

György Csomós – Zsófia Viktória Vida – Balázs Lengyel*

Exploring the Changing Geographical Pattern of International Scientific Collaborations Through the Prism of Cities

Introduction

The decrease of communication and travel costs since the 1990s has enabled interactions between distant partners. However, and despite the early visions developed in the geography literature on the decreasing significance of distance (Batty 1993; Cairncross 2001; Friedman 2005; Kolko 2000; Ohmae 1995; O'Brien 1992), it is repeatedly found that the majority of social interactions are spatially bounded (Lambiotte et al. 2008; Lengyel et al. 2015; Liben-Nowell et al. 2005). Research is no exception: the probability of collaborations decreases as distance grows, as has been found for co-authorship relations (Frenken et al. 2009; Katz 1994; Pan-Kaski–Fortunato 2012), EU-supported research collaboration (Maggioni 2009), and inventor collaboration (Tóth et al. 2020). What is not entirely clear yet is how the quality of collaborative output influences the distance-dependence of scientific collaboration. Are high-impact collaborations similarly constrained by distance in the same manner as for collaborations of lower impact?

High-impact research—reflected by the number of citations a publication receives—is increasingly a multi-university phenomenon (Jones–Wuchty–Uzzi 2008), in which the combination of diverse knowledge located in many departments pays off in better-received publications. However, it is still debated whether spatial concentration or spatial diversity produces more high-impact papers. Jones et al. (ibid) claim that high-impact publications concentrate in spatially concentrated elite universities. In this line, Abbassi and Jaafari (2013) find that national collaboration favors citations more than international collaboration. On the contrary, scientific collaboration in Europe has been found to have higher impact when collaborators are from many countries (Glänzel–Schubert–Czerwon 1999a; Narin–Stevens–Whitlow 1991).

* The research article was published in PLOS ONE under the CC-BY licence in 2020 and is available here: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242468>
Please cite the article as: Csomós, G. – Vida, Z.V. – Lengyel, B. (2020) *Exploring the changing geographical pattern of international scientific collaborations through the prism of cities*. PLoS ONE, 15/11, e0242468. DOI10.1371/journal.pone.0242468

International collaboration in science is gaining importance (Pan–Kaski–Fortunato 2012; Hoekman J. 2012; Glänzel 2001) and is further supported by research funders, primarily in the European Union (Hoekman et al. 2013; Ovalle–Perandones et al. 2013). In recent decades, international scientific collaboration in terms of the number of co-authored papers has experienced a remarkable growth rate (Wagner–Whetsell–Leydesdorff 2017). For example, Wagner et al. (Wagner–Park–Leydesdorff 2015) found that, between 1990 and 2011, the ratio of internationally co-authored records in the SCI dataset increased from 10 to 25 percent. It is well studied that international collaboration is not only highly beneficial for participants (Barjak–Robinson 2008; Katz–Martin 1997) but in some cases (e.g., for “big science”) it is essential (Esparza–Yamada 2007; Hallonsten 2014; Jordan 1992). In addition, those publications that were produced in international research projects generally received more citations (Glänzel–Schubert 2001; Persson–Glänzel–Danell 2004). It is, however, important to notice that according to research conducted by Maisonobe et al. (2016), in many countries, domestic collaborations increased faster than international collaborations.

In this paper, we focus on international co-publication links between global leader cities in science production and evaluate the changing role of distance over the last three decades by analyzing Web of Science (WoS) data. To varying degrees, cities are major sites of science production in terms of the number of publications and citations. The question regarding how cities participate in global science is extensively analyzed in the growing field of spatial scientometrics.

The pioneer work of spatial analysis focusing on the city level was produced by Matthiessen and Schwarz (Matthiessen–Schwarz 1999), who examined the scientific strength in terms of publication output of “greater” urban regions of Europe. Since the beginning of the 2010s, this rather quantitative approach has been replaced by a new paradigm focusing on the geographical context of the production of research excellence. Bornmann et al. (2011) and Bornmann and Leydesdorff (2011, 2012) identified and mapped cities that were considered to be centers of excellence in scientific research on the basis of the size and frequency of the production of top 1% highly cited papers. Bornmann and de Moya–Anegón (2019a) mapped German cities, with most papers belonging to the 1% most frequently cited papers, within their subject area and publication year. Bornmann and de Moya–Anegón (2019b) detected hot and cold spots in the United States based on bibliometric data produced by institutions. Other researchers investigated some additional aspects of cities’ participation in science. Grossetti et al. (2014) examined the global and national deconcentration of scientific activities through the lens of cities. Csomós and Tóth (2016) investigated the spatial distribution of scientific publications produced by the industry. Csomós (2018a) examined the publication dynamics, collaboration pattern, and disciplinary profile of more than 2,000 cities worldwide, and in another contribution, Csomós (2018b) revealed factors that may influence cities’ high impact efficiency. Leydesdorff and Persson (2010) displayed co-authorship, collaboration networks between cities by using mapping and network visualization software. The article produced by Wu (2013) proposed a citation rank

based on spatial diversity in terms of cities and countries, focusing on the measurement of the spatial aspect in citation networks. Jiang et al. (2014) investigated the spatial patterns of R&D collaborations of Chinese cities by using co-patent data, and Andersson et al. (2014) revealed the internal spatial structure of China's scientific output. Catini et al. (2015) explored spatially concentrated innovation clusters within metropolitan areas by geocoding publication data. Maisonobe et al. (Maisonobe et al. 2017, Maisonobe–Jégou–Cabanac 2018; Maisonobe–Jégou–Eckert 2018) investigated cities' publication output and collaboration network from different aspects. They found “that cities located in scientific emerging countries tended to favor domestic interurban co-authorships whereas cities located in more traditionally English-speaking countries internationalized” (Maisonobe et al. 2017).

These papers cover varying topics in the domain of spatial scientometrics, focusing on the city level.

However, scientific interaction between cities has been analyzed by only a few papers (Maisonobe et al. 2017; Csomós–Lengyel 2019; Ma et al. 2014), primarily due to the problems of data collection and processing (Csomós 2020).

We make several new contributions to this literature. We document that the intensity of inter-national collaborations of cities is gradually growing over time. In the meantime, we observe a shift in the average geographical distance that occurs for both low- and high-intensity city–city collaborations as well. However, this shift leaves distance decay unchanged: the average distance of collaborations decreases monotonically as the intensity of collaborations between cities increases, and the pattern is stable over time. Most importantly, we find that a similar distance decay spans over larger distances for high-impact collaborations, meaning that the most important global collaborations require large geographical coverage. Results show that cities located in the European Union construct the most intense international research collaboration. Since the mid-2000s, the creation of the European Research Area (ERA) and the enlargements of the Community have given significant impetus to the deepening of intracontinental research collaborations. However, major scientific actors, that is the United States, the European Union, and Japan, tend to carry out big science projects separately from each other.

Materials and Methods

Data

We have collected a number of co-authored publications between the top 245 global science producer cities for the periods 1994–1996, 2004–2006, and 2014–2016. That is, we capture scientific cooperation between two cities if a publication is produced by at least two authors located in those cities; and it is international if those cities are located in different countries (Katz–Martin 1997; Luukkonen et al.

1993; Schubert–Braun 1990). We also analyzed the geography of scientific collaboration in highly cited papers (HCPs). HCPs are those papers that receive sufficient citations to belong to the top 1% of their academic fields, taking the most recent 10-year period into account.

To conduct the bibliometric analysis, the WoS database was employed, which is considered to be one of the most prestigious abstract and citation databases (Martín-Martín et al. 2018; Mikki 2010; Mongeon–Paul-Hus 2016) and is widely used for carrying out spatial analysis (see, for example, Glänzel 2001; King 2004; Bornmann–Wagner–Leydesdorff 2018; Wang–Wang–Philipsen 2017; Gazni–Sugimoto–Didegah 2012; Leydesdorff–Wagner 2008; Leydesdorff–Wagner 2009; Frenken–Hardeman–Hoekman 2009). The WoS provides four major indexing databases for journal articles (SCIE, SSCI, A&HCI, and ESCI), out of which the SCIE (Science Citation Index-Expanded) and SSCI (Social Sciences Citation Index) were employed. The SCIE and SSCI together list more than 12,800 journals and cover such broader areas as life sciences and biomedicine, physical sciences, technology, and social sciences.

We consider only those cities where at least 10,000 articles were published during 2014–2016. The publication history of these selected cities was investigated in the periods of 1994–1996 and 2004–2006 as well. Naturally, one of the major problems of spatial scientometric analysis focusing on the city level is that it is rather challenging to delineate cities according to the same spatial standards (Csomós 2020). For example, in urban geography, the name “Tokyo” can correspond to both the Tokyo Metropolis with an area of 2,200 square kilometers and a population of 14 million and the Tokyo Major Metropolitan Area covering 32,700 square kilometers and containing 36.3 million people (i.e., they may produce highly different publication outputs). To remain consistent, in our analysis, the “city” corresponds to the spatial unit that is reported by the author(s) in the affiliation field of the article, being placed between the name of the country (state/prefecture/etc.), and that of the organization the author(s) are affiliated with. *Table 1* demonstrates the distribution of selected cities across macro-regions (in addition, a more thorough explanation on the topic can be found in Section 3.3). As can be seen, more than two thirds of the cities being involved in this analysis are located in Western Europe and Northern America (with a dominance of the United States). Based on the number of cities, Asia (with the major proportion of Chinese and Japanese cities) comes third. The contribution of the European Union to the total number of cities is 37.5 percent.

