

RESEARCH

Open Access



Cities and countries in the global scientist mobility network

Luca Verginer^{1*}  and Massimo Riccaboni^{2,3}

*Correspondence:

lverginer@ethz.ch

¹Chair of Systems Design, D-MTEC, ETH, Weinbergstrasse 56/58, 8092 Zürich, Switzerland

Full list of author information is available at the end of the article

Abstract

Global mobility and migration of scientists is an important modern phenomenon with economic and political implications. As scientists become ever more footloose it is important to identify general patterns and regularities at a global scale and how it impacts a country's scientific output. The analysis of mobility and brain circulation patterns at global scale remains challenging, due to difficulties in obtaining individual level mobility data. In this work we trace intercity and international mobility through bibliographic records. We reconstruct the intercity and international mobility network of 3.7 million life scientists moving between 5 thousand cities and 189 Countries. In this exploratory analysis we offer evidence that international scientist mobility is marked by national borders and show that international mobility boosts the scientific output of selected countries.

Keywords: Global mobility, Mobility, Research policy

Introduction

Scientists are highly mobile professionals, especially in the early phase of their careers. The tendency to move has been observed in the past (Cardwell 1972; Mokyr 2016), but the size of the phenomenon has drastically increased over the years in a globalized market for high-skill labour (Culotta 2017; Geuna 2015; OECD 2017). Modern economies require a highly skilled labour force to maintain their competitive advantage and grow (Chambers et al. 1998; Solimano 2008; Ozden and Rapoport 2018; Zucker and Darby 2007). The economic relevance makes it essential to understand the structure and the evolution of this kind of mobility at a global scale. However, individual-level mobility data is challenging to collect and is the primary reason for the lack of high resolution and large scale investigations of the phenomenon. Despite the importance of understanding the global mobility of high-skill labour for education, migration and innovation policies, evidence and literature are scant (Fortunato et al. 2018). Previous research on the mobility of scientists has relied on large-scale surveys (Franzoni et al. 2012; Franzoni et al. 2014; Scellato et al. 2017; Franzoni et al. 2018; Petersen 2018), and more recently massive bibliographic databases (Bohannon and Doran 2017; Deville et al. 2014; Graf and Kalthaus 2018).

There are other sources of mobility information (e.g. Job search portals, social media). However, papers offer the most direct and high-frequency signal of scientific activity.

We contribute to the literature on scientific and high skill labour mobility by constructing and analyzing a large scale and global scientist mobility dataset of 3.7 Million scientists working in 189 Countries and 5,531 cities. In this work, we address through an exploratory analysis three questions addressing how cities and countries are affected by international scientist mobility. Specifically, we look at (1) how the centrality of cities in the global mobility network has evolved, (2) how national borders and cultural similarity constrain intercity and international mobility and finally (3) which countries benefit most from the international exchange.

We take advantage of the fact that scientists, especially in some disciplines, regularly publish in their career, and the affiliations listed on publications can be geo-referenced. We are taking inspiration from bibliographic approaches and use MEDLINE, a large open-access publications repository primarily covering research in the life sciences. We reconstruct the mobility paths of scientists through their publication history in MEDLINE, using disambiguated author names (AUTHOR-ITY Torvik and Smalheiser (2009)) and geo-referenced affiliation records (MAPAFFIL Torvik (2015)) as well as journal impact scores (SCIMAGO [SCImago Journal & Country Rank \[Portal\]](#)). With this data, we reconstruct individual level publication histories with affiliation and impact scores. Moreover, we look not only at international mobility but also at intercities moves to capture within-country mobility.

International mobility of high skill-labour is associated with “Brain Drain”, the idea that high skill labour leaves their home country to its detriment and benefit for the receiving country. Several authors (Saxenian 2005; Agrawal et al. 2006; Agrawal et al. 2011) have pointed out that there are positive spillover effects to the sending country, highlighting that global mobility is not a zero-sum game, suggesting that a more fitting term to describe the mobility of high skill labour is “brain circulation”. In this work, we will not assess the causal link between scientist mobility and spillover effects, e.g., from diasporas or international collaboration. We take a high-level view of the international mobility of scientists looking at the effect of national borders, the evolution of the centrality of cities and the benefits to countries to characterize this data for future research. Moreover, we quantify and discuss the impact of international and intercity scientists mobility and how it relates to scientific output. We address first which “mobility communities” are present in the data and subsequently discuss which countries benefit most from international turnover. Note that we do not have information on the nationality of the authors, and when talking about mobility, we do not talk about *migration*, which would require this information.

The rest of the paper is structured as follows. We introduce first the data and methodology to extract individual career trajectories and the mobility network. We then describe which cities lie at its centre using the widely used PageRank centrality measure. To highlight how national borders constraint mobility, we use a community detection approach. Then we discuss which countries benefit most from international mobility of scientists by estimating scientific output growth due to international turnover. Finally, we discuss the implications of these results and offer ideas for future research.

