
47	Montreal, QC	Ericsson	10.9	Digital communication	11.9	9.6	New York, NY	6.9	15.4
48	Toronto, ON	University Health Network	3.0	Computer technology	7.4	10.0	San Jose–San Francisco, CA	4.5	12.6
49	Austin, TX	University of Texas System	11.0	Computer technology	19.6	12.6	San Jose–San Francisco, CA	15.3	9.2
50	Lyon	IFP Energies Nouvelles	9.5	Organic fine chemistry	8.0	9.0	Paris	13.8	21.1
51	Wilmington, DL	Du Pont	47.1	Basic materials chemistry	8.2	3.9	Philadelphia, PA	21.1	15.5
52	Barcelona	Hewlett-Packard	8.7	Pharmaceuticals	9.4	17.3	Madrid	7.6	24.0
53	Regensburg	Osram Opto Semiconductors	36.7	Semiconductors	25.8	1.2	Munich	9.8	6.7
54	Brussels–Leuven	Solvay	4.7	Pharmaceuticals	6.1	12.3	Frankfurt–Mannheim	3.8	17.6
55	Cambridge	Cambridge University	6.7	Computer technology	8.1	10.4	London	17.6	14.9
56	Grenoble	CEA	44.3	Semiconductors	10.8	49.2	Paris	11.6	16.0
57	Moscow	Siemens	1.9	Pharmaceuticals	6.1	1.9	San Jose–San Francisco, CA	1.8	13.8
58	Milan	Pirelli	8.5	Pharmaceuticals	5.3	4.3	London	1.5	15.6
59	Hamburg	Henkel	11.0	Organic fine chemistry	14.1	3.1	Cologne–Dusseldorf	5.8	20.1
60	Melbourne	Monash University	5.1	Pharmaceuticals	5.8	16.3	Sydney	9.0	15.2
61	Madrid	Telefonica	13.3	Digital communication	11.1	25.7	Barcelona	9.0	26.9
62	Malmö	Ericsson	19.5	Digital communication	12.6	0.8	Stockholm	18.1	9.5
63	Guangzhou	South China Univ. of Tech.	6.8	Computer technology	6.8	19.3	Shenzhen–Hong Kong	10.4	29.2
64	Indianapolis, IN	Dow Agrosciences	22.6	Basic materials chemistry	8.6	6.8	New York, NY	3.4	16.0
65	Lausanne	Nestec	27.6	Food chemistry	7.5	12.4	Zurich	2.9	17.4
66	Ottawa, ON	Huawei Technologies	16.6	Digital communication	30.2	4.3	Plano, TX	13.6	17.4
67	Hartford, CT	United Technologies	65.7	Engines, pumps, turbines	39.6	1.4	Boston–Cambridge	4.9	9.7
68	Busan	Pusan National University	5.6	Medical technology	5.2	22.2	Seoul	48.6	24.7
69	Gothenborg	Ericsson	22.2	Digital communication	9.4	0.3	Stockholm	12.8	11.4
70	Rochester, NY	Eastman Kodak	38.2	Textile and paper machines	9.9	10.1	San Jose–San Francisco, CA	3.9	15.4
71	Vienna	Technische Universität Wien	4.3	Pharmaceuticals	7.8	10.4	Munich	2.9	12.7
72	Phoenix, AZ	Intel	15.4	Semiconductors	11.8	1.7	Portland, OR	9.0	13.0
73	Vancouver, BC	University of British Columbia	6.8	Pharmaceuticals	5.5	11.7	San Jose–San Francisco, CA	8.9	12.9
74	Heidenheim–Aalen	Carl Zeiss	21.9	Optics	15.9	0.2	Stuttgart	9.9	5.7
75	Cleveland, OH	Cleveland Clinic Foundation	9.7	Medical technology	11.1	19.9	New York, NY	2.5	11.2
76	Boulder, CO	University of Colorado	5.8	Medical technology	11.6	7.0	San Jose–San Francisco, CA	8.6	14.4
77	Yokkaichi	Autonetworks Technologies	39.1	Electr. mach., appar., energy	32.3	0.7	Tokyo–Yokohama	33.8	2.9
78	Haifa	Intel	10.8	Medical technology	18.6	8.7	Tel Aviv	46.9	12.9
79	Salt Lake City, UT	University of Utah	14.9	Medical technology	19.3	16.0	San Jose–San Francisco, CA	7.3	10.8
80	Ann Arbor, MI	University of Michigan	27.3	Pharmaceuticals	7.1	29.5	San Jose–San Francisco, CA	4.2	14.1
81	Pittsburgh, PA	University of Pittsburgh	12.8	Medical technology	9.0	21.3	Boston–Cambridge	4.0	14.0