Table 1: Summary statistics of cities involved in the analysis

Macro-region	Number of cities	Percentage of cities in the dataset	Number of papers, 1994-1996	Number of papers, 2004-2006	Number of papers, 2014-2016	Number of HCPs, 2014-2016
Africa	2	0.816	5,601	6,942	21,914	471
Asia	44	17.959	251,565	628,930	1,596,164	19,571
Australia	8	3.265	42,625	83,144	208,525	3,914
Eastern Europe	9	3.673	92,549	119,813	200,878	2,100
Latin America	7	2.857	28,251	74,296	160,619	1,598
Middle East	9	3.673	35,711	75,644	198,067	2,236
Northern America	74	30.204	581,229	1,273,062	1,967,595	45,110
Western Europe	92	37.551	791,967	1,223,389	2,145,969	43,846
Total	245	100.000	1,829,498	3,485,220	6,499,731	118,846
<i>European Union*</i>	92	37.551	778,285	1,211,981	2,125,969	41,862

* In the case of the European Union, the community of 28 member states is considered, irrespective of which of the periods is examined

The dataset that demonstrates the geographical classification and publication outputs of cities, as well as the number of co-authored papers and the Jaccard indexes of the top 3,000 collaboration links by each period, is available at Harvard Dataverse (<https://doi.org/10.7910/DVN/WRGHHT>). The dataset regarding the disciplinary breakdown of publications produced by cities is also available at this site.

Methods

When investigating the international scientific collaboration trends of cities, a threshold was set in the case of each period. Theoretically, the collaboration matrix of cities contains $29,890$ ($(n \times (n-1))/2$, where $n = 245$) links, out of which the maximum number of international collaboration links is $26,990$. However, most city-dyads produced a rather weak collaboration in terms of the number of co-produced articles. In addition, primarily in the first period (1994–1996), but also in the second period (2004–2006), many cities did not maintain international collaboration, and they only collaborated with their domestic peers. Therefore, it was reasonable to establish minimum collaboration values regarding each period between city-dyads,

which are as follows: 1994–1996: 10 co-produced articles per year; 2004–2006: 30 co-produced articles per year; and 2014–2016: 90 co-produced articles per year. When choosing the above threshold values, the increase in the world’s publication output was considered (68). Based on these thresholds, cities produced a total number of 3,122, and 3,111 collaboration links in the periods of 1994–1996, and 2004–2006, respectively. In the most recent period, however, the total number of collaboration links increased to 7,827, regardless of the fact that the threshold was set high. In each period, based on their relative strength, the top 3,000 collaboration links were considered. Naturally, it is highly likely that if two cities, irrespective of where they are located, produce high publication outputs, they will build stronger collaboration in terms of the number of co-authored publications, compared to cities with smaller publication outputs. To reduce the size effect and determine the relative strength of a particular collaboration link, the Jaccard similarity index was employed (56):

$$J_{x,y} = \frac{C_{x,y}}{(C_x + C_y - C_{x,y})} \quad (1)$$

where $J_{x,y}$ is the relative strength of a given collaboration link, $C_{x,y}$ is the number of co-produced publications of cities x and y , and C_x and C_y are the total publication outputs of city x and y , respectively.

The main reason for applying the Jaccard index for analyzing the role of distance is that we do not have access to the number of scientists located in the cities we analyze. Therefore, we cannot turn to conventional gravity models and compare the observed volume of collaboration with the potential number of collaborations (Liben-Nowell et al. 2005) or with the expected number of collaborations retrieved from regression estimation (Tóth et al. 2020). Instead, the Jaccard-like measures, in which the strength of nodes are used to scale down the dyad weight, have been shown to produce distance-decay patterns (Lengyel et al. 2015) and are therefore appropriate for our problem.

Let us take the example of the Boston–London pair to illustrate the relation between the Jaccard coefficient and the raw number of collaborations. In the period of 2014–2016, there were 4,735 co-authored publications including authors from Boston and London, which is the second-highest number of collaborations. In that period, with 153,725 publications, London’s output was the second largest in the world, and Boston was ranked fifth with 105,769 publications. After calculating the relative strength of that collaboration of the city-dyad ($J_{x,y} = 0.018586$), it turned out that it was occupying only 207th place in the ranking.

We consider it important to demonstrate how high-impact research collaborations relate to geographical distance, and whether geographical proximity affects the intensity of those collaborations. In the past few years, a number of studies have been published focusing on identifying and ranking the centers of excellence across the world (see, e.g., Bornmann 2011; Bornmann–Leydesdorff 2011; Glänzel et al. 2009; Tijssen–Leeuwen 2006; Tijssen–Visser–Van Leeuwen 2002). In

those works, the number and/or ratio of highly cited papers are employed as a proxy to express research excellence. In the case of the period of 2014–2016, we compared the relative strengths of international scientific collaborations based on the Jaccard indexes derived from all papers, and HCPs exclusively.

By employing this method, it was possible to compare how international scientific collaboration between cities had developed over time. In addition, in the case of the period of 2014–2016, it was also examined how cities participate in the production of highly cited papers; that is, which of the collaboration links were considered the relatively strongest when producing excellent papers. Finally, each collaboration link was mapped to explore the changes in the geographical pattern of those collaborations.

Results

Distance and intensity of international scientific collaboration between cities

The intensity of international collaboration between cities, measured by the Jaccard index, has witnessed a rather small change from 1994–1996 ($\mu_j = 0.003547$) to 2004–2006 ($\mu_j = 0.004599$), but the magnitude of the increase was observed by 2014–2016 ($\mu_j = 0.013212$). These results illustrated in Fig 1A suggest that besides the previously reported general rise of international scientific collaboration (Wagner–Whetsell–Leydesdorff 2017), the pairwise intensity of city-city collaboration has increased since the mid-2000s. In other words, not only the magnitude of international collaboration has risen, but also its intensity of collaboration already controlled for the size effect of cities.

The geographical reach of intensifying international collaboration has widened, while distance decay remained an important factor of collaboration intensity between two cities. Fig 1B demonstrates that distance decay curves have shifted up and to the right as well over the decades. For example, the mean distance of the weakest collaboration links at a $10^{-2.7}$ Jaccard value had covered approximately 6,000 kilometers on average in 1994–1996, which almost increased to 8,000 kilometers by 2004–2006. At the same time, the general increase of Jaccard shifts the decay curves to the right: the smallest value of Jaccard $10^{-1.9}$ in 2014–2016, for which the average distance is around 6,000 kilometers. These observations mean that recently much stronger collaborations (in terms of the number of co-produced papers scaled down by city production) have been established between cities even if they are located at an increased distance from each other. Yet, taking each time period, even the latest one into account, the mean distance curves are sloping downwards from the lowest Jaccard index category to the highest one. This finding implies that those cities that are located a further distance from each other, particularly if they are located on different continents, establish relatively less intense scientific cooperation in the given period. The stable patterns of distance decay are

due to the dominance of European collaborations with other European and Northern American cities (on this issue, see a more thorough explanation in Section 3.3).

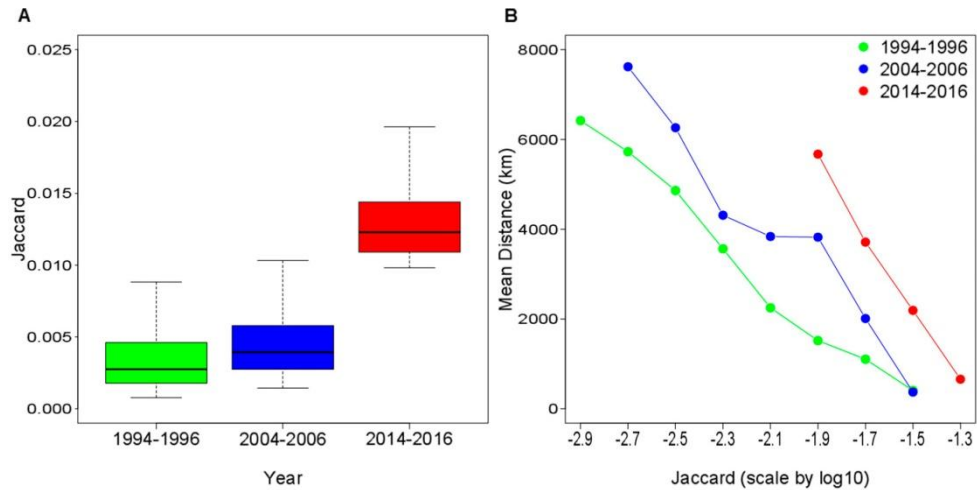


Figure 1: Intensity and distance decay of international scientific collaboration between cities
A. Changing intensity of international scientific collaboration between cities is reflected by the distribution of the Jaccard index. **B.** Changing mean distance of top 3,000 collaboration links between cities in the periods of 1994–1996, 2004–2006, and 2014–2016.