Data

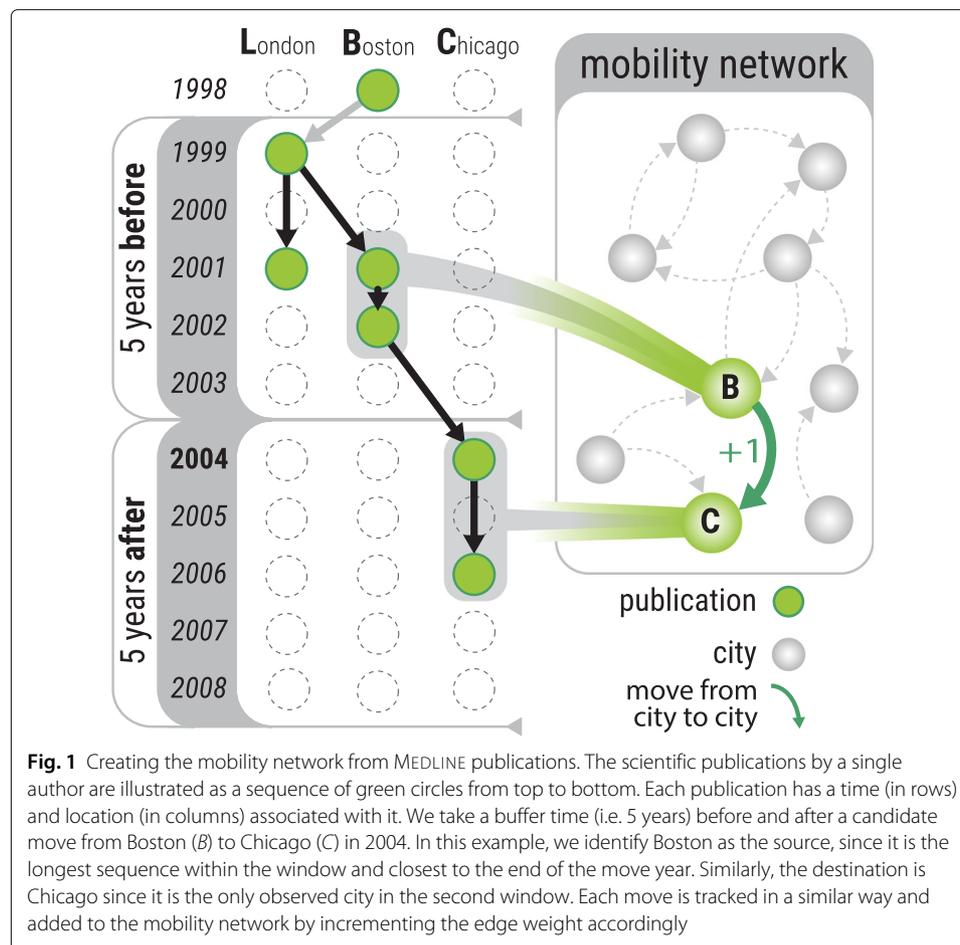
For the analysis of scientist mobility we use four datasets, MEDLINE, AUTHOR-ITY, MAPAFFIL, and SCIMAGO. MEDLINE provides open access to more than 26 million records of scientific publications, with most of the corpus covering research in the life

sciences. The data goes as far back as 1867 (earliest publication in the dataset) and is updated continuously. However, we focus on papers in the period from 1990 to 2009. We restrict our analysis to this period to have good coverage and make use of existing high-quality disambiguations of scientists (AUTHOR-ITY) and affiliations (MAPAFFIL), which are restricted to this time interval. MAPAFFIL lists for a large portion of MEDLINE papers the disambiguated city corresponding to the affiliation of each author as listed on the paper (ca. 37,396,671 author-locations) and is freely available for download from www.nlm.nih.gov. AUTHOR-ITY contains the disambiguated names of 61,658,514 appearances of names on MEDLINE papers (author-name instances). These author-name instances have been mapped to 9,300,182 disambiguated authors. MAPAFFIL, is a disambiguation of affiliations listed on MEDLINE papers. This dataset allows us to map the affiliation string to the city this affiliation is located in. SCIMAGO is a publicly accessible dataset of annual journal impact scores (SCImago Journal & Country Rank [Portal]) and is freely accessible at <https://scimagojr.com>.

With the extracted publication, we can reconstruct the path for a given author over time, as witnessed by the affiliations on the papers the author publishes. In other words, we have a path for author i over several years, indicating where she passed through. It might and does happen, that an author has multiple publications in the same year as well as multiple locations. Possible reasons for multiple locations are that the author had multiple affiliations, or that the publication took some time to publish and an earlier affiliation is listed. Here we define what a move is and how we extract it from the empirically observed publication sequences. To determine a move, and just as importantly a non-move, we define mobility by determining the location of an author within a given time window before a year of interest (t) (i.e. the move year) and assess where she is located, in the window after.

More specifically to determine the source and destination of a move, for a given time interval we chose a candidate move-year (t) and several buffer years (b) around it (see Fig. 1). To transform a publication path into a single edge representing a move, we proceed as follows. We chose a “move year” t of interest. The move year represents the year around which the decision to move happened. Next, we choose a number b of years around t defining two windows: **before** $[t - b, t)$ and **after** $[t, t + b)$. Given these two windows, we proceed to determine in which location any given author was before and after. If the locations differ, then the author moved. Otherwise, she stayed.

To determine a unique starting position in window $[t - b, t)$, we choose the longest uninterrupted sequence of locations closest to t . Take, for example, the observed publication sequence as illustrated in Fig. 1. Here we have the publication history $\{B_{1998}, L_{1999}, L_{2001}, B_{2001}, B_{2002}, C_{2004}, C_{2006}\}$, move year 2004 and a buffer of $b = 5$ years before and after. The Uppercase letter indicates the city and index the year. To determine the starting location, we take all publications in the interval $[1999, 2004)$ and chose the locations with the longest sequence closest to 2004. In this example, we observe 3 publications in B , but only 2 of these are within the $[1999, 2004)$ window, so we discard B_{1998} . On the other hand, we observe 2 publications in L and one simultaneously with B . According to the rule mentioned above, we chose B as the source since it is closest to 2004 even though both L and B have 2 observations. As the destination of the move we chose C since in this case, it is the only observed location in the window $[2004, 2009)$.



We chose this method since it discards ambiguous affiliations in publication sequences with spurious affiliations (e.g. multiple affiliations in the same year but either of these appears only once). This definition allows us to carry out several robustness checks in generating the network. For example, we can increase the number of publications required in each location before and after to reduce the chance that a move was only temporary (e.g. visiting or double affiliations). For this method we require precisely one source and one destination location, which jointly define a move, it would certainly be possible to include double affiliations, but then our definition of “move” would no longer be unambiguous. For example, an author has two locations before, and two locations after, in this case, we would need to employ a convention of how to treat this case, e.g. all 4 possible links but instead of a weight of 1 we use a weight of 1/4. To reduce the number of assumptions and keep the method as simple as possible, we have opted to use the proposed method, which yields only simple moves, i.e. one source and one target. Similarly, we can restrict the size of the windows, thus requiring that authors have fewer holes in their publication history, however, doing so will drop any scientist not publishing at least once in the two periods. Note that a mobility network is a snapshot of aggregated inventor level simple moves (i.e. one source and one target). Thus we consider only *one* move.

When analysing the impact of mobility on the scientific output, we will also rely on the impact factor of the journal the paper was published in using SCIMAGO.