Rank	Cluster name	Largest applicant: Name	Largest applicant: Share of PCT filings	Main field of technology: Name	Main field of technology: Share of PCT filings	Share of universities & PROs	Largest co-inventing top-100 cluster: Name	Largest co-inventing top-100 cluster: Share of co-inventors	Share of women inventors
82	Aachen	Ericsson	13.3	Digital communication	9.0	10.5	Cologne–Dusseldorf	16.7	8.9
83	Shizuoka	Fujifilm	48.1	Optics	11.2	0.3	Tokyo–Yokohama	41.2	8.5
84	Buhl	Schaeffler Technologies	48.6	Mechanical elements	44.0	0.5	Frankfurt–Mannheim	28.0	3.6
85	Hangzhou	Alibaba Group	26.5	Computer technology	16.9	12.0	Shanghai	12.2	27.1
86	Albany, NY	General Electric	55.0	Semiconductors	9.9	6.5	New York, NY	9.6	13.0
87	St. Louis, MO	Monsanto Technologies	11.5	Biotechnology	10.4	13.6	Seattle, WA	6.6	17.4
88	Oxford	Oxford University Limited	27.6	Pharmaceuticals	8.3	31.3	London	15.8	18.1
99	Baltimore, MD	Johns Hopkins University	45.3	Pharmaceuticals	15.0	51.9	Washington, DC	13.0	20.7
90	Daegu	Kyungpook National University	12.1	Medical technology	7.7	26.1	Seoul	51.1	26.3
91	Amsterdam	Shell	29.1	Basic materials chemistry	8.6	9.2	Rotterdam–The Hague	13.6	13.8
92	Kuala Lumpur	Mimos Berhad	50.0	Computer technology	11.4	68.0	Houston, TX	8.0	25.5
93	Clermont-Ferrand	Michelin	74.1	Transport	26.3	3.0	Paris	13.0	17.0
94	Nanjing	Southeast University	10.1	Digital communication	8.7	30.9	Beijing	10.1	31.5
95	Mumbai	Piramal Enterprises	6.7	Organic fine chemistry	15.4	5.9	Bengaluru	11.1	16.8
96	Pune	CSIR	23.2	Organic fine chemistry	15.7	24.5	San Jose–San Francisco, CA	9.8	12.4
97	Shikokuchuo	Unicharm Corporation	90.0	Medical technology	52.3	0.6	Tokyo–Yokohama	34.5	15.5
98	Toulouse	Continental	10.1	Transport	10.0	17.9	Paris	13.8	19.2
99	Hannover	Continental	14.3	Transport	15.3	7.1	Cologne–Dusseldorf	4.1	8.1
100	Suzhou	Ecovacs Robotics	7.7	Furniture, games	7.9	6.0	Shanghai	9.5	25.4

Notes: PCT filing shares refer to the 2011-2015 period and are based on fractional counts, as explained in the text. We use the location of inventors to associate patent applicants to clusters; note that addresses of applicants may well be outside the cluster(s) to which they are associated. The identification of technology fields relies on the WIPO technology concordance table linking International Patent Classification (IPC) symbols with thirty-five fields of technology (available at http://www.wipo.int/ipstats/en/). The identification of universities and public research organizations (PROs) relies on keyword-based searches of PCT applicant names, which encompasses all types of educational and public research entities, including universities, colleges, polytechnics, and university hospitals; it also take account of the different languages of PCT applicant names.

The largest co-inventing top-100 cluster refers to the cluster hosting the highest share of co-inventors. The share of co-inventors is relative to the total number of co-inventors located outside the cluster in question.

The identification of women inventors relies on the name dictionary described in Lax-Martínez *et al.* (2016). With this dictionary, we can attribute gender for more than 90 percent of listed inventors for each cluster except for Beijing, Bengaluru, Guangzhou, Hangzhou, Kuala Lumpur, Seoul, Shanghai, and Suzhou, for which we attribute gender for 84-90 percent of listed inventors. The share of women inventors is calculated on the basis of listed inventors, so inventors listed in multiple applications are counted multiple times. The calculation ignores inventors whose gender could not be attributed.