In *Fig 2*, we compare the collaboration intensity of publications that belong to the top 1% based on the number of citations they received with the general collaboration patterns in the 2014–2016 period. As can be seen in *Fig 2A*, the distribution of inter-city collaboration intensity is much higher in the case of HCP production. In the period of 2014–2016, the mean Jaccard index of the top 3,000 collaboration links producing HCPs ($\mu_j = 0.068663$) was more than five times higher than that of the top 3,000 collaboration links. These findings suggest that the production of HCPs that are deemed to be the outcomes of large-scale research projects requires deeper cooperation from international actors.

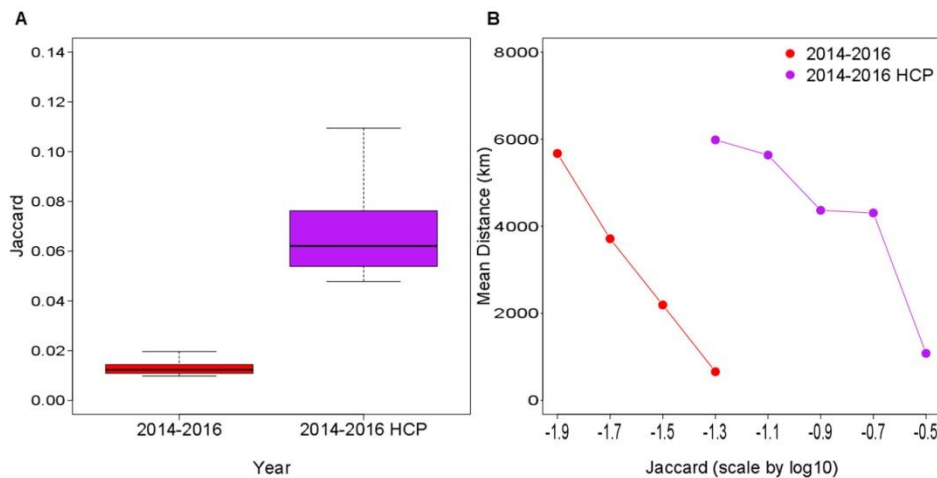


Figure 2: Comparing intensity of international scientific collaborations between cities in the cases of all papers and HCPs exclusively

- A.** International collaboration intensity is higher for HCP than for general papers.
B. Mean distances of the top 3,000 collaboration links if all papers and HCPs are considered, respectively.

In addition, as demonstrated in Fig 2B, in the case of HCPs, international collaborations between cities are less dependent on the effect of geographical proximity. The mean-distance curve of the HCP collaborations starts from an approximately same mean distance level but at a higher Jaccard index category than the general curve. Hence, in the case of HCPs, stronger international collaborations are created between cities at distances that allow for relatively weak collaborations otherwise. Naturally, geographical proximity still matters in the case of HCP collaborations; the curve of those collaborations follows a similar slope.

Disciplinary profiles of high scientific impact

Citation trends differ across disciplines. Therefore, we examine the distribution of papers across major scientific fields considering the difference between average- and high-impact papers and examine the disciplinary profiles of cities. This approach will enable us to better explain what is behind the continental distribution of high-impact collaboration in the next section.

Fig 3 illustrates that taking all papers from selected cities, the largest proportion of papers are published in the fields of life sciences, physical sciences, and technology (the classification is based on the WoS Research Area classification). The contribution ratio of these fields to the total output of the 245 cities is 91 percent, with life sciences holding the dominant position. If focusing on the HCP outputs, the fields of life sciences, physical sciences, and technology produce almost

the same contribution ratio, but their internal ratios have changed: fewer HCPs are published in the field of life sciences, whereas the field of physical sciences produces more HCPs. In addition, the contribution ratio of social sciences is much smaller in the case of HCPs, whereas that of multidisciplinary sciences has increased.

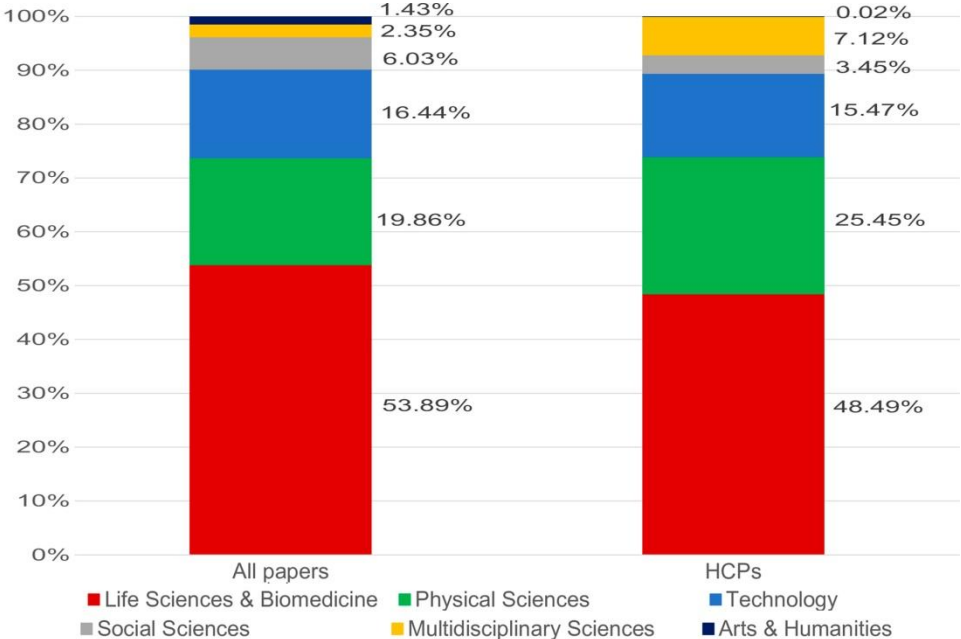


Figure 3: Disciplinary profile of selected cities' publication outputs
A. For all papers. **B.** For papers with top 1% citations.

In the WoS, the research papers published in multidisciplinary journals, such as Nature, Science, PNAS, and PLoS ONE, are classified as multidisciplinary, irrespective of their exact disciplinary profiles. Approximately 47–48 percent of the research papers published in Nature and Science can be classified in the field of life sciences and 21–22 percent of them belong to the fields of physics and chemistry (i.e., physical sciences); in contrast, nearly 95 percent of the research papers in PNAS are published in the field of life sciences (Glänzel–Schubert–Czerwon 1999b). In sum, the growing share of multidisciplinary sciences among HCPs intensifies the overrepresentation of life sciences and physical sciences.

Examining the breakdown of broader disciplinary categories enables us to realize that the majority of top disciplines in terms of the number and share of HCPs belong to the fields of life sciences, physical sciences, and technology (Table 2). These disciplines combined with multidisciplinary sciences (i.e., 20 out of the 233 disciplines) provide more than 50 percent of HCPs that have been produced in the selected cities over 2014–2016.

Table 2. *Disciplinary classification of HCPs produced by cities*

Disciplines	Broader discipline categories	Total number of HCPs produced in 2014–2016	Ratio of HCPs (%)	Number of cities in which particular HCPs are produced
Multidisciplinary Sciences		13,214	7.119	245
General Internal Medicine	Life Sciences	12,037	6.485	243
Oncology	Life Sciences	8,203	4.419	241
Chemistry, Multidisciplinary	Physical Sciences	7,402	3.988	237
Physics, Particles & Fields	Physical Sciences	7,055	3.801	214
Astronomy & Astrophysics	Physical Sciences	6,501	3.502	222
Materials Science, Multidisciplinary	Technology	5,240	2.823	232
Environmental Sciences	Life Sciences	4,128	2.224	239
Physics, Multidisciplinary	Physical Sciences	4,116	2.217	231
Chemistry, Physical	Physical Sciences	4,058	2.186	224
Cardiac & Cardiovascular Systems	Life Sciences	4,053	2.183	225
Physics, Applied	Physical Sciences	3,695	1.991	225
Nanoscience & Nanotechnology	Technology	3,645	1.964	217
Biochemistry & Molecular Biology	Life Sciences	3,536	1.905	236
Cell Biology	Life Sciences	3,162	1.703	239
Neurosciences	Life Sciences	3,010	1.622	224
Public Environmental Occupational Health	Life Sciences	2,956	1.593	224
Engineering, Electrical Electronic	Technology	2,554	1.376	197
Physics, Condensed Matter	Physical Sciences	2,533	1.365	207
Energy Fuels	Technology	2,455	1.323	216
Other Disciplines		82,067	44.212	
Total (233 disciplines)		185,620	100.000	

Recently, research projects both in the case of life science disciplines, such as oncology and neuroscience, and physical science disciplines, including particle physics, astronomy, and astrophysics have been carried out by international collaborations in large-scale research teams. Some of these projects, particularly those carried out in various branches of physics are often labeled as “big science” projects because they are highly complex and expensive and require a research team of hundreds or thousands of scientists and engineers, as well as major research infrastructure, including research facilities, machines, and services (Castelvecchi 2015; Cronin 2001; Hallonsten 2016). In addition, over the past 50 years, research projects, particularly in the fields of natural sciences and life sciences but also in technology and social sciences, have been experiencing a substantial increase in terms of team size (Milojevic 2014; Hsiehchen–Espinoza–Hsieh 2015). A study by Larivière et al. (2015) found that “collaborative research results in higher citation rates”; that is, those papers that are produced by large teams will receive more citations and are, thus, more likely to be highly cited, in contrast to those being produced by single authors or small research teams (Abt HA. 2017; Wuchty–Jones–Uzzi 2007). The nexus between team size and citation rates is reinforced by Wu et al. (Wu–Wang–Evans 2019), who assert that ten-person teams are 50% more likely to score a high-impact paper than those produced by solo authors and small research teams.