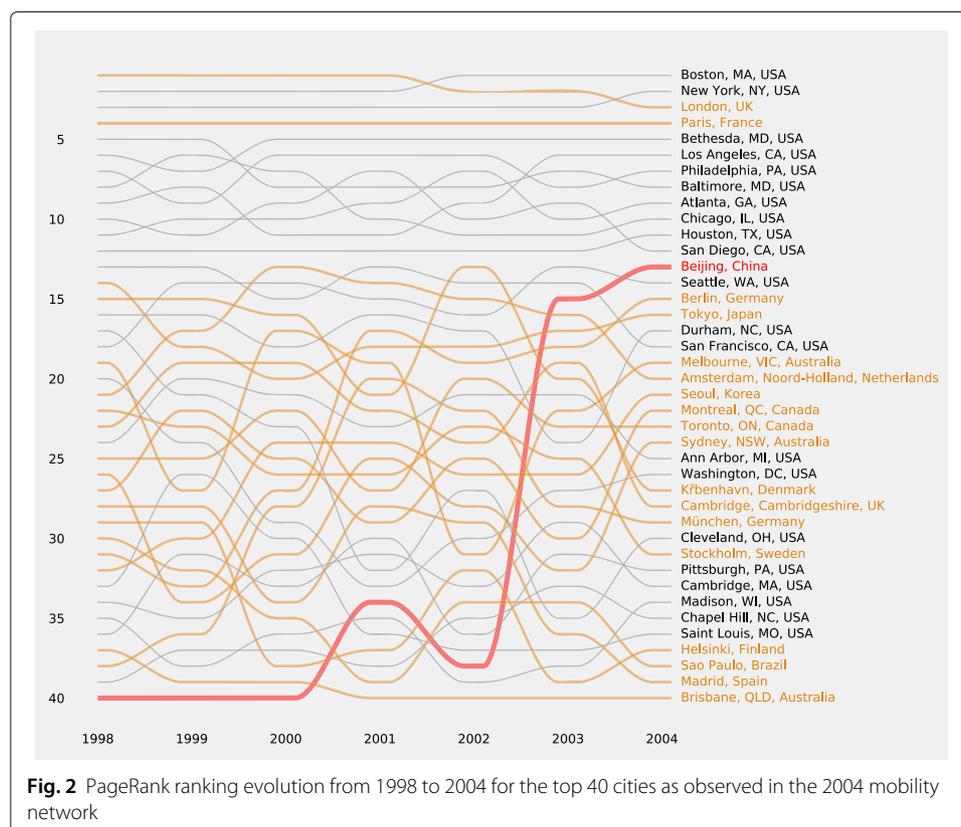
Global cities as hubs

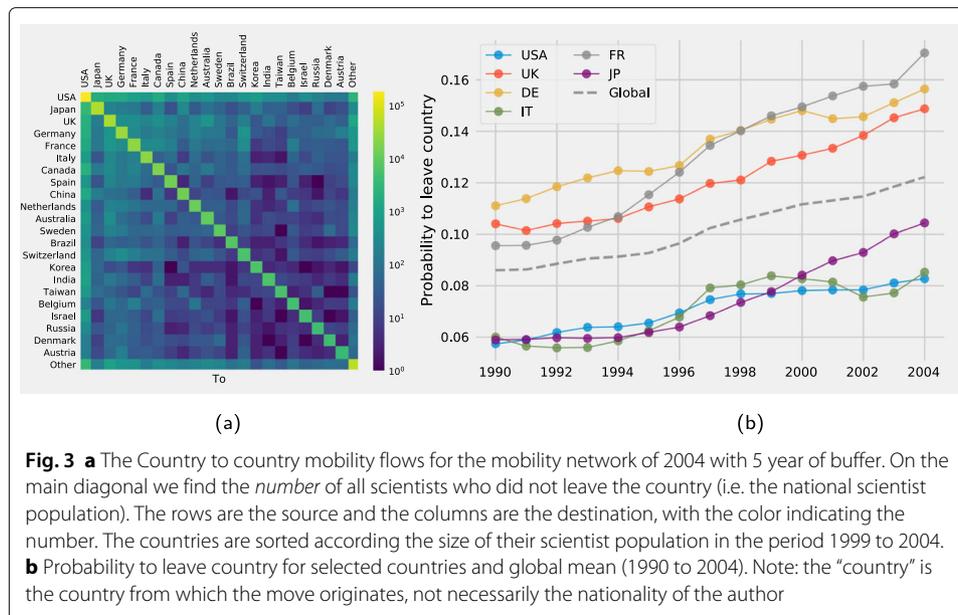
Which cities are at the centre of the exchange of life scientists? How do different countries fare in this comparison? To answer these questions, we look at the topological centrality of cities in the international mobility network extracted as described before. Explicitly, we compute the PageRank centrality of cities in this weighted and directed network from 1998 to 2004. A bump-plot, i.e., a plot showing the changes in ranking over time, for this measure is shown in Fig. 2.

The top 5 cities by centrality in the mobility network are Boston, New York, London, Paris and Bethesda in that order. Except for Boston and New York overtaking London and Paris, the top 5 cities in the international mobility network did not change. Among the top 10 cities, there have been some changes in ranking, but overall the cities in this group have remained the same from 1998 to 2004. Note that 8 out of these top 10 locations are situated in the United States. The dominance of the US in the ranking suggests that the global mobility network is influenced in large part by US cities. However, looking at the top 40 cities, we see that the rest of the world, is better represented, but that the positions in the rankings are changing significantly over time. Among these cities, Beijing stands out by going from lower ranks in 1998 to 11th places in 2004.

National border effects

Co-authorship networks have been found by Hoekman et al. (2010); Chessa et al. (2013) to be influenced by national borders resulting in collaborations being more likely within

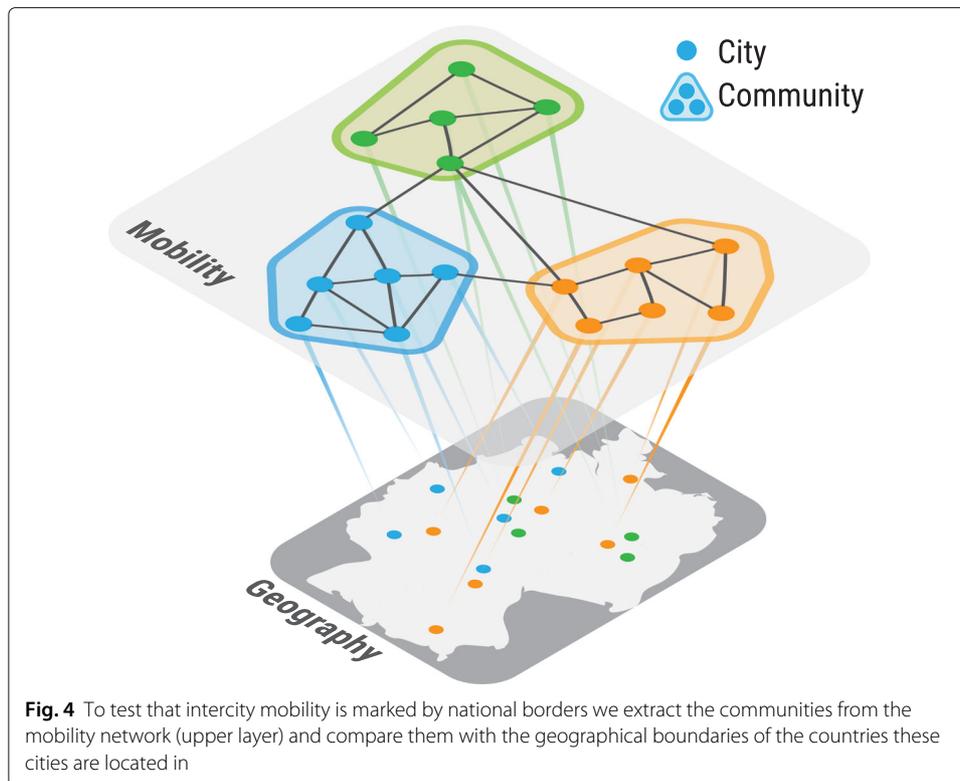




than across countries. In line with these findings, we test the hypothesis that countries have a stronger within than across mobility.

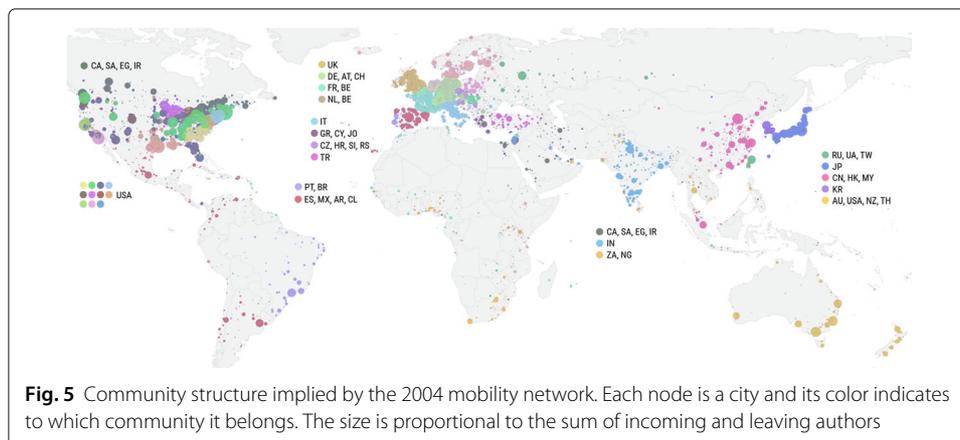
Figure 3a shows the pattern of cross country mobility in 2004. Most scientists do not leave their country (as indicated by the main diagonal). Note also that certain countries have few exchanges with all other countries, as indicated by having only a few off-diagonal elements brighter than the rest. This means that while the network is dense (i.e. all major countries have at least one exchange), there are preferences. Note also that the probability of leaving the country has increased steadily year by year as can be seen in Fig. 3b. The global probability of observing a move, i.e. that any given scientist moves abroad if we look at five years before and after, has never dipped since 1990. The listed countries fall into two categories, below the global mean and above. The US, Japan and Italy are below the global average, indicating a stronger within mobility. Moves originating from the US tend to be mostly within the US. This number had gone from 5% in 1990 to 8.1% in 2004. However, compared to France (16.8%) and the global average (12%), it is low. Note that scientists based in the US do not leave the country as often as most other countries, but there is a substantial domestic mobility.

The international mobility patterns in Fig. 3 suggest that international mobility varies by country and that there is more mobility within than across. The notion of “more within” and “less across” is made precise by the measure of *modularity* (Newman and Girvan 2004). At a high level, modularity is a quality score of how well a given partitioning of nodes (i.e. set of cities) separates nodes which are well connected but have few ties to members of other partitions. More specifically, modularity measures the ratio of links falling within a given partition minus the ratio of links we would expect from a random network (see Newman and Girvan (2004) for more detail). Thus this null model represents a mobility network where scientists move without regard for geographic proximity or national borders. Conceptually we carry out the analysis shown in Fig. 4, where we want to see if the community structure in the topology of the mobility network (mobility layer) coincides with national borders in the (geography layer). We estimate the communities



by maximizing the modularity of the partition following the Louvain algorithm (Blondel et al. 2008) implemented by Traag (2017).

If the null hypothesis that scientists move without regard for national borders were correct, we should find that the community structure we obtain by maximizing the modularity does not coincide with any geographic or political boundaries. The spatial organization of the communities; however, as shown in Fig. 5 reveals that national borders are geographically clustered and respect national borders. A breakdown of countries as they fall within the various communities in 2004 is available in the Appendix (Table 4). For example, we find, especially in Europe, that national borders coincide with the spatial boundaries of the mobility communities. However, the picture changes when looking at North America. Here we also observe a national component in the form of Canada and



Mexico being identified as separate communities. However, within the US, the identified communities are less spatially segregated than in the rest of the world. Beyond the pure border effect, the community structure reveals some new patterns. We see that countries sharing a language or more generally are culturally similar are more likely to fall within the same community. For example, three majority German-speaking countries, Germany, Austria and Switzerland are identified as belonging to the same mobility community. Several former French colonies in North Africa are placed in the majority French community, suggesting that a scientific exchange persists. Belgium's cities, on the other hand, are split along the countries language divide (French, Dutch), mirroring findings on the same divide using mobile phone data (Sobolevsky et al. 2013). Even more strikingly is the placement of Spain and Portugal in different communities. The two countries share a border but not a language. However, as Fig. 5 shows, Portugal and Brazil have more exchange among themselves than Portugal has with Spain even though one is across the ocean and the other a next-door neighbour. Similarly, Spain and Mexico are placed in the same community, both countries share a colonial history and language, as do Portugal and Brazil. This result would suggest that scientific mobility is influenced by language and possibly by cultural similarity.