Big science and many large-scale research projects are typically carried out in international collaborations. For example, the Manhattan Project (1942–1946), which is generally accepted to be the earliest big science project (Hughes 2002; Schatz 2014) was coordinated by the United States and supported by the United Kingdom and Canada. Following projects in the fields of particle physics, astronomy, and astrophysics using the infrastructure of such mega research facilities as the Large Hadron Collider operated by the pan-European research organization, CERN (Giudice 2012), the Spallation Neutron Source located in the Oak Ridge National Laboratory, Tennessee (Crease 2019), and the Very Large Array of the National Radio Astronomy Observatory in New Mexico. In life sciences and biology, the Human Genome Project was the world’s largest collaborative project, taking place between 1990 and 2003 (Hood–Rowen 2013; Lander et al. 2001). This was followed by such highly complex multinational research projects as the Human Epigenome Project in the field of epigenomics (Bradbury 2003), and the European Union’s flagship neuroscience project, the Human Brain Project, launched in 2013 (Underwood 2016). In addition, there is evidence that international research collaboration has also been becoming increasingly important in the fields of medicine (Butrous 2008), cancer research (Stefan–Seleiro 2016; Tang et al. 2019), and neuroscience (Koch–Jones 2016).

Geographical patterns and global regions in city-city collaborations

Now, we turn to investigate the detailed geographical patterns of international scientific collaboration between cities and pay special attention to continental

distributions. To outline the changes in the geographical pattern of international scientific collaborations between cities and investigate the patterns of high-impact collaborations, we classified each link into quarters based on the Jaccard index (Table 3). Each quarter contains 750 collaborations links.

Table 3: Classification of collaboration links into quarters by periods

	1994–1996				2004–2006			
	Jaccard category quartile	Mean distance (km)	Standard deviation of distance (km)	Mean Jaccard index	Jaccard category quartile	Mean distance (km)	Standard deviation of distance (km)	Mean Jaccard index
Q4	0.000783 – 0.001780	6,140	3,927	0.0014	0.001450 – 0.002750	6,934	3,782	0.0022
Q3	0.001780 – 0.002760	5,017	3,644	0.0022	0.002750 – 0.003940	5,257	3,878	0.0033
Q2	0.002760 – 0.004610	3,934	3,595	0.0036	0.003940 – 0.005790	3,603	3,530	0.0048
Q1	0.004610 – 0.022800	2,149	2,624	0.0070	0.005790 – 0.021000	3,860	3,446	0.0081
	2014–2016				2014–2016 HCP			
	Jaccard category quartile	Mean distance (km)	Standard deviation of distance (km)	Mean Jaccard index	Jaccard category quartile	Mean distance (km)	Standard deviation of distance (km)	Mean Jaccard index
Q4	0.009820 – 0.010900	6,019	4,208	0.0104	0.047800 – 0.053900	5,832	4,446	0.0506
Q3	0.010900 – 0.012300	5,449	4,319	0.0116	0.053900 – 0.062100	5,820	4,463	0.0577
Q2	0.012300 – 0.014400	4,106	3,989	0.0133	0.062100 – 0.076200	5,431	4,492	0.0685
Q1	0.014400 – 0.045300	3,123	3,618	0.0176	0.076200 – 0.212000	4,262	4,147	0.0977

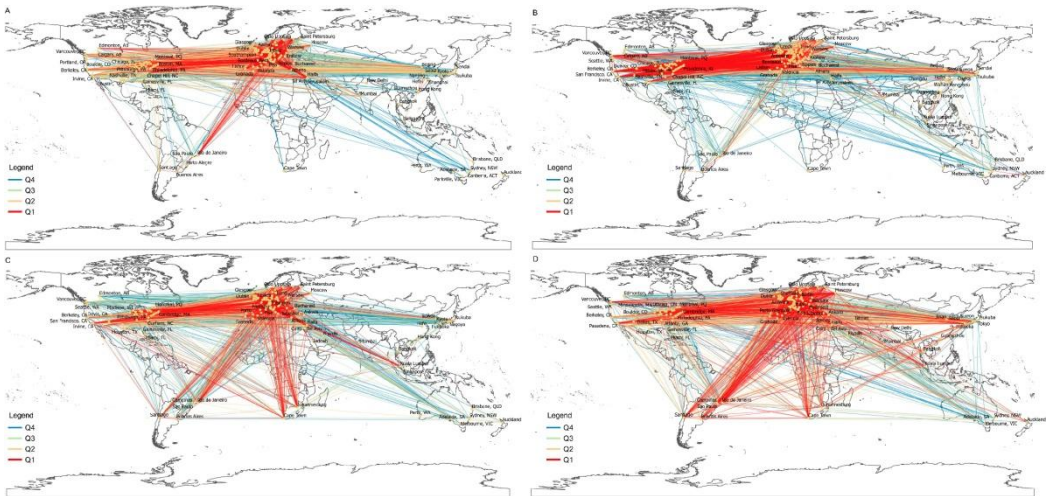


Figure 4: Geographical pattern of international scientific collaborations between cities based on the Jaccard index (Quartile ranges are reported in Table 3).

A. 1994–1996. **B.** 2004–2006. **C.** 2014–2016. **D.** 2014–2016 HCP

Fig 4 illustrates the international scientific collaborations between cities by quarters of the Jaccard index. A clear observation is the increase of collaboration links between Western European cities and Northern American and Asian cities from 1994–1996 to 2004–2006 (Figs 4A and 4B). This observation is in line with previous findings on the rapid globalization of science (Gui–Liu–Du 2019; Waltman–Tijssen–Eck 2011). Second, in the case of the Q1 (the strongest) collaboration links, between 1994–1996 and 2004–2006, the Western European–United States links became dominant among the strongest city–city links. By the period of 2014–2016, the strongest inter-city links became more diffused across continents, with an emerging presence of African, Latin American, and Middle Eastern cities due to which the ranks of some links among Northern America, Europe, and East Asia lowered (Fig 4C). In contrast, high-impact collaborations across Northern America, Europe, and Asia were ranked higher than average collaborations, whereas collaboration between Europe and the emergent cities in Latin America and Africa did not lose importance compared to average collaborations (Fig 4D).

To examine more closely the emergence of cities in the strongest international collaborations, we aggregate the number of links by continents and macro-regions and report the ratios of these aggregates in Table 4. In each period, the share of Western Europe was highest, which is not particularly surprising because the highest number of cities in the network are from Western Europe (i.e., 92 cities, the 37.55 percent of all cities in the dataset). However, the dominance of Western Europe is even larger in Q1 collaboration links, signaling that international collaboration is a European phenomenon, which is partly due to the large number of cities distributed across many countries in Europe. However, the dynamics of the

network implies that this dominance is not automatic. In the first period, EU cities had 3,469 links that decreased to 3,249 by 2004–2006 but rose again to 3,675 by 2014–2016. EU cities (both Eastern and Western) have even a larger share among the strongest international collaborations that have high Jaccard values as well (Q1).