We should note that community detection through modularity maximization may fail to separate communities which are "too small" due to the method's "resolution limit" (Fortunato and Barthelemy 2007). Ground truth communities, which are not of comparable size to the identified communities, may be lumped together with larger communities or split up. In practice, this could mean that we have lumped "small" communities together, which probably should be kept separate, for example, Greece, Cyprus and Jordan are placed in the same community. While Greece and Cyprus share a language, the inclusion of Jordan in this community is most likely because Jordan has had an exchange with the other two but was "erroneously" placed in the same community.

National gains

The ability of a country to be at the forefront of research and innovation is in part determined by its ability to attract bright and talented scientists, in addition to retaining the highly trained individuals already working for national institutions. For this reason, high skill labour mobility and brain circulation is a significant concern at the country level. To estimate which countries are the primary beneficiaries of international mobility, we estimate the contribution to the national scientific output growth coming from the mobile scientist population. By doing so, we can compare scientific output across countries and identify the primary direct beneficiaries from international mobility.

Before we can discuss the impact of mobility on scientific output, we need to define how to measure scientific output. Only relying on the number of publications is not a good proxy for scientific relevance. We, therefore, rely on the impact score of the journal a paper was published in. Specifically, we use the number of citations per document in the two years before. We obtain this information from SCIMAGO. To compute the scientific production of a given location for a given period, we obtain the papers listing that city among the affiliations of any of the authors and obtain the impact factor for that journal in that year. If multiple authors are working in different cities, we apportion this score equally. For example, for a paper published in a journal with an impact of 6 and 3 authors, 2 of which reside in city A and 1 in city B. We would then add $6/3 * 2 = 4$ to the running

total of city A and $6/3 * 1 = 2$ to B. For given year y and period, i.e., *before* $[y - 5, y)$ and *after* $[y, y + 5)$, the total scientific output is the sum of the fractional impacts associated with that city. At country-level, these impact factors are summed up.

The total scientific output produced within a country can be accounted for in the following way. Scientific output produced by authors *staying in their city* (S), *moving domestically* (D), *coming in from abroad* (I) and *leaving the country* (L). The total output for a given time period within a country before A_0 and after A_1 are given by $A_0 \equiv S_0 + D_0 + L_0$ and $A_1 \equiv S_1 + D_1 + I_1$ respectively. Note that in A_0 the output contains the production of those individuals who will leave the country L_0 in the second period and A_1 the production of those that will come in the second period I_1 . Based on this breakdown, we can define indicators identifying the growth due to the four mobility types, i.e., S , D , I and L . Specifically this is the output after O_1 net of the output before O_0 divided by the output O_0 before, i.e., $g = (O_1 - O_0)/O_0$. The indicators are defined in detail in Table 1. Overall growth g_A for the country, g_S growth due to stationary scientists, g_D growth due to *domestically* mobile scientists and most relevant for the brain circulation discussion g_I , the gain due to *international* turnover. Moreover, the growth of a country g_A can be expressed as a weighted sum of the individual growth rates g_S , g_D and g_I as shown in Eq. 1, where the weights are w_S , w_D and w_I respectively.

$$g_A = (A_1 - A_0)/A_0 = \frac{(S_1 + D_1 + I_1) - (S_0 + D_0 + L_0)}{A_0} \tag{1}$$

$$g_A = \frac{S_1 - S_0}{A_0} + \frac{D_1 - D_0}{A_0} + \frac{I_1 - L_0}{A_0}$$

$$g_A = \frac{S_1 - S_0}{S_0} \frac{S_0}{A_0} + \frac{D_1 - D_0}{D_0} \frac{D_0}{A_0} + \frac{I_1 - L_0}{L_0} \frac{L_0}{A_0}$$

$$g_A = g_S * w_S + g_D * w_D + g_I * w_I$$

We also report these weights to understand the importance of the contribution of each growth component. Additionally, to indicate the generational turnover, we also report the mean difference between incoming and leaving scientists, $\Delta \overline{age}$. The results for the

Table 1 Country level Brain Circulation indicators

	Definition	Description
θ		0 before the move and 1 after
S_θ		Output of stationary scientist in domestic cities
D_θ		Output of scientist moving between domestic cities
I_θ		Output of scientist coming from a foreign city to a domestic city.
L_θ		Output of scientist leaving the country for a foreign city.
A_0	$S_0 + D_0 + L_0$	Total Output in the country before the move year
A_1	$S_1 + D_1 + I_1$	Total Output in the country after the move year
g_A	$(A_1 - A_0)/A_0$	National output growth of output
g_S	$(S_1 - S_0)/S_0$	Output growth from stationary scientists
g_D	$(D_1 - D_0)/D_0$	Output growth from domestically mobile scientists
g_I	$(I_1 - L_0)/L_0$	Output growth from international exchange
\overline{age}_I		Average age (years from first publication) for incoming
\overline{age}_L		Average age of leaving scientists
$\Delta \overline{age}$	$\overline{age}_L - \overline{age}_I$	Age difference between leaving and incoming scientists

largest countries in the dataset for the intervals before [1999, 2004) and after [2004, 2009) are reported in Table 2.

We do find that the US has increased its overall output by 14%, but that the gain due international exchange was as high as 61%. While overall, only 6% of the scientific production comes from these scientists, it is clear that their contribution is a net benefit to the US research system. Similarly, the contribution to the national scientific output is positive for most countries, with several exceptions like Argentina, India and Israel, to name a few. In this comparison, China stands out, with an astonishing growth rate of 141% overall and 158% due to international exchanges. Overall, this result suggests that China is a prime example of a country gaining by participating in the international exchange of scientists. If we look at the $\Delta\overline{\text{age}}$ value in Table 2 we also note that the scientists moving to the US are younger than the scientists leaving, on average 1.2 years younger. On the other hand and line with the idea that the increasing scientific output of China is due to returnees, we find that scientists moving to China are on average 1.16 years older. Give these two statistics we argue that the US can “rejuvenate” their scientific labour force and China can entice more senior ex-pats to move back. As for the proportion of scientific output due to international mobility, i.e., w_I , we see that for several countries it is a significant proportion. So, for example the scientific output growth of China, India, Argentina, Russia and Switzerland has a weight of more than 20% ($w_I > 0.2$), suggesting a strong exposure to international mobility. From these statistics, a picture emerges that certain countries are more exposed than others, and the gain is not unequivocally positive. Note that this is only a direct measure of scientific output and disregards any other benefits to

Table 2 National scientific output growth for selected countries

	g_A	$g_S (w_S)$	$g_D (w_D)$	$g_I (w_I)$	$\Delta\overline{\text{age}}$
Argentina	0.05	0.23 (0.71)	0.79 (0.02)	-0.48 (0.27)	+0.85
Australia	0.32	0.28 (0.80)	0.24 (0.07)	0.56 (0.13)	+0.40
Austria	0.18	0.22 (0.77)	-0.07 (0.04)	0.09 (0.18)	+0.01
Belgium	0.23	0.28 (0.79)	0.15 (0.07)	0.00 (0.14)	+0.02
Brazil	0.46	0.53 (0.78)	0.57 (0.08)	0.01 (0.14)	+0.77
Canada	0.20	0.16 (0.73)	0.19 (0.11)	0.36 (0.16)	+0.06
China	1.41	1.17 (0.62)	2.12 (0.15)	1.58 (0.23)	+1.16
Denmark	0.18	0.18 (0.81)	0.18 (0.06)	0.17 (0.13)	+0.07
Finland	0.03	0.08 (0.77)	0.02 (0.10)	-0.21 (0.13)	+0.14
France	0.10	0.14 (0.76)	0.12 (0.09)	-0.10 (0.16)	+0.66
Germany	0.16	0.18 (0.66)	0.16 (0.19)	0.08 (0.15)	-0.02
India	0.42	0.65 (0.66)	0.73 (0.10)	-0.33 (0.24)	+2.27
Israel	0.14	0.22 (0.74)	0.27 (0.09)	-0.23 (0.17)	+2.54
Italy	0.32	0.32 (0.82)	0.31 (0.10)	0.36 (0.08)	+0.49
Japan	0.09	0.12 (0.66)	0.13 (0.24)	-0.21 (0.10)	+1.85
Korea	0.71	0.66 (0.63)	0.87 (0.19)	0.76 (0.18)	+0.85
Netherlands	0.26	0.28 (0.74)	0.18 (0.13)	0.19 (0.12)	-0.29
Russia	0.13	0.27 (0.73)	0.41 (0.01)	-0.27 (0.26)	+0.66
Spain	0.35	0.35 (0.81)	0.21 (0.07)	0.44 (0.11)	+1.23
Sweden	0.10	0.18 (0.74)	0.13 (0.08)	-0.25 (0.18)	+0.40
Switzerland	0.14	0.08 (0.67)	0.02 (0.08)	0.33 (0.24)	-0.42
Taiwan	0.37	0.37 (0.74)	0.51 (0.15)	0.17 (0.11)	-1.04
UK	0.16	0.14 (0.74)	0.25 (0.13)	0.17 (0.13)	-0.12
USA	0.14	0.09 (0.71)	0.17 (0.23)	0.61 (0.06)	-1.20

a country from participating in the international exchange of scientists such as diasporas and sustained international collaborations (Saxenian 2005; Agrawal et al. 2006; Agrawal et al. 2011). These measures do, however, offer a first-level approximation of the primary beneficiaries of international scientific mobility.

Discussion

The mobility of scientists and its impact on scientific production are still poorly understood, mainly due to the lack of harmonized international statistics. We contribute to filling this gap by providing a description of scientist mobility at the city and country level. The mobility network highlights the existence of a highly connected core of cities. However, the reach and mobility of scientists is constrained by geographical factors. These hubs are predominantly found in the US, which may benefit from high levels of domestic and low cultural and linguistic barriers to international mobility. This hypothesis is supported by the community structure of the mobility network, which, unlike in many other countries, is not delineated by a clear spatial segregation. Moreover, we find that not all countries benefit equally from international exchange. In most cases, the direct benefit, without considering any indirect benefits such as international collaborations and diasporas, is ambiguous. In other words, gains in national scientific output, as highlighted by the output growth due to international turnover, do not provide a clear signal that international exchange is unequivocally beneficial to all participants. The list of countries negatively affected by international mobility is not limited to less developed and emerging countries, such as India. However, it includes scientific powerhouses such as Japan and Sweden, for which the balance of international scientist turnover is negative. Among the listed countries, China stands out by having the most substantial overall growth by far (141%), while also benefiting from international exchange. As argued above, this is in part explained by the fact that on average incoming scientists are more senior than the leaving scientists and that these leaving scientists are less prolific than the incoming scientists. In other words, China did benefit exceptionally from international mobility by sending scientists abroad and attract them back. With the returning scientists, China has been able to build up their local innovation system substantially. These observations highlight the importance to investigate international mobility not purely as “brain drain” but rather as “brain circulation” as argued by Agrawal et al. (2006); Saxenian (2005). The analysis on community structure implied in the mobility network, reveals that national borders do affect the mobility patterns we observe, a finding which mirrors the observed tendency to form collaboration within national borders (Cerina et al. 2014; Hoekman et al. 2010). Further analysis on the determinants on which scientists move abroad and the potential benefit of this turnover, can inform policies to enhance international collaborations, such as the European Research Area (ERA) initiative.

This is an exploratory investigation of the international and intercity mobility network and does only provide a high-level overview of the phenomenon. The macro patterns, however, we have identified raise several questions and possible future research directions regarding, e.g., the push and pull forces driving mobility and how emerging and developed economies have fared during the globalization and integration of research. The primary take-home message from this analysis is that scientist mobility has a strong spatial component and individual level career trajectories offer a window into this phenomenon.