Table 4: International collaboration links of cities by macro-regions

	Ratio of collaboration links by macro-regions (%)				Ratio of Q1 collaboration links by macro-regions (%)			
	1994–1996	2004–2006	2014–2016	2014–2016, HCPs	1994–1996	2004–2006	2014–2016	2014–2016, HCPs
Africa	0.050	0.067	2.317	2.033	0.000	0.000	3.267	2.733
Asia	5.917	8.267	2.583	5.850	0.867	2.600	1.467	3.867
Australia	0.517	1.033	1.333	1.583	0.000	0.400	0.000	0.267
Eastern Europe	7.533	5.683	10.417	11.733	10.067	4.267	13.533	20.400
Latin America	1.600	1.700	3.917	6.383	1.867	0.667	2.600	7.667
Middle East	1.783	0.467	2.900	5.033	1.867	0.000	1.600	6.133
Northern America	24.467	27.817	16.933	15.500	8.333	26.933	7.867	8.067
Western Europe	58.133	54.967	59.600	51.883	77.133	65.133	69.667	50.867
TOTAL	100.000	100.000	100.000	100.000	100.000	100.000	100.000	100.000
<i>European Union*</i>	<i>57.817</i>	<i>54.150</i>	<i>61.250</i>	<i>54.683</i>	<i>78.933</i>	<i>63.533</i>	<i>71.400</i>	<i>57.267</i>

* In the case of the European Union, the community of 28 member states is considered, irrespective of which of the periods is examined

This dynamic for Northern American and Asian cities was the opposite over the three decades. Northern American cities had 1,468 links in 1994–1996, 1,669 links in 2004–2006, and 1,016 links in 2014–2016. Asian cities have increased their links from 1994–1996 to 2004–2006 in China (82→137), Japan (170→191), and South Korea (55→119). However, and irrespective of the increasing participation of East Asian cities in international scientific collaborations, most of these links have low Jaccard values (Fig 4B). More surprisingly, there were only five links of Chinese cities among the strongest international collaborations in 2014–2016, none of which had a high Jaccard. This finding is due to the fact that recently, Chinese cities have experienced a substantially more robust increase in their total publication output as compared to the number of their internationally co-authored publications (i.e., in the case of Chinese cities, the value of the Jaccard index has become smaller over time) (Csomós 2018a; Maisonobe–Jégou–Cabanac 2018; Nature Index 2018). In contrast to the Northern American and Asian trends, the ratio of African, Latin American, and Middle Eastern links has risen by 2014–2016.

The European Union dominates the international collaboration of cities partly due to the research policy of the Community. In the beginning of the 2000s, the European Research Area (ERA) was established, which was motivated by efficiency gains of developing a pan-European science base instead of coordinating national efforts in order to avoid lagging behind other major global players and create a “new European-level funding mechanism to support the very best research carried out at the frontiers of knowledge” (European Commission 2005; 2019; European Parliament 2017; Nedeva–Stampfer 2012). Since the launch of the Sixth Framework Programme (2002–2006), the funding instrument to support and foster the construction of the ERA, the key goals of research funding are deepening the research collaboration between institutions located in the Member States (Ortega–Aguillo 2010; Hoekman–Frenken–Tijssen 2010). However, critics argue that the distance decay of research collaboration in Europe is a sign that ERA is not functioning optimally (Hoekman–Frenken–Tijssen 2010).

Comparing the participation of macro-regions in high-impact collaborations with participation in lower-impact collaborations, we observe that participation of the European Union, Western Europe in particular, and Northern America underperform, whereas Latin America, the Middle East, and Eastern Europe account for high ratios, particularly when links of the highest Jaccard values are only considered. Certainly, the Q1 set of links is biased towards cities that have few links, which produces an even stronger underrepresentation of Northern America and Western Europe, an important artifact of the analysis to keep in mind.

In the following section, we build on the findings in Section 3.2 and attempt to interpret the distribution of high-impact international collaboration as a result of big science and large-scale research projects because the highest proportion of high-impact international collaborations are materialized in fields where big science and other highly complex research projects are increasingly dominant. We offer three interpretations of high-impact distributions.

First, recently, core regions (i.e., the United States and the European Union) tend to establish more intensive research collaboration with developing countries involving researchers from the latter ones to participate in big science and other highly complex research projects (Tang et al. 2019; Ellis 2003; Harris 2004; U.S. Congress 1995). This collaboration is important for core regions because some infectious diseases (e.g., Ebola and Malaria), geological phenomenon, and environmental problems can be best studied in developing countries, which requires the participation of local experts and researchers (Harris 2004). In addition, the involvement of developing countries in collaborative projects can serve to improve international political stability as well as transfer vital skills and technologies to other parts of the world (U.S. Congress 1995).

Second, the ERA is an effective tool for producing strong collaboration links for high-impact output. The relatively large ratio of Eastern Europe in such projects is a sign of this ability. Further, findings presented in Section 3.1 suggest that most intensive high-impact collaboration occurs across cities that are less than 4,000 kilometers away from each other on average, suggesting that there are many such links across European cities.

Third, Northern America remains relatively isolated from international high-impact science. US cities are high-impact producers themselves, which decreases the relative importance of collaborations. In addition, due to the fact that the United States has the largest science system in the world with many actors (e.g., universities, research institutes, and corporate labs) within that system, the ratio of the national collaboration is remarkably high (Melin 1999). In the case of big science, even the traditionally strong connections between the United States and Western European cities (Kato–Ando 2017; Leydesdorff et al. 2013) become less cooperative, and the two large science systems tend to carry out such large-scale research projects in parallel. For example, two neuroscience initiatives were launched in 2013 with almost equal budgets: the BRAIN Initiative of the US National Institutes of Health and the Human Brain Project, the flagship project of the European Commission (Abbott 2013; Theil 2015). Similar parallel investments occurred in the construction of next-generation neutron sources that an OECD report in 1998 strongly recommended to carry out in America, Europe, and Asia (Garoby 2018). In 2006, the United States put the SNS, a pulsed spallation neutron source into operation in the Oak Ridge National Laboratory, and in 2009 Japan followed it with the Japan Proton Accelerator Research Centre in Tokai. In the European Union, the Lund-based European Spallation Source (ESS) is currently under construction and is intended to be the world's most powerful next-generation neutron source.

As per the official statements coming from representatives located on both sides, the United States and the European Union are committed to maintaining strong trans-Atlantic scientific cooperation (European Commission 2006; Zerhouni–Potočnik 2008). That is, the question remains: What is the reason for the United States and the European Union each intending to run big science projects with similar scientific goals in parallel and not in cooperation if the collaboration is supported by (science) politicians? In fact, if digging more deeply, we can find evidence of sharp competition between the United States and the European Union. Taking beam physics as an example, Kaiserfeld adds (2013) that “when European expressed hopes that the new spallation sources in Japan and the US might also accommodate the need for neutrons among European scientists, representatives from the SNS and the US Department of Energy ‘firmly contradicted’ them”. As a matter of fact, “competition was the word now used to inject courage into the struggling ESS project—not competition between European countries, but between Europe and other countries.” It is assumed that such competition exists in other fields as well, and it could be one reason why there is only weak relative collaboration between US and EU cities.

Discussion and Conclusions

In this paper, we demonstrated the effect of geographical proximity on the relative strength of international scientific collaboration between cities over time. Our research was centered on three research questions:

1) Due to multiple factors, but the rapid development of information and communication technologies in the first place, the intensity of international scientific collaboration has significantly increased recently; yet, geographical proximity has still remained a restrictive factor for the actors involved in scientific cooperation.

First, we found that, in the past 30 years, particularly since the mid-2000s, the relative strength of international scientific collaboration in general had increased to a significant extent; that is, as compared to their total publication outputs, cities tend to produce a growing number of internationally co-authored publications. Second, the mean geographical distance of international scientific collaborations between cities, even in the case of the relative strongest collaborations, has become substantially higher over time. This finding suggests that recently, cities have been constructing scientific collaborations with their peers even if they are located at an increased geographical distance from each other. In addition, in the past two decades, a growing number of cities from developing countries has joined the scientific realm created by core countries, subsequently contributing to an increase in the mean geographical distance of collaborations. Yet, irrespective of which time period is observed, the geographical proximity still impacts the collaborations between cities. That is, by acknowledging the overall increase in the mean geographical distance of international scientific collaborations, we experienced that relatively strong collaborations still required smaller geographical distances. More precisely: the relatively strongest collaborations were generally created between cities located in neighboring countries.

2) The supranational policies fostering international scientific collaborations, particularly in the case of the European Union, help lessen the restrictive effect of geographical proximity.

Until the mid-2000s, a growing number of cities across the world, but particularly those located in Asia and Eastern Europe, had started to join the international arena of science and construct relatively strong collaborations with other cities. Furthermore, Northern American cities also became more collaborative in terms of the number of internationally co-authored papers. During the mid-1990s to the mid-2000s, it was the European Union (i.e., the totality of the old and new members of the EU-28) that experienced a decreasing number of cities in international collaborations. Then, since the mid-2000s, radical changes have taken place: Cities from the European Union have occupied the vast majority of collaboration links, whereas the ratio of Northern American cities (US cities in the first place) in those collaboration links has almost been halved and the participation ratio of Asian cities has become rather insignificant.

We propose two major reasons behind these changes. First, the European Union's largest single enlargement in terms of people and the number of countries took place in 2004, when eight Central and Eastern European (CEE) and two Mediterranean countries joined the Community. This was followed by the accession of two more CEE countries in 2007. By the mid-2010s, the European Union became the political and economic integration of 28 member states. After the accession of CEE countries to the European Union, they were able to receive

support from the EU's Structural Funds and the Cohesion Fund, allowing those countries to improve the infrastructure of their national science system and pay additional money to researchers. Second, since the beginning of the 2000s, by the establishment of the ERA, the Community has made significant efforts to reduce the fragmentation of the European research landscape, and the isolation and compartmentalization of national research systems (Busquin 2000). In addition, in 2009, the legal framework for the European Research Infrastructure Consortium (ERIC) was put in force to facilitate the establishment and operation of research infrastructure with European interest. Under the umbrella of the ERIC, a number of large-scale multi-Member States research infrastructure projects have been implemented, one of which is the ESS in Lund.

Due to these developments, international scientific collaborations between cities located in the European Union have been given significant impetus. We also found, however, that even in the case of the European Union, the geographical proximity still affects the relative strength of collaborations. This observation suggests that irrespective of the positive impact of supranational policies and the number of financial incentives, cities mostly tend to collaborate with their peers located in neighboring countries.

3) In the case of international scientific collaborations resulting in excellent papers, due to the high complexity of research projects, the restrictive effect of geographical proximity will be less significant.

To answer this question, in the period of 2014–2016, we compared the relative strength of international scientific collaborations and the geographical pattern of those collaborations in the case of all papers and highly cited papers (HCPs). The results demonstrate that, based on a disciplinary analysis of cities' outputs, the majority of HCPs are the outcomes of big science and other large-scale research projects carried out in the fields of life sciences and physical science. These projects have common features in that they require the most cutting-edge research infrastructure and the cooperation of huge researcher teams, being sometimes constituted by hundreds or thousands of individuals. In addition, due to the extremely high costs generally characterizing big science projects, they might require co-funding of multiple nations. Considering these factors, it is not surprising that in the case of HCPs, the intensity of international collaborations in terms of output vs. co-produced papers ratio is substantially higher than in the case of all papers. Another observation is that HCP collaborations are less constrained by the effect of geographical proximity; that is, cities construct relatively strong collaborations with their peers located at a significantly increased physical distance from them. Yet, from a threshold interval of 4,000–5,000 kilometers, the intensity of HCP collaborations begins to lessen. These facts suggest that the United States and the European Union (even Japan in some cases), the global leaders in science, tend to carry out big science projects on their own and not in cooperation. Now, "big science" can be labelled by such terms as prestige (Gilady 2018), nationalism (Sassower 2015), and competition (Kaiserfeld 2013). One exception is considered the quite successful Human Genome Project (HGP), which was carried out in

collaboration with major scientific actors (i.e., the United States, some member states of the European Union, and China and Japan). The implementation of the HGP demonstrates the manner in which big science should be approached to surmount the challenges posed by the new coronavirus (COVID-19) (Berkley 2020; Cohen 2020).

In addition, by examining the impact of distance on big science collaborations, a further research question emerges. Big science projects can be carried out either in a research lab that has a specific geographical location (e.g., CERN and the Oak Ridge National Laboratory), or by international research teams of whom members are located geographically separated (e.g., the HGP and the Human Brain Project). Gibbons et al. (Gibbons et al. 2015) suggest that due to the development of such platforms as the Internet, we are now experiencing the emergence of a socially distributed knowledge production system. Follow-up research should focus on investigating the differences in the evolution of geographically concentrated big science and distributed knowledge production because these modes have varying effects on the distance of collaborations.

References

- Abbasi, A. – Jaafari, A. (2013) Research impact and scholars' geographical diversity. *J Informetr.* 7. 683–692.
- Abbott, A. (2013) Neuroscience: Solving the brain. *Nature*, 499/7458, 272–274.
- Abt HA. (2017) Citations and team sizes. *Publ Astron Soc Pac.* 129/972, 024008
- Andersson, D. E. – Gunessee, S. – Matthiessen, C. W. – Find. S. (2014) The geography of Chinese science. *Environ Plann A.*, 46, 2950–2971.
- Barjak, F. – Robinson, S. (2008) International collaboration, mobility and team diversity in the life sciences: Impact on research performance. *Soc Geogr.* 3. 23–36.
- Batty, M. (1993) The geography of cyberspace. *Environ Plann B.*, 20, 615–616.
- Berkley, S. (2020) We need a 'Big Science' approach for developing corona vaccines. *The Spectator*, April 7 (Cited 2020 April 22). <https://www.spectator.co.uk/article/we-need-a-big-science-approach-to-finding-a-corona-vaccine>
- Bornmann, L. – De Moya-Anegón, F. (2019) Hot and cold spots in the US research: A spatial analysis of bibliometric data on the institutional level. *J Inform Sci.*, 45, 84–91.
- Bornmann, L. – De Moya-Anegón, F. (2019) Spatial analysis of bibliometric data on the institutional level. *J Inform Sci.*, 45, 416–425.
- Bornmann, L. – Leydesdorff, L. (2011) Which cities produce more excellent papers than can be expected? A new mapping approach, using Google Maps, based on statistical significance testing. *J Am Soc Inf Sci Tec.*, 62, 1954–1962.
- Bornmann, L. – Leydesdorff, L. (2012) Which are the best performing regions in information science in terms of highly cited papers? Some improvements of our previous mapping approaches. *J Informetr.*, 6, 336–345.

- Bornmann, L. – Mutz, R. (2015) Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J Assoc Inf Sci Tech.*, 66, 2215–2222.
- Bornmann, L. – Wagner, C. – Leydesdorff, L. (2018) The geography of references in elite articles: Which countries contribute to the archives of knowledge? *PLoS ONE*. 13/3, e0194805.
- Bradbury, J. (2003) Human epigenome project – Up and running. *PLoS Biology*. 1/3, e82.
- Busquin, P. (2000) Addressing strategic challenges and enhancing professional skills. Nov 18 (cited 16 April 2020). In: *Ecsite 2000 Conference* (Internet). European Commission Press Corner. – (about 12 pages). https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_00_443
- Butrous, G. (2008) International cooperation to promote advances in medicine. *Ann Thorac Med.*, 3, 79–81.
- Cairncross, F. (2001) *The death of distance 2.0*. London: Texere Publishing Limited.
- Castelvecchi, D. (2015) Physics paper sets record with more than 5,000 authors. May 15 (cited 7 April 2020). In: *Nature News* (Internet). London: Nature Research – (about 5 screens). <https://www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567>
- Catini, R. – Karamshuk, D. – Penner, O. – Riccaboni, M. (2015) Identifying geographic clusters: A network analytic approach. *Res Policy*. 44, 1749–1762.
- Cohen, J. (2020) With record-setting speed, vaccinemakers take their first shots at the new coronavirus. *Science*. March 31 (Cited 2020 April 22). <https://www.sciencemag.org/news/2020/03/record-setting-speed-vaccinemakers-take-their-first-shots-new-coronavirus>
- Crease, R. P. (2019) The rise and rise of ‘new big science’. April 9 (cited 7 April 2020). In: *Physics World. Projects and Facilities. Opinion and Reviews* (Internet). Bristol: IOP Publishing – (about 6 screens). <https://physicsworld.com/a/the-rise-and-rise-of-new-big-science/>
- Cronin, B. (2001) Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *J Am Soc Inf Sci Tec.*, 52, 558–569.
- Csomós, G. (2018a) A spatial scientometric analysis of the publication output of cities worldwide. *J Informetr.*, 12, 547–566.
- Csomós, G. (2018b) Factors Influencing Cities’ Publishing Efficiency. *J Data Inf Sci.*, 3, 43–80.
- Csomós, G. (2020) On the challenges ahead of spatial scientometrics focusing on the city level. *Aslib J Inform Manag.*, 72, 67–87.
- Csomós, G. – Lengyel, B. (2019) Mapping the efficiency of international scientific collaboration between cities worldwide. *J Inform Sci.*, Apr 10. DOI: 10.1177/0165551519842128
- Csomós, G. – Tóth, G. (2016) Exploring the position of cities in global corporate research and development: A bibliometric analysis by two different geographical approaches. *J Informetr.*, 10, 516–532.