In summary, this study makes three contributions. First, it introduces an approach to extract mobility networks from bibliographic data and augments it with quality indicators. Second, it characterizes the international flows of scientists highlighting the importance of national barriers. Third, it quantifies the gains from mobility to countries. This study has several strengths. We reconstruct intercity mobility networks for specific time intervals, making it potentially useful for evaluating the impact of research policies. The dataset has extensive coverage of life scientists spanning multiple countries, career stages and productivity levels (i.e. not only star scientists). However, it has also some limitations. First, by using PUBMED, there are linguistic and field biases, containing mostly English journals in the life sciences. More importantly, however, we lack biographical details, such as nationality and ethnicity, which would better inform migration policies. Our contribution, which is based on open access data, is meant to stimulate follow-on research to investigate the determinants of mobility, including the impact of reverse brain drain and research policies to attract the best talent. Future research can shed more light on the role of collaboration networks, both as drivers of mobility and as vectors of knowledge diffusion. Moreover, more work is needed to understand the attractiveness of global cities and their role in the global knowledge economy. In this work, we have limited ourselves to a descriptive analysis of the mobility network, omitting causality claims. However, the richness of the dataset makes it potentially useful for use in determining causal relocation factors. The global nature and good temporal coverage mean that several natural experiments can be identified, which can help to isolate the determinants of mobility. An example of this is the estimation of the impact of stem cell legislation on the US on stem cell scientist mobility (US states offer various degrees of support). Similarly, the effect of regional projects (e.g. opening a new research campus), aiming to improve scientific output or innovation, can be quantitatively analyzed as part of the national or international research system. This dataset, in conjunction with natural language processing techniques and text mining, can also be used to follow the mobility and diffusion of new ideas and concepts in the life sciences. By estimating the relative importance of mobility and collaboration research policies, optimizing diffusion could be devised. Moreover, by exploiting the available information on collaborations and the data on mobility, we can estimate the propensity of scientists to remain in contact with their home country and city. This approach could be used to replicate findings on the benefit of international mobility and their spillovers which go deeper into the issue of brain drain and circulation.

Appendix

Table 3 Example of career path of a specific author (Lisa M. Shulman). For each record we have the year of publication, the city of the affiliation and the relative PubMed ID identifying the paper. The algorithm would record a move from Miami to Baltimore in 2001 and no moves in any other years

Year	locations	pmids
1995	Miami, FL, USA	7651440
1995	Miami, FL, USA	7542376
1995	Miami, FL, USA	7675237
1995	Miami, FL, USA	7481139
1996	Miami, FL, USA	8618701
1996	Miami, FL, USA	8771085

Table 3 Example of career path of a specific author (Lisa M. Shulman). For each record we have the year of publication, the city of the affiliation and the relative PubMed ID identifying the paper. The algorithm would record a move from Miami to Baltimore in 2001 and no moves in any other years (*Continued*)

Year	locations	pmids
1996	Miami, FL, USA	8576519
1996	Miami, FL, USA	8771060
1997	Miami, FL, USA	8990066
1998	Miami, FL, USA	9596026
1999	Miami, FL, USA	11096691
1999	Miami, FL, USA	10025817
1999	Miami, FL, USA	10563642
2000	Miami, FL, USA	11054153
2000	Miami, FL, USA	10928576
2000	Miami, FL, USA	10714670
2000	Miami, FL, USA	10634252
2001	Baltimore, MD, USA	11763581
2001	Baltimore, MD, USA	11391746
2002	Baltimore, MD, USA	15177058
2002	Baltimore, MD, USA	12039431
2002	Baltimore, MD, USA	12210879
2003	Baltimore, MD, USA	12578945
2006	Baltimore, MD, USA	16894111
2006	Baltimore, MD, USA	16482533
2006	Baltimore, MD, USA	16533144
2007	Baltimore, MD, USA	17584622
2008	Baltimore, MD, USA	18361474
2008	New York, NY, USA	18695158
2009	Baltimore, MD, USA	19597081
2009	Baltimore, MD, USA	18693062

Table 4 Breakdown of communities by country. For each modularity class the number of cities belonging to a country are listed along with their proportion of cities in the class. For example community 16 is composed of 25% Swedish, 22% Finnish, 20% Norwegian and 16% Danish cities as well as 17% smaller cities. Note: only countries with at least 5% member cities are listed

Community (x)	Country (c)	% of cities of c in x	No. Cities
1	USA	92%	392
2	USA	91%	161
3	Spain	27%	65
	Mexico	22%	53
	Argentina	11%	27
	Chile	6%	14
	USA	6%	15
4	USA	94%	355
5	USA	98%	197
6	USA	85%	197
7	France	57%	225
	Belgium	10%	40
8	Germany	69%	444
	Switzerland	14%	88
	Austria	7%	46

Table 4 Breakdown of communities by country. For each modularity class the number of cities belonging to a country are listed along with their proportion of cities in the class. For example community 15 is composed of 25% Swedish, 22% Finnish, 20% Norwegian and 16% Danish cities as well as 17% smaller cities. Note: only countries with at least 5% member cities are listed (*Continued*)

Community (<i>x</i>)	Country (<i>c</i>)	% of cities of <i>c</i> in <i>x</i>	No. Cities
9	Russia	45%	53
	Taiwan	21%	25
	USA	20%	23
	Ukraine	9%	10
10	USA	95%	151
11	USA	91%	192
12	Australia	41%	81
	USA	12%	24
	Thailand	11%	22
	New Zealand	10%	20
13	Czech Republic	36%	39
	Croatia	16%	17
	Slovenia	12%	13
	Serbia	11%	12
	Slovakia	9%	10
	USA	6%	6
14	Netherlands	60%	110
15	USA	51%	41
16	Sweden	25%	86
	Finland	22%	75
	Norway	20%	69
	Denmark	16%	55
	Korea	16%	55
17	Korea	88%	43
	USA	10%	5
18	USA	89%	71
19	USA	98%	122
20	Japan	81%	91
21	Canada	54%	127
	Iran	11%	25
	Saudi Arabia	9%	22
	USA	8%	19
	Egypt	7%	17
	Greece	7%	17
22	Greece	80%	35
	Jordan	9%	4
	Cyprus	7%	3
	Belgium	7%	3
23	Brazil	14%	26
	Brazil	58%	55
24	Portugal	21%	20
	Portugal	21%	20
24	Israel	70%	39
	USA	21%	12
25	China	70%	171
	Hong Kong	7%	17
	Malaysia	7%	18
26	Poland	92%	57
27	UK	66%	315
28	Turkey	91%	53
29	Italy	75%	121
	USA	6%	10

Table 4 Breakdown of communities by country. For each modularity class the number of cities belonging to a country are listed along with their proportion of cities in the class. For example community 15 is composed of 25% Swedish, 22% Finnish, 20% Norwegian and 16% Danish cities as well as 17% smaller cities. Note: only countries with at least 5% member cities are listed (*Continued*)

Community (<i>x</i>)	Country (<i>c</i>)	% of cities of <i>c</i> in <i>x</i>	No. Cities
30	USA	83%	84
31	Nigeria	47%	34
	South Africa	39%	28
32	India	73%	122
	USA	8%	14

Acknowledgements

We would like to thank Vetle Torvik and Neil Smalheiser for granting us access to the Authority and MapAffil Datasets and Prof. Frank Schweitzer and Prof. Pietro Panzarasa for comments on an earlier version of this manuscript.