- Ellis, J. (2003) Developing countries and CERN. Jun 30 (cited 9 April 2020). In: *CERN Courier* (Internet). Bristol: IOP Publishing – (about 16 screens). <https://cerncourier.com/a/developing-countries-and-cern/>
- Esparza, J. – Yamada, T. (2007) The discovery value of “Big Science”. *J Exp Med.*, 204, 701–704.
- European Commission (2005) *Frontier Research: The European Challenge. High-Level Expert Group Report*. Brussels: European Commission. Directorate-General for Research.
- European Commission (2006) *Directorate-General for External Relations. The European Union and the United States: Global partners, global responsibilities*. Brussels: European Commission. Directorate-General for External Relations. http://eeas.europa.eu/archives/docs/us/docs/infopack_06_en.pdf
- European Commission (2019) *The European Research Area. European Commission*. https://ec.europa.eu/info/sites/info/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_factsheet-era_2019.pdf
- European Parliament (2017) *EU Research Policy: Tackling the major challenges facing European society*. Briefing: European Added Value in Action. European Parliament. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2017/598597/EPRS_BRI\(2017\)598597_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2017/598597/EPRS_BRI(2017)598597_EN.pdf)
- Frenken, K. – Hardeman, S. – Hoekman, J. (2009) Spatial scientometrics: Towards a cumulative research program. *J Informetr.*, 3, 222–232.
- Frenken, K. – Hoekman, J. – Kok, S. – Ponds, R. – Van Oort, F. – Van Vliet, J. (2009) Death of Distance in Science? A Gravity Approach to Research Collaboration. In: Pyka, A. – Scharnhorst, A. (eds): *Innovation Networks, Understanding Complex Systems*. Berlin–Heidelberg: Springer-Verlag, 43–57.
- Friedman, T. L. (2005) *The World is Flat: A Brief History of the Twenty-First Century*. New York: Farrar – Straus and Giroux.
- Garoby, R. – Danared, H. – Alonso, I. – Bargallo, E. – Cheymol, B. – Darve, C. et al. (2018) The European Spallation Source Design. *Phys Scripta.*, 93/1, 014001
- Gazni, A. – Sugimoto, C. R. – Didegah, F. (2012) Mapping world scientific collaboration: Authors, institutions, and countries. *J Am Soc Inf Sci Tec.*, 63, 323–335.
- Gibbons, M. – Limoges, C. – Nowotny, H. – Schwartzman, S. – Scott, P. – Trow, M. (2015) *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London–Thousand Oaks: Sage.
- Gilady, L. (2018) *The Price of Prestige: Conspicuous Consumption in International Relations*. Chicago–London: University of Chicago Press.
- Giudice, G. F. (2012) Big Science and the Large Hadron Collider. *Physics in Perspective*, 14, 95–112.
- Glänzel, W. (2001) National characteristics in international scientific co-authorship relations. *Scientometrics*, 51, 69–115.
- Glänzel, W. – Schubert, A. (2001) Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50, 199–214.

- Glänzel, W. – Schubert, A. – Czerwon, H-J. (1999a) A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, 45, 185–202.
- Glänzel, W. – Schubert, A. – Czerwon, H-J. (1999b) An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics*, 44, 427–439.
- Glänzel, W. – Thijs, B. – Schubert, A. – Debackere, K. (2009) Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, 78, 165–188.
- Grossetti, M. – Eckert, D. – Gingras, Y. – Jégou, L. – Larivière, V. – Milard, B. (2014) Cities and the geographical deconcentration of scientific activity: A multilevel analysis of publications (1987–2007). *Urban Stud.*, 51, 2219–2234.
- Gui, Q. – Liu, C. – Du, D. (2019) Globalization of science and international scientific collaboration: A network perspective. *Geoforum*, 105 1–12.
- Hallonsten, O. (2014) How expensive is Big Science? Consequences of using simple publication counts in performance assessment of large scientific facilities. *Scientometrics*, 100, 483–496.
- Hallonsten, O. (2016) *Big Science Transformed: Science, Politics and Organization in Europe and the United States*. London: Palgrave Macmillan.
- Harris, E. (2004) Building scientific capacity in developing countries. *EMBO Rep.*, 5, 7–11.
- Hoekman, J. (2012) Science in an age of globalisation: the geography of research collaboration and its effect on scientific publishing. Ph.D. Thesis. Technische Universiteit Eindhoven. <https://research.tue.nl/en/publications/science-in-an-age-of-globalisation-the-geography-of-research-coll>
- Hoekman, J. – Frenken, K. – Tijssen, R. J. W. (2010) Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Res Policy*, 39, 662–673.
- Hoekman, J. – Scherngell, T. – Frenken, K. – Tijssen, R. (2013) Acquisition of European research funds and its effect on international scientific collaboration. *J Econ Geogr.*, 13, 23–52.
- Hood, L. – Rowen, L. (2013) The human genome project: Big science transforms biology and medicine. *Genome Med.*, 5/9, 79.
- Hsiehchen, D. – Espinoza, M. – Hsieh, A. (2015) Multinational teams and diseconomies of scale in collaborative research. *Sci. Adv.*, 1/8, 1500211.
- Hughes, J. (2002) *The Manhattan Project and the Birth of Big Science*. New York: Columbia University Press.
- Jiang, S. – Shi, A. – Peng, Z. – Li, X. (2014) Major factors affecting cross-city R&D collaborations in China: evidence from cross-sectional co-patent data between 224 cities. *Scientometrics*, 111, 1251–1266.
- Jones, B. F. – Wuchty, S. – Uzzi, B. (2008) Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, 322/5905, 1259–1262.

- Jordan, E. (1992) Invited editorial: The human genome project: Where did it come from, where is it going? *Am J Hum Genet.*, 51, 1–6.
- Kaiserfeld, T. (2013) The ESS from neutron gap to global strategy. Plans for an international research facility after the cold war. In: Kaiserfeld, T. – O'Dell, T. (eds.): *Legitimizing ESS: Big Science as a Collaboration Across Boundaries*. Lund: Nordic Academic Press, 25–42.
- Kato, M. – Ando, A. (2017) National ties of international scientific collaboration and researcher mobility found in Nature and Science. *Scientometrics*, 110, 673–694.
- Katz, J. S. (1994) Geographical proximity and scientific collaboration. *Scientometrics*, 31, 31–43.
- Katz, J. S. – Martin B. R. (1997) What is research collaboration? *Res Policy*, 26, 1–18.
- King, D. A. (2004) The scientific impact of nations. *Nature*, 430/6997, 311–316.
- Koch, C. – Jones, A. (2016) Big Science, Team Science, and Open Science for Neuroscience. *Neuron*, 92, 612–616.
- Kolko, J. (2000) The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage. In: Vogelsang, I. – Compaine, B. J. (eds.): *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*. Cambridge, MA: MIT Press, 73–98.
- Lambiotte, R. – Blondel, V. D. – De Kerchove, C. – Huens, E. – Prieur, C. – Smoreda, Z. – Van Dooren, P. (2008) Geographical dispersal of mobile communication networks. *Physica A*, 387, 5317–5325.
- Lander, E. S. – Linton, L. M. – Birren, B. – Nusbaum, C. – Zody, M. C. – Baldwin, J. – Devon, K. et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, 409/6822, 860–921.
- Larivière, V. – Gingras, Y. – Sugimoto, C. R. – Tsou, A. (2015) Team size matters: Collaboration and scientific impact since 1900. *J Am Soc Inf Sci Tec.*, 66, 1323–1332.
- Lengyel, B. – Varga, A. – Ságvári, B. – Jakobi, Á. – Kertész, J. (2015) Geographies of an online social network. *PLoS ONE*, 10/9, e0137248.
- Leydesdorff, L. – Persson, O. (2010) Mapping the Geography of Science: Distribution Patterns and Networks of Relations Among Cities and Institutes. *J Am Soc Inf Sci Tec.*, 61, 1622–1634.
- Leydesdorff, L. – Wagner, C. S. (2008) International collaboration in science and the formation of a core group. *J Informetr.*, 2, 317–325.
- Leydesdorff, L. – Wagner, C. S. (2009) Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics*, 78, 23–36.
- Leydesdorff, L. – Wagner, C. – Park, H-W. – Adams J. (2013) International collaboration in science: The global map and the network. *Prof Inform.*, 22, 87–95
- Liben-Nowell, D. – Novak, J. – Kumar, R. – Raghavan, P. – Tomkins, A. (2005) Geographic routing in social networks. *Proc Natl Acad Sci USA*, 102, 11623–11628.