Authors' contributions

LV and MR conceived the idea. LV collected the data and carried out the analysis. LV created the graphics. LV and MR wrote the manuscript. Both authors read and approved the final manuscript.

Funding

Not applicable

Availability of data and materials

The raw XML data on all Pubmed articles is available for download from the NIH at

- https://www.nlm.nih.gov/databases/download/pubmed_medline.html
- <https://ftp.ncbi.nlm.nih.gov/pubmed/baseline>

The Impact factor data has been obtained from SCIMAGO.

- <https://www.scimagojr.com/>

The disambiguation of authors (AUTHOR-ITY) and affiliations (MAPAFFIL) has been obtained from

- <http://abel.lis.illinois.edu/downloads.html>

Access to this resource can be requested for free from the maintainers through the online form on the same page. The merged datasets can be requested from the authors, on reasonable request. Note that due to an agreement with the providers of AUTHOR-ITY and MAPAFFIL, these datasets may only be shared by requesting access through the previously mentioned online form.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Chair of Systems Design, D-MTEC, ETH, Weinbergstrasse 56/58, 8092 Zürich, Switzerland. ²IMT School for Advanced Studies Lucca, Piazza S. Francesco, 19, 55100 Lucca, Italy. ³KU Leuven, Department of Managerial Economics, Strategy and Innovation (MSI), Naamsestraat 69, 3000 Leuven, Belgium.

Received: 6 February 2020 Accepted: 19 June 2020

Published online: 16 July 2020

References

- Agrawal A, Cockburn I, McHale J (2006) Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *J Econ Geogr* 6(5):571–591. <http://dx.doi.org/10.1093/jeg/lbl016>
- Agrawal A, Kapur D, McHale J, Oettl A (2011) Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *J Urban Econ* 69(1):43–55. <https://doi.org/10.1016/j.jue.2010.06.003>
- Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bohannon J, Doran K (2017) Introducing ORCID. *Science* 356(6339):691–692. <https://doi.org/10.1126/science.356.6339.691>
- Cardwell DSL (1972) *Turning Points in Western Technology; a Study of Technology, Science and History*. Science History Pub, New York
- Cerina F, Chessa A, Pammolli F, Riccaboni M (2014) Network communities within and across borders. *Sci Rep* 4:4546
- Chambers E, Foulon M, Handfield-Jones H, Hankin S, Michael III E (1998) The war for talent. *McKinsey Q* 3:44–57. <https://doi.org/10.1080/03071840308446873>
- Chessa A, Morescalchi A, Pammolli F, Penner O, Petersen AM, Riccaboni M (2013) Is Europe Evolving Toward an Integrated Research Area? *Science* 339(6120):650–651. <https://doi.org/10.1126/science.1227970>
- Culotta E (2017) People on the move: The science of migrations. *Science*. <https://doi.org/10.1126/science.aan6884>

- Deville P, Wang D, Sinatra R, Song C, Blondel VD, Barabási A-L (2014) Career on the move: Geography, stratification, and scientific impact. *Sci Rep* 4:4770
- Fortunato S, Barthelemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci* 104(1):36–41. <https://doi.org/10.1073/pnas.0605965104>
- Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B, et al. (2018) Science of science. *Science* 359(6379):0185
- Franzoni C, Scellato G, Stephan P (2012) Foreign-born scientists: mobility patterns for 16 countries. *Nat Biotechnol* 30(12):1250
- Franzoni C, Scellato G, Stephan P (2014) The mover's advantage: The superior performance of migrant scientists. *Econ Lett* 122(1):89–93
- Franzoni C, Scellato G, Stephan P (2018) Context Factors and the Performance of Mobile Individuals in Research Teams. *J Manag Stud* 55(1):27–59. <https://doi.org/10.1111/joms.12279>
- Geuna A (2015) *Global Mobility of Research Scientists: The Economics of Who Goes Where and Why*. Elsevier Academic Press, London
- Graf H, Kalthaus M (2018) International research networks: Determinants of country embeddedness. *Res Policy* 47(7):1198–1214. <https://doi.org/10.1016/j.respol.2018.04.001>
- Hoekman J, Frenken K, Tijssen RJW (2010) Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Res Policy* 39(5):662–673. <https://doi.org/10.1016/J.RESPOL.2010.01.012>
- Mokyr J (2016) *A Culture of Growth: the Origins of the Modern Economy*. Princeton University Press, Princeton
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E* 69(2):026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- OECD (2017) *OECD Science, Technology and Industry Scoreboard 2017*. OECD Publishing, Paris
- Ozden C, Rapoport H (2018) Cross-country perspectives on migration and development: Introduction. *Econ J*
- Petersen AM (2018) Multiscale impact of researcher mobility. *J R Soc Interf* 15(146):20180580. <https://doi.org/10.1098/rsif.2018.0580>
- Saxenian A (2005) From Brain Drain to Brain Circulation: Transnational Communities and Regional Upgrading in India and China. *Stud Comp Int Dev* 40(2):35–61. <https://doi.org/10.1007/BF02686293>
- Scellato G, Franzoni C, Stephan P (2017) A mobility boost for research. *Science* 356(6339):694–2694. <https://doi.org/10.1126/science.aan4052>
- SCImago Journal & Country Rank [Portal]. <https://www.scimagojr.com>. Retrieved 01 Feb 2020
- Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti C (2013) Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries. *PLoS ONE* 8(12):81707. <https://doi.org/10.1371/journal.pone.0081707>
- Solimano A (2008) *The International Mobility of Talent: Types, Causes, and Development Impact* vol. 394. Oxford University Press, Oxford
- Torvik VI (2015) MapAffil: A Bibliographic Tool for Mapping Author Affiliation Strings to Cities and Their Geocodes Worldwide. *D-Lib Mag* 21((11/12)). <https://doi.org/10.1045/november2015-torvik>
- Torvik VI, Smalheiser NR (2009) Author name disambiguation in MEDLINE. *ACM Trans Knowl Disc Data* 3(3):1–29. <https://doi.org/10.1145/1552303.1552304>
- Traag V (2017) *vtraag/louvain-igraph: 0.6.1*. <https://doi.org/10.5281/zenodo.1054103>
- Zucker LG, Darby MR (2007) Star scientists, innovation and regional and national immigration. Technical report. Natl Bur Econ Res

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)