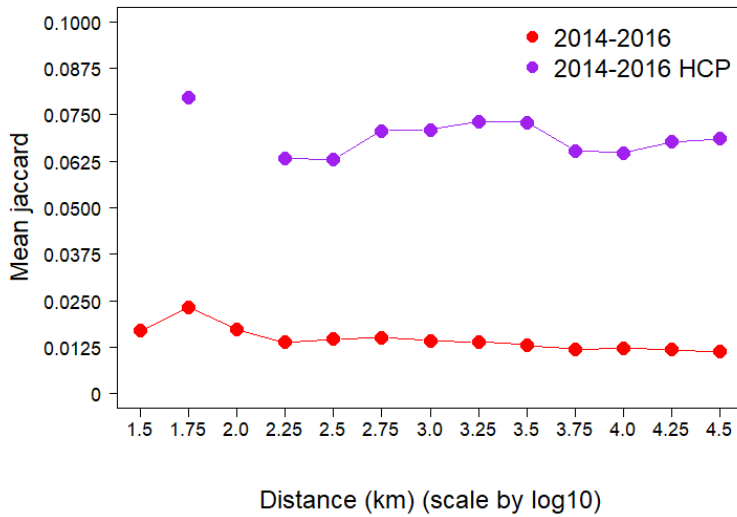
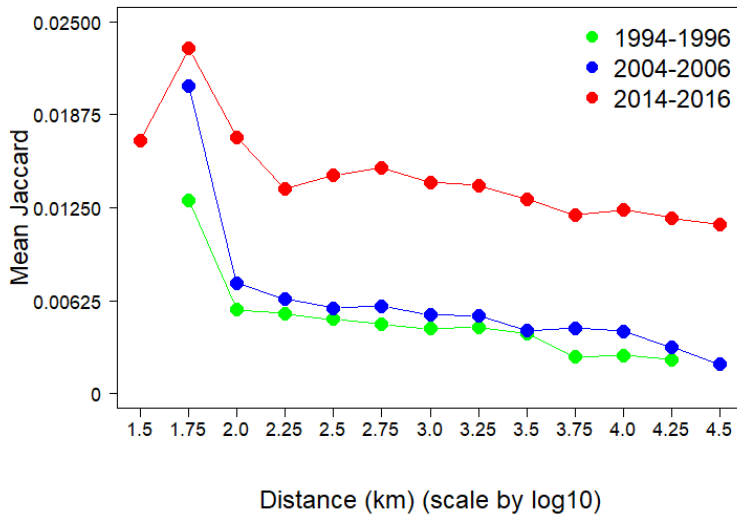
- Luukkonen, T. – Tijssen, R. J. W. – Persson, O. – Sivertsen, G. (1993) The measurement of international scientific collaboration. *Scientometrics*, 28, 15–36.
- Ma, H. – Fang, C. – Pang, B. – Li, G. (2014) The Effect of Geographical Proximity on Scientific Cooperation among Chinese Cities from 1990 to 2010. *PLoS ONE*, 9/11, e111705. doi: 10.1371/journal.pone.0111705
- Maggioni, MA. – Uberti, T. E. (2009) Knowledge networks across Europe: Which distance matters? *Ann Regional Sci.* 43 (3 SPEC. ISS.), 691–720.
- Maisonobe M. – Eckert, D. – Grossetti, M. – Jégou, L. – Milard, B. (2016) The world network of scientific collaborations between cities: domestic or international dynamics? *J Informetr.*, 10, 1025–1036.
- Maisonobe, M. – Grossetti, M. – Milard, B. – Jégou, L. – Eckert, D. (2017) The global geography of scientific visibility: a deconcentration process (1999–2011). *Scientometrics*, 113, 479–493.
- Maisonobe, M. – Jégou, L. – Cabanac, G. (2018) Peripheral Forces Nature Index 2018. Science Cities. *Nature*, 563, S18–S19.
- Maisonobe, M. – Jégou, L. – Eckert, D. (2018) Delineating urban agglomerations across the world: a dataset for studying the spatial distribution of academic research at city level. *Cybergeo*, 871.
- Martín-Martín, A. – Orduna-Malea, E. – Thelwall, M. – Delgado López-Cózar, E. (2018) Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *J Informetr.*, 12, 1160–1177.
- Masselot, A. (2016) Geo localizing Medline citations. Jan 12 (cited 2020 March 21). *Le Blog de Octos* (Internet) – (about 22 screens). <https://blog.octo.com/en/geo-localizing-medline-citations/>
- Matthiessen, C. W. – Schwarz, A. W. (1999) Scientific Centres in Europe: An Analysis of Research Strength and Patterns of Specialisation Based on Bibliometric Indicators. *Urban Stud.*, 36, 453–477.
- Matthiessen, C. W. – Schwarz, A. W. – Find, S. (2002) The Top-level Global Research System, 1997–99: Centres, Networks and Nodality. An Analysis Based on Bibliometric Indicators. *Urban Stud.*, 39, 903–927.
- Melin, G. (1999) Impact of national size on research collaboration: A comparison between Northern European and American universities. *Scientometrics*, 46, 161–170.
- Mikki, S. (2010) Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics*, 82, 321–331.
- Milojevic, S. (2014) Principles of scientific research team formation and evolution. *P Natl Acad Sci USA*, 111, 3984–3989.
- Mongeon, P. – Paul-Hus, A. (2016) The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, 213–228.
- Narin, F. – Stevens, K. – Whitlow, E. S. (1991) Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21, 313–323.

- Nature Index (2018). Earth and Environmental Sciences. The top 500 cities by scientific output in the Earth and Environmental Sciences from 2015 to 2017. 558/7711.
- Nedeva, M. – Stampfer, M. (2012) From “science in Europe” to “European science”. *Science*, 336/6084, 982–983.
- O'Brien, R. (1992) *Global Financial Integration: The End of Geography*. London: Pinter.
- Ohmae, K. (1995) *The Borderless World: Power and Strategy in an Interdependent Economy*. New York: Harper Business.
- Ortega, J. L. – Aguillo, I. F. (2010) Shaping the European research collaboration in the 6th Framework Programme health thematic area through network analysis. *Scientometrics*, 85, 377–386.
- Ovalle-Perandones, M-A. – Gorraiz, J. – Wieland, M. – Gumpenberger, C. – Olmeda-Gómez, C. (2013) The influence of European Framework Programmes on scientific collaboration in nanotechnology. *Scientometrics*, 97, 59–74.
- Pan, R. K. – Kaski, K. – Fortunato, S. (2012) World citation and collaboration networks: uncovering the role of geography in science. *Scientific Reports*, 2,902.
- Persson, O. – Glänzel, W. – Danell, R. (2004) Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60, 421–432.
- Sassower R. (2015) *Compromising the Ideals of Science*. New York: Palgrave Macmillan.
- Schatz, G. (2014) The faces of Big Science. *Nature Reviews Molecular Cell Biology*, 15, 423–426.
- Schubert, A. – Braun, T. (1990) International collaboration in the sciences 1981–1985. *Scientometrics*, 19, 3–10.
- Stefan, D. C. – Seleiro, E. (2016) International Collaboration in Cancer Research. In: Stefan D. (ed.): *Cancer Research and Clinical Trials in Developing Countries*. Cham: Springer, 191–199.
- Tang, M. – Joensuu, H. – Simes, R. J. – Price, T. J. – Yip, S. – Hague, W. – Sjoquist, K. M. – Zalcborg, J. (2019) Challenges of international oncology trial collaboration—a call to action. *Brit J Cancer*, 121, 515–521.
- Theil, S. (2015) Why the Human Brain Project Went Wrong—and How to Fix It. Oct 1 (cited 14 April 2020). *Scientific American* (Internet). New York, NY: Springer Nature – (about 26 screens). <https://www.scientificamerican.com/article/why-the-human-brain-project-went-wrong-and-how-to-fix-it/>
- Tijssen, R. – Van Leeuwen, T. (2006) Centres of research excellence and science indicators. Can ‘excellence’ be captured in numbers? In: Glänzel, W. (ed.): *Ninth International Conference on Science and Technology Indicators*. Leuven, Belgium: Katholieke Universiteit Leuven, 146–147.
- Tijssen, R. – Visser, M. – Van Leeuwen, T. (2002) Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54, 381–397.

- Tóth, G. – Juhász, S. – Elekes, Z. – Lengyel, B. (2018) Inventor collaboration and its persistence across European regions. arXiv,1807.07637 (Preprint). (cited 2020 May 19). <https://arxiv.org/abs/1807.07637>
- U.S. Congress, Office of Technology Assessment (1995) *International. Partnerships in Large Science Projects*. Washington, DC: U.S. Government Printing Office.
- Underwood, E. (2016) International brain projects proposed. *Science*, 352/6283, 277–278
- Wagner, C. S. – Park H. W. – Leydesdorff, L. (2015) The continuing growth of global cooperation networks in research: A conundrum for national governments. *PLoS ONE*, 10/7, e0131816.
- Wagner, C. S. – Whetsell, T. A. – Leydesdorff, L. (2017) Growth of international collaboration in science: revisiting six specialties. *Scientometrics*, 110, 1633–1652.
- Waltman, L. – Tijssen, R. J. W. – Eck, N. J. V. (2011) Globalisation of science in kilometres. *J Informetr.*, 5, 574–582.
- Wang, L. – Wang, X. – Philipsen, N. J. (2017) Network structure of scientific collaborations between China and the EU member states. *Scientometrics*, 113, 765–781.
- Wu, J. (2013) Geographical knowledge diffusion and spatial diversity citation rank. *Scientometrics*, 94, 181–201.
- Wu, L. – Wang, D. – Evans, J. A. (2019) Large teams develop and small teams disrupt science and technology. *Nature*, 566/7744, 378–382.
- Wuchty, S. – Jones, B. F. – Uzzi, B. (2007) The increasing dominance of teams in production of knowledge. *Science*, 316/5827, 1036–1039.
- Zerhouni, E. A. – Potočník, J. (2008) European union and NIH collaborate. *Science*, 322/5904, 1048

S1 Table: Correlation between inter-city link strength

		1	2	3	4	5	6	7
1	C_{94-96}	1						
2	J_{94-96}	0.47	1					
3	C_{06-06}	0.66	0.01	1				
4	J_{04-06}	0.34	0.58	0.42	1			
5	C_{14-16}	0.69	-0.17	0.83	-0.01	1		
6	J_{14-16}	0.35	0.48	0.23	0.50	0.31	1	
7	$C_{14-16HCP}$	0.59	-0.23	0.74	-0.03	0.93	0.02	1
8	$J_{14-16HCP}$	0.25	0.26	0.12	0.21	0.15	0.47	0.02



S1 Fig: Mean Jaccard index by